EMERGING TRENDS IN INFORMATION TECHNOLOGY

Dr.T.VELUMANI
Assistant Professor & Head,
Department Of Information Technology,
Rathinam College of Arts and Science(Autonomous),
Coimbatore, Tamil Nadu, India.

Dr.R.LAKSHMI DEVI
Assistant Professor,
Department of BCA,
Women's Christian College,
Chennai, Tamil Nadu, India.

Dr.K.PRABAVATHY
Assistant Professor,
Department of Data Science,
Sree Saraswathi Thyagaraja College (Autonomous),
Pollachi, Tamil Nadu, India.

Dr.V.KALAIMANI
Professor & Head,
Department of MCA,
Rathinam Technical Campus,
Coimbatore, Tamil Nadu, India.

Published by
SK Research Group of Companies
The International Journals, Conferences and Books – SKRGC Publication
142, Periyar Nagar, Madakulam,
Madurai – 625 003, Tamil Nadu, India.
skrgc.publisher@gmail.com / +91 9790120237
www.skrgcpublication.org/isbn
SKRGC Since 2012
SKRGC Publication Read | Write | Teach

TITLE: **EMERGING TRENDS IN INFORMATION TECHNOLOGY**

Edited by: Dr.T.Velumani

Dr.R.Lakshmi Devi

Dr.K.Prabavathy

Dr.V.Kalaimani

Published by: SK Research Group of Companies - SKRGC Publication

Madurai - 625 003, Tamil Nadu, India.

Printers Details: Leilani Katie Press,

Madurai - 625 003, Tamil Nadu, India.

Edition Details: I

ISBN: 978-81-965283-7-9

Month & Year September, 2023

Pages 304

Price Rs.700/-

ACKNOWLEDGEMENT

First and foremost, praises and thanks to the God, the Almighty, for His showers of blessings throughout my research work to complete the **Book Chapter** Under the title ***"Emerging Trends in Information Technology"*** successfully.

I am extremely grateful to my beloved Father **Lt.A.Thiyagarajan** and My Mother **S.Thilakam** for their love, prayers, caring and sacrifices for educating and preparing me for my future. I am very much thankful to my wife **R.Seetha** and my daughter **V.Kiruthiga Sri** for their love, understanding, prayers and continuing support to complete this research work. Also I express my thanks to my sister **T.JothiAruna** and uncle R**.Srinivasan**, brothers, father in law and mother in laws for their support and valuable prayers. My Special thanks goes to my Spouse **Mrs.R.Seetha** for the keen interest shown to complete this Book Chapter successfully.

I would like to say thanks to **Dr. Madan A. Sendhil**, **Chairman**, Rathinam Group of Institutions, **Mrs**.**Shima Sendhil, Director**, Rathinam Group of Institutions, **Dr.R.Manickam, CEO and Secretary**, Rathinam Group of Institutions, **Dr.S.Balasubramanian, Principal**, Rathinam College of Arts and Science (Autonomous), Coimbatore, **Dr. S. N. Suresh, Vice Principal & Dean, School of Science**, **Dr.Sabareesh KPV, Dean of Research & Development** Rathinam College of Arts & Science, for their encouraging and guidance to Publish a **book Chapter.**

I thank the management of Rathinam College of Arts and Science (Autonomous), Coimbatore for their support to do this work. I thank the **School of computer Science** at Rathinam College of Arts and Science (Autonomous) for their genuine support to complete this book Chapter successfully.

The overall Aim of this book Chapter is to compile the latest research and development, up-to-date, challenges in the field of "***Emerging Trends in Information Technology***" with some modern approaches. Last but not least we thank all those who are directly and indirectly supported me to complete the book Chapter Successfully.

Dr.T.VELUMANI
Editor
M.Sc(CS).B.Ed,M.Phil,MBA,M.Sc(Psychology),Ph.D

PREFACE

"Emerging Trends in Information Technology" is a popular subject evolving continuously. This book Chapter covers topics like all computer science streams subjects .The outcome of our teaching experience and research findings at the Department of Computer Science and provides a basic foundation for various topics in Information Technology. In general, the field of **"*Emerging Trends in Information Technology*"** requires sound theoretical knowledge and extensive experimental work involving software simulation and testing with large sets of sample projects. The book Chapter has been designed to meet.

The book Chapter presents a unique overview of recent as well as what are the ***"Emerging Trends in Information Technology"***, it focuses the Integration of all computer science techniques and Applications also. Recent results on invariant Concepts Integration are clearly presented and well explained. Identifying Feature Points is demonstrated. We reviewed Understanding Uncertain Spatial Relationships

The book chapter ***"Emerging Trends in Information Technology"*** from teaching the graduate courses on moment invariants and related fields of all domains and presenting several tutorials on moments at major international conferences.

The target readership of the book Chapter are academic researchers and R&D engineers from all application areas who need to recognize objects extracted Industry 4.0 as well as specialists in moment-based interested in a new development on this field. Last but not least, the book chapter is also intended for university lecturers and graduate students of all computer science fields are analyzed. This knowledge then motivates the presentation of the techniques needed to reverse distortions, minimize artifacts and enhance important features.

Emerging Trends in Information Technology

Editors

Dr.T.VELUMANI

Dr.R.LAKSHMI DEVI

Dr.K.PRABAVATHY

Dr.V.KALAIMANI

CONTENT

ID: 01

DETECTION AND CLASSIFICATION OF LEAF DISEASE USING SVM & IMAGE PROCESSING

Dr.S.GOVINDARAJU
Associate Professor,
Department of Computer Science,
Sri Ramakrishna College of Arts & Science,
Coimbatore, Tamil Nadu, India

Dr.M.INDIRANI
Assistant Professor,
Department of Computer Science and Engineering,
Adithya Institute of Technology,
Coimbatore, Tamil Nadu, India

ABSTRACT

Agricultural production is something on which the economy significantly relies. Agriculture is affected by a variety of factors, biotic such as diseases resulting from bacteria, fungi, and viruses and non-biotic such as: water and, temperature and other environmental factors. Detection of these diseases require people to experts in addition to a set of equipment and it is expensive in terms of time and money Therefore, the development of a computer based system helps the detection of the plants' diseases is very helpful for farmers As well as to specialists in the field of plant protection.

The proposed plant disease detection system consists of two phases, in the first phase, the knowledge base is established by introducing a set of training samples in a series of processing that include first use pre-processing techniques such as: cropping, resizing, extracting a set of color and texture features and used to great the knowledge base that used as training data for support vector machine classifier. In the second phase, use the classifier that was trained using the knowledge base for detection and diagnosis of plant leaf diseases. To create the knowledge base, 'N' number of sample images that divided it by 80% training and 20% testing are used and three crops each yield three diseases in addition to the proper state of each crop are also used.

KEYWORDS: *SVM (Support Vector Machine), Image Processing, Leaf Disease Detection Classification.*

I. INTRODUCTION

The agriculture industry is one of the most vital sectors for contribution to the national income in many countries. Throughout the years, lot of agriculture components and processes has become robotic to ensure faster production and to ensure products are of highest quality standards. Because of the high demand in the agricultural industry, it is essential that agriculture produce is cultivated using a systematic process. Diseases and defects found in plants and crops have a large impact on production in the agriculture industry, and lead to remarkable economic losses. A loss of an estimated 33 billion dollars every year was the consequence of plant pathogens found in crops in the United States. Pathogenic species affect plants significantly, initiating diseases such as chestnut blight fungus and huanglongbing citrus disease. Insect infestation along with bacterial, fungal, and viral infections are other important contribution to diseases exists in plants. Changes in climate and temperature arealso a few components that may contributeto the accelerate in diseases found in plants.

Once a plant has been infected, symptoms develop on separate segments of the plant, basically degrading the growth of the subsequent fruit or vegetable. Apple production is a very large industry remarkably in China with over 17 million tons of produce every year. Apple infections do not only of course reducegrade and yield, but can also affect thereturn bloom of the resulting season. Thesefactors have radical impact on countriesthat rely heavily on its agriculture sector as its main program of income. To overcome these losses and issues of plant diseases, farmers tend to see to chemical pesticides as a remedy solution. This solution may bepowerful in eliminating plant diseases but has harsh drawbacks. As well as being costly, the largest use of pesticides creates dangerous levels of toxic residue levels on agriculture products. This leads to burden about wholesomeness and healthiness ofproducts raised by the public when pesticides are commonly used in the produce they purchase. Therefore, the use of pesticides must be controlled, and used only when needed.

II. RELATED WORK

Various techniques are proposed by the researchers in the recent past. In these methods have own merits and demerits, which are discussed as follows. S. Rameshet al. proposed system is used to detect the healthy and diseased leaf. The datasets is created and trained under the Random Forest to classify the healthy and diseased leaves. Once the image is inserted it undergoes several processes and displays the result. The final output shows both the healthy and diseased leaf of the particular image given. It displays the 70% accuracy level.

Kumar, P. Manoj, C. M. Surya, andVarun P. Gopi. executed a system that explores the feature vectors from both frontand back side of the green leaf along with the morphological features to improve at a unique optimum combination that maximizes the identification rate. Using classifier or matching algorithm the performance analysis is executed and display the name of the medicinal plantwith its accuracy.

Ibrahim, Zaidah, Nurbaity Sabri, and NN Abu Mangshor. executed system gives the application of texture features for the leaf recognition of herbal plant identification. Then pre-processing is done on image for resizing and converting the image into gray-scale image. The features are extracted and leaf recognition is executed. Finally, it displays the name of the herbal plant with its accuracy level.

Kumar, Santhosh S., and B. K. Raghavendra. describes that they have done survey on the different plant disease and the various advance techniques to detect the plant diseases. Disease in the plants that takes place when some of the virus, bacteria infect a plant and disturb its normal growth. Farmers need the fast and efficient techniques to detect all the types of diseases of the plants that can save time. These systems that can reduce the efforts and use of pesticides.

Dhaware, Chaitali G., and K. H.Wanjale. executed system includes the techniques for the image pre-processing, image segmentation algorithm used for the automatic recognition that may be used for the leaves disease classification. Plant leaf disease identification and detectionincludes the stages such as imageacquisition, image pre-processing, imagesegmentation, feature extraction and classification. It displays the name of the disease affected by the leaf.

Barbedo, Jayme Garcia Arnal, et al. proposed system performs a deep-learning based approach to detect the tomato diseases and pests in the tomato plants using images captured in-place by using camera devices with various resolutions. Experimental results show that the proposed system can effectively recognize the nine different types of diseases and pests, with the ability to deal with the complex scenarios from a plant's surrounding area.

Based on the above study, the following challenges are identified:

- There is no separate system for classification and disease detection for herbs.
- The existing method for plant disease detection is simple naked eye observation by specialists, which allows professionals to identify and detect plant illnesses. To doso, a huge team of experts as well as continual plant monitoring are necessary, both of which are quite expensive and only possible with large farms.

III. PROBLEM DEFINITION

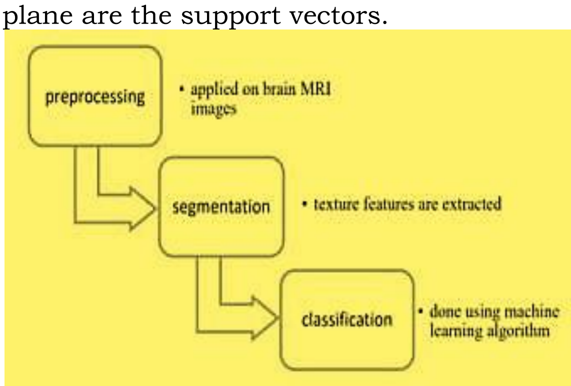
Bacterial and fungal viral infectionshave a serious impact on plant health and introduce diseases that affect growth of produce. In addition, the over reliance on fungicides and pesticides to remedy this issue, is not only costly, but has a considerably negative impact on the environment. So, there is a need to predict and target plant diseases at an early stage to aid farmers to take appropriate precautions to help preserve the defective plant. The purpose of this project is to discuss and compare the current plant disease detection techniques that feature visible imaging. An easy-to-use system to predict crop disease harshness is designed for farmers and agriculturists to find disease severity levelsof plants.

IV. PROPOSED SYSTEM

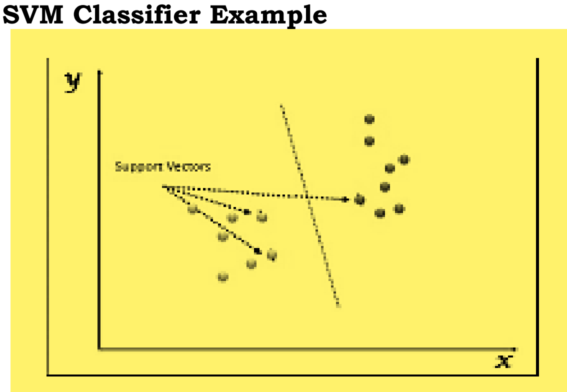
Our proposed system involves training of leaf images dataset and detection of disease from given image. Leaf images dataset is trained by SVM. It is a special technique widely used to extract unique features from no of images.

Phases of Leaf Disease Detection

SVM Classifier: SVM may be a regulated AI algorithmic program that may be utilized for every classification challenges in image processing. Nonetheless, it's essentially utilized in classification tasks. A Support Vector Machine (SVM) gives the analysis of classification by obtaining the hyper plane that maximizes the margin between the two classes. The vectors that outline the hyper plane are the support vectors.



SVM Classifier Example



Testing and Labeling the data

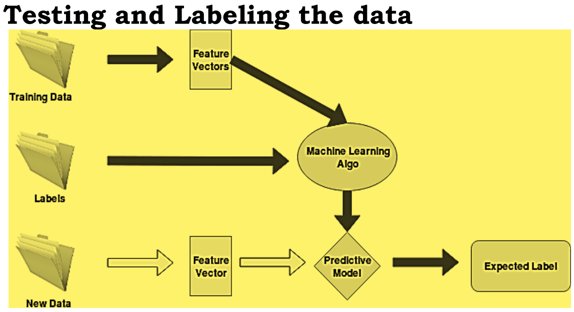


Image Acquisition

Image acquisition is the first step which requires the capturing of an image with the help of a camera or pick a leaf from local storage.

Classification

Machine learning algorithms are used for the classification of leaf images either as healthy or diseased. The major aim of ml algorithms is to automatically learn and make intelligent decisions the classification is done based on the below features:

(a) Feature processing: before classification,the extracted feature set can be refined to make it more appropriate for achieving high classification accuracies:

(b) Classifier training: pixels that are labeled as normal and cancerous are used with the extracted features to automatically learn a classification model that predicts labels based on the features;

(c) Pixel classification: the learned classification model can then be used to predict the labels for pixels with unassigned labels, based on their extracted features;

(d) Relaxation: since the learned classification model may be noisy, a relaxation of the classification results which takes into account dependencies in the labels (i.e., Classification) of neighboring pixels can be used to refine the classification predictions and yield a final segmentation.

This method requires only a small amount of training data to estimate the parameters which are needed for classification. Thetime taken for training and classification isless. This can extract useful attributes from trained weights by feeding data by levels and tune for the specific task.

V. IMAGE PROCESSING

1. Image processing is a field of computer vision that involves manipulating digital images to enhance, analyze, or extract information from them.
2. Digital images are represented as grids of pixels, where each pixel contains color orintensity information.
3. Image processing can be broadly categorized into two types: 2D image processing and 3D image processing.

ISBN: 978-81-965283-7-9

4. 2D image processing involves operations on 2D images, such as filtering, thresholding, and edge detection.
5. 3D image processing involves processing of volumetric data, such as medical CT scans or MRI images.
6. Image processing techniques can be classified as either point-based or neighborhood-based operations.
7. Point-based operations are applied to individual pixels and typically involve operations such as brightness and contrastadjustments.
8. Neighborhood-based operations are applied to a group of pixels and include operations such as blurring and sharpening.
9. Common image processing tasks include image enhancement, such as brightness/contrast adjustment, color correction, and gamma correction.
10. Image restoration, such as noise reduction, image deblurring, and image denoising using techniques like Gaussian filtering or median filtering.
11. Image segmentation, which involves dividing an image into regions based on certain criteria, such as color, intensity, ortexture.
12. Feature extraction, where relevant information is extracted from an image for further analysis or recognition, such as edgedetection, corner detection, and texture analysis.
13. Object detection, which involves identifying and locating objects within an image, commonly used in computer vision applications like object recognition and autonomous vehicles.
14. Image compression, where the size of an image is reduced while preserving important features, such as JPEG and PNGcompression.
15. Morphological operations, whichinvolve operations such as dilation and erosion, used for tasks like noise removal, image segmentation, and shape analysis.
16. Image analysis, which involves extracting meaningful information from images, such as measuring object properties (e.g., size, shape, color) or identifying patterns.
17. Image synthesis, where new images are generated based on existing images or other input data, such as image in painting, style transfer, and image generation using deep learning techniques like Generative Adversarial Networks (GANs).
18. Image recognition, which involves identifying objects or patterns within an image using machine learning or pattern recognition algorithms, such as imageclassification, object recognition, and face detection.
19. Image registration, which involves aligning and transforming images to a common coordinate system, commonly used in medical imaging and remote sensing applications.
20. Image stitching, which involves combining multiple images to create a panoramic or wide-field-of-view image, commonly used in photography and virtualreality applications.
21. Image warping, which involves geometric transformation of images, such as rotation, scaling, and shearing, commonly used in image alignment and image transformation tasks.
22. Image color correction, which involves adjusting the color balance, saturation, andhue of an image to correct for color variations due to lighting conditions or camera settings.
23. Image filtering is a common digital image processing operation that involves modifying the intensity values of pixelsbased on their neighborhood.
24. Image filtering can be used for tasks such as noise reduction, edge detection, andimage enhancement.
25. Convolution is a widely used image filtering technique that involves applying afilter mask to an image to compute new pixel values.
26. Convolutional Neural Networks (CNNs) are a popular deep learning technique used for image processing tasks such as image classification, object detection, and image segmentation.
27. Image enhancement techniques include histogram equalization, contrast stretching, and adaptive histogram equalization, which can improve the visual quality of images by enhancing their contrast and brightness.

VI. MODULE DESCRIPTION

- Data Collection and Data Visualization
- Data Splitting
- Feature Scaling and Model Training
- Model Evaluation and ModelTesting

ISBN: 978-81-965283-7-9

DATA COLLECTION

Collecting data for training the ML model is the basic step in the machine learning pipeline. The predictions made by ML systems can only be as good as the data on which they have been trained. Following are some of the problems that can arise in data collection:

- Inaccurate data. The collected data could be unrelated to the problem statement.
- Missing data. Sub-data could be missing. That could take the form of empty values in columns or missing images for some class of prediction.
- Data imbalance. Some classes or categories in the data may have a disproportionately high or low number of corresponding samples. As a result, they risk being under- represented in the model.
- Data bias. Depending on how the data, subjects and labels themselves are chosen, the model could propagate inherent biases on gender, politics, age or region, for example. Data bias is difficult to detect and remove.

DATA VISUALISATION

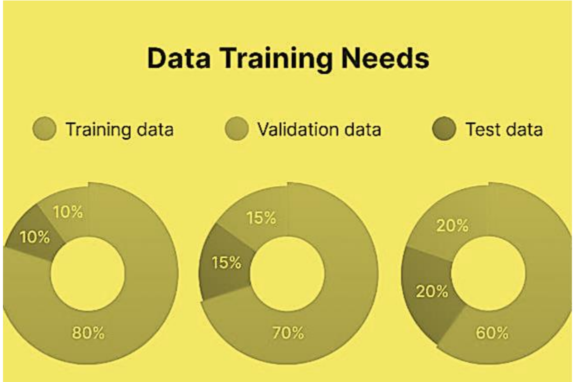
Data visualization is the graphical representation of information and data in a pictorial or graphical format (Example: charts, graphs, and maps). Data visualization tools provide an accessibleway to see and understand trends, patterns in data, and outliers. Data visualization tools and technologies are essential to analyzing massive amounts of information and making data-driven decisions. The concept of using pictures is to understand data that has been used for centuries. General types of data visualization are Charts, Tables, Graphs, Maps and Dashboards.

SPLITTING THE DATA

Train-Test Split for Evaluating Machine Learning Algorithms

The train-test split procedure is used to estimate the performance of machine learning algorithms when they are used to make predictions on data not used to train the model.

It is a fast and easy procedure to perform, the results of which allow you to compare the performance of machine learning algorithms for your predictive modeling problem. Although simple to use and interpret, there are times when the procedure should not be used, such as when you have a small dataset and situations where additional configurationis required, such as when it is used forclassification and the dataset is not balanced.



Training the Model

FEATURE SCALING

Feature scaling is a method used to normalize the range of independentvariables or features of data. In data processing, it is also known as data normalization and is generally performed during the data pre-processing step. Just togive you an example — if you have multiple independent variables like age, salary, and height; With their range as (18– 100 Years), (25,000–75,000 Euros), and(1–2 Meters) respectively, feature scaling would help them all to be in the same range, for example- cantered around 0 orin the range (0,1) depending on the scalingtechnique.

TRAINING THE MODEL

Training a machine learning (ML) model is a process in which a machine learning algorithm is fed with training data from which it can learn. ML models can betrained to benefit businesses in numerous ways, by quickly processing huge volumes of data, identifying patterns, finding anomalies or testing correlations that would be difficult for a human to do unaided.

MODEL EVALUATION

The train/test/validation split

The most important thing you can do to properly evaluate your model is tonot train the model on the entire dataset. A typical train/test split would be to use 80% of the data for training and 20% of the data for testing. The main metric used to evaluate a classification model is accuracy

Accuracy is defined as the percentage of correct predictions for the test data.

ISBN: 978-81-965283-7-9

TESTING THE MODEL

Machine Learning testing helps spot problems in models that regularevaluation metrics might miss. These problems can come from the code that enables each element of the ML system to function or from the data where outliers and an uneven distribution split, amongother things, can affect the model's performance. It can also help mitigate future issues with the model during deployment and provide a certain level of quality assurance for the model.

ML models are frequentlydeployed to production, and while evaluation metrics are used to tell the performance of the dataset, model testing focuses on checking the expected behaviorof the model. It is needed because therecan be unexpected events in production. When models are deployed, they canencounter challenges not previously seen, and testing enables your model to perform better in varying conditions.

Evaluation vs. Testing

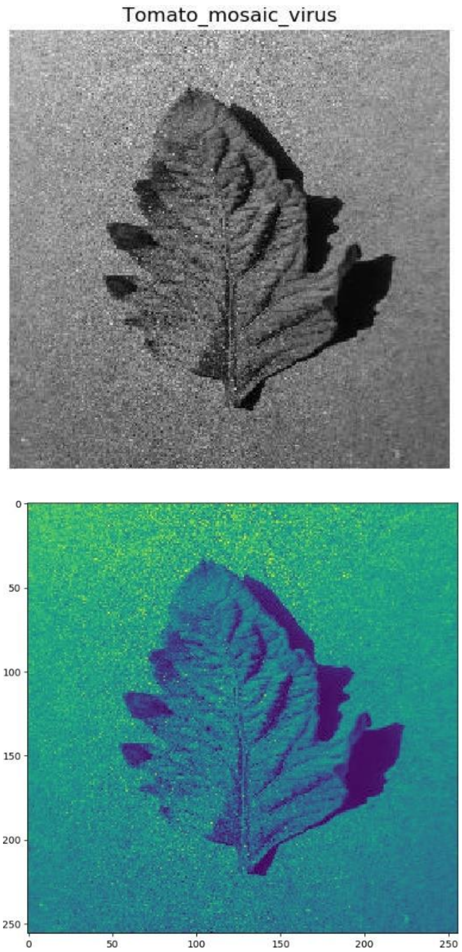
Many practitioners may rely solely on machine learning model performance evaluation metrics. Evaluation, however,is not the same as testing. It is important to know the difference.

ML model evaluation focuses onthe overall performance of the model. Such evaluations may consist of performance metrics and curves, and perhaps examples of incorrect predictions. This model evaluation is a great way to monitor your model's outcome between different versions.

VII. EXPERIMENTAL RESULTS

All the experiments are performed in Spider. For input disease leaf data, samples of plant leaves are considered. Thefirst image is original images which arefollowed by the contrast image. The enhanced contrast image which adds extra filter, brighten and gives the clear structure of the image. Followed by the diseasedimage chooses for the classification stage.

From the results it can be seen that the detection accuracy is enhanced by the SVM with proposed algorithm compared to the other approaches reported. The belowFigure , shows the segmented ROI of the respected leaf image and as result thedisease name.



Original Image

VIII. CONCLUSION ANDFUTURE WORK

The proposed method, Optimal Plant Leaf Disease Detection Using SVM Classifiers, was created to detect plant diseases at an early stage. This system saves the farmer time and money by detecting sickness for the plant. Using various methodologies and algorithms, the damaged region, healthy part of the leaf, and disease affected area will be recognized. The future works are, the disease identification will be done in real time utilizing a live camera, and it will alsoassist the Agriculture Officer in checking the quality of a crop without the need for manual monitoring. The suggested system's applicability is that data gathered from farmers would assist the government in identifying the disease that has been developed in a specific region and taking appropriate measures to prevent or manage it.

IX. REFERENCES

1. S. Ashwinkumar a, S. Rajagopal a, V. Manimaran a, B. Jegajothi, "Automated plant leaf disease detection and classification using optimal MobileNet

ISBN: 978-81-965283-7-9

based convolutional neural networks", Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2021.05.58 4, May 2021.
2. Olmos, A., et al. "High-throughput sequencing the technologies for plant pest diagnosis: opportunities and the challenges." in EPPO Bulletin 48.2 (2018):219-224.
3. Ristaino, Jean B., et al. "The persistent threat of emerging plant disease pandemics to global food security." Proceedings of the National Academy of Sciences 118.23 (2021).
4. S. Ramesh et al., "Plant Disease Detection Using Machine Learning," 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C), 2018, pp. 41-45, doi: 10.1109/ICDI3C.2018.00017.
5. Arsenovic, Marko, et al. "Solving current limitations of deep learning based approaches for plant disease detection." Symmetry 11.7 (2019): 939.
6. Ramcharan, Amanda, et al. "Deep learning for image-based cassava disease detection." Frontiers in plant science 8 (2017): 1852.
7. Kulkarni, Pranesh, et al. "Plant Disease Detection Using Image Processing and
8. Machine Learning." arXiv preprint arXiv:2106.10698 (2021).
9. Chohan, Murk, et al. "Plant disease detection using deep learning." International Journal of Recent Technologyand Engineering 9.1 (2020): 909-914.
10. Pantazi, Xanthoula Eirini, Dimitrios Moshou, and Alexandra A. Tamouridou. "Automated leaf disease detection in different crop species through image features analysis and One Class Classifiers." Computers and electronics in agriculture 156 (2019): 96-104.
11. Kumar, P. Manoj, C. M. Surya, and Varun P. Gopi. "Identification of ayurvedic medicinal plants by image processing of leaf samples." 2017 Third International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN). IEEE, 2017.
12. Bhong, Vijay S and Pawar B.V, Study and Analysis of Cotton Leaf Disease Detection Using Image Processing, International Journal of Advanced Research in Science, Engineering and Technology, 3 (2), 2016.
13. Sasirekha N, Swetha N, An Identification of Variety of Leaf Diseases Using Various Data Mining Techniques, International Journal of Advanced Research in Computer and Communication Engineering, 4 (10), 2015.
14. Peng Jiang , Yuehan Chen ,Bin Liu , Dongjian He , Chunquan Liang ,' Real-Time Detection of Apple Leaf Diseases Using Deep Learning Approach Based on Improved Convolutional Neural Networks', (Volume: 7), pp. 06 May 2019
15. Tavares JMRS, Image processing and analysis: applications and trends, AES-ATEMA'2010, Fifth international conference, Canada, June 27–July 03; 2010. pp. 27–41, ISBN: 978-0-9780479-7-9.
16. Khirade SD, Patil AB (2015) Plant Disease Detection Using Image Processing. Int Conf on Comp Communication Control and Automation IEEE 153: 768-771.

ID: 02

A REVIEW ON SDN BASED ON MOBILTY, SECURITY AND SCALABILITY MANAGEMENT

Ms.T.G.GANGA
Ph.D. Research Scholar,
Sri Krishna Arts and Science College,
Coimbatore, Tamil Nadu, India.
Assistant Professor,
Department of Computer Science,
Nilgiri College of Arts and Science, Thaloor, Tamil Nadu, India.

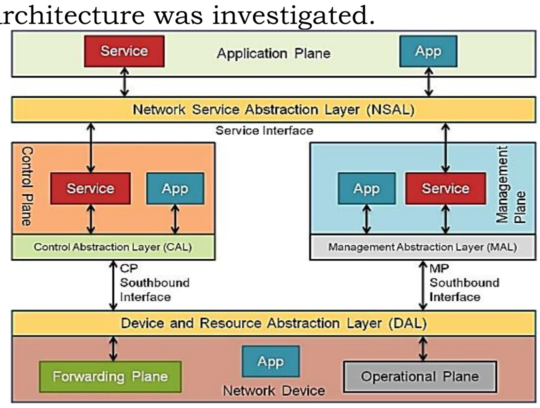
Dr.B.ANUJA BEATRICE
Associate Professor & Head,
Department of Computer Science,
Sri Krishna Arts and Science College,
Coimbatore, Tamil Nadu, India.

ABSTRACT

In the new, impending era, the Internet hosts all business applications. Nearly all corporate apps went online simultaneously during this pandemic. The COVID-19 epidemic has resulted in a massive bandwidth flood on the Internet. This condition raises concerns for security, load balance, and traffic management. The revolutionary concept of Software Defined Networking (SDN), which is part of the next-generation trend, is developing alongside other network technologies. In SDN, the control and data planes are separated, allowing for the emergence of novel network features like centralized flow management and network programmability. These features promote the introduction of new and improved network functions in an effort to enhance key aspects of network deployment, such as flexibility, scalability, network-wide visibility, and cost-effectiveness. SDN is rapidly evolving, which is transforming it into a significant enabler for implementations in diverse network scenarios, such as datacenters, ISPs, corporate, academic, and home environments. However, the technology is still far from being seen as secure and dependable, which prevents its quick adoption. To decrease the adoption gap for SDN, the scientific community has been drawn to investigate SDN security in recent years. This paper reviews the machine learning techniques that are currently used in SDN for traffic classification and traffic prediction. The recent methods are compared and its merits and demerits are discussed.

1. INTRODUCTION

The network architecture known as a legacy network has been used for a very long time. As a result, it is unable to handle problems with contemporary systems. In order to introduce programmability into the conventional network, legacy network mechanisms were investigated in the 1980s, followed by lively networks in the 1990s. According to, it is now difficult for the legacy network to manage new, powerful, and traditional systems applications effectively and efficiently. The dynamic application with central/programmable features now performs worse in the conventional network. Software-defined networking (SDN), a new dynamic and scalable technology that has emerged as a revolutionary framework for the impending current Internet, was created to address the shortcomings of the legacy networks design. In order to process, store, and distribute data and applications effectively, the introduction of cloud computing and virtualization in database-centric architecture or data-centric architecture was investigated.



SDN Layers and its Architecture

ISBN: 978-81-965283-7-9

The management and setup of internet and ICT (information-centric technology) advancements, such as mobile, cloud, social networking, big data, multimedia, and the trend towards a digital society, have become extremely complex, difficult, and time-consuming. Additionally, having high bandwidth access, being able to expand, and having dynamic control are crucial, specifically when network devices are fully integrated. Therefore, it is recommended to employ a particular group of predefined line commands and an operating system or firmware. By separating the control plane from the data plane, SDN (software-defined networking) is a system created to streamline and improve network management with high flexibility. As a result, network programmability is improved, creating more potential for innovation. Despite being a relatively recent study topic, SDN has caught the interest of many academic and industrial research institutions.

To effectively perform network optimization, data analysis, and automated network service provisioning, machine learning models are trained on historical network data. Recent advances in machine learning offer potential directions for using it in networking, according to the literature. SDN systems benefit from increased performance, effectiveness, and security thanks to machine learning. The performance, effectiveness and security, of the network can all be enhanced through machine learning-based SDN. Figure 1 shows the layers of SDN and its architecture.

II. MACHINE LEARNING ALGORITHMS FOR SDN

SDN controller provides centralized network control and management with a global view of the networks. To add intelligence to SDN, machine learning methods can be employed separately or in conjunction with the various northbound applications of the SDN controller. Machine learning-based SDN controllers analyse and automate network data as well as optimize the network. This section surveys SDN-related machine learning-based research studies. We group the studies into four categories: resource management, route optimization, quality of service (QoS) prediction, and quality of experience (QoE) prediction.

2.1 Traffic Classification

The classification of traffic is essential to network administration. Network administrators can manage and distribute various services thanks to traffic classification. Dynamic port-based methods and machine learning are currently the most popular traffic classification techniques . An OpenFlow-based SDN system is suggested for the data collecting in enterprise networks . The advantage of port-based methods, including deep packet inspection (DPI), is their high categorization accuracy. The categorization of applications based on the availability of patterns and the high processing cost to check network flow is DPI limitations. The traffic that is encrypted is invisible to DPI. As a result, machine learning-based algorithms are employed to analyze encrypted traffic at a minimal computational cost as opposed to classic DPI approaches. Massive amounts of traffic flow data must also be collected, and knowledge must be derived from the traffic data using machine learning techniques.

A two-stage approach with an effective learning cost in SDN is proposed by Xiao et al. for the identification of elephant flow. The first step in separating questionable elephant flow from mic flow is to use a head packet measuring approach. Secondly, to determine if the suspicious traffic is an elephant flow or not, a decision tree classification model is used. To identify applications based on traffic flow, application-aware traffic classification is suggested in the literature.

A behavioral engine for UDP protocol-based application-aware traffic classification was proposed by Rossi et al. UDP traffic is classified using an SVM-based model with a classification accuracy of more than 90% using its Netflow records. For traffic classification that takes into account mobile applications, TrafficVision, an SDN-enabled edge network, is suggested. Traffic Vision Engine is the primary element of Traffic Vision (TV Engine). TV Engine collects data from access devices and end devices and stores it, as well as extracts flow statistics and data for ground truth training. Different applications are classified using a decision tree classifier model. The classification of different sorts of flow, such as video content, audio files, video chats, etc., however, is done using a KNN classifier-based model.

The traffic flow is divided into QoS classes using QoS-aware traffic classification. Various applications are given

QoS classes depending on QoS criteria including jitter, latency, and loss rate. The most effective method for traffic flow classification, according to QoS classes, is traffic flow classification. For QoS aware traffic classification, a semi-supervised learning method and DPI-based approach are suggested. The widely used applications are labeled using a DPI method. To categorize apps into known and unknown QoS classes, Laplacian SVM or other semi-supervised learning-based models are trained on the labeled data from DPI.

The SDN controller is subjected to machine learning techniques for the analysis of gathered traffic data. Traffic classification methods based on machine learning, like elephant flow-aware (EF), application-aware and QoS-aware methods. Traffic flow is categorized into elephant flow and mice flow using the EF traffic classification system. Elephant flows are strong, persistent, and bandwidth-hungry, whereas mice flows are weak and delay-intolerant. In data centers, there are 80 mouse flows while the remaining traffic is an elephant flow. Elephant flow identification is crucial in such settings for effectively managing traffic flow. In paper a hybrid strategy combining a multi-classifier and a DPI-based classifier is suggested for identifying and categorizing apps.

2.3 TRAFFIC PREDICTION

Data is forwarded using several routes utilizing the dynamic routing technique depending on the circumstances or available communication lines. The Highest Throughput Minimal Cost Dynamic Routing Issue is solved by Neu Route a framework for dynamic routing for SDN that makes use of ML. NeuRoute achieves the same outcome as existing dynamic routing algorithms while taking less time to execute. NeuRoute is a dynamic, controller-independent architecture that use a neural network to learn traffic characteristics. To maximise network throughput, forwarding rules are developed according to a real-time predict traffic matrix. It is standard practise to provide more network resources than are strictly necessary, based on estimates of peak traffic loads, to achieve a particular level of QoS. Because peak demands are foreseeable, this QoS strategy is fairly straightforward, but it is not long-term economically viable. The main justification behind Neu-Route is that it is impractical to employ conventional algorithmic solutions for dynamic routing because of the significant computing cost involved. The traffic matrix predictor and the traffic routing unit are two of its primary fundamental building pieces that are built on DNN. A LSTM used in the traffic matrix predictor makes precise predictions about what will happen next. The FFN in the traffic routing unit learns how to match the routing paths to the traffic demands.

For traffic matrix forecasting, NeuTM, also put forth by Azzouni et al. uses LSTM-RNNs. In order to collect the input-output pairs needed to feed the neural networks, it uses a sliding window approach. For time-series predictions, the LSTM is a powerful self-learning algorithm that can recognize intricate non-linear patterns. The outcomes demonstrate that LSTMs outperform conventional RNNs and achieve excellent prediction accuracy in a relatively short training period.

Due to the volume of data conveyed by existing telecommunications networks, the Machine Learning Routing Computation (MLRC) module, designed by Troia et al., views it as a significant problem to offer accurate and efficient quality communications to end-users. In order to train and set up the optimizer responsible for locating the various paths in the SDN network, the authors used the ONOS controller to create a machine learning model called MLRC. A logistic regression classifier is used by MLRC because it is straightforward and comprehensible. Their findings show that the SDN network can quickly calculate a new routing configuration and execute it in response to any incoming changes in the traffic matrix. The authors also note that their findings were constrained and that more sophisticated models for efficient routing in actual networks with industrial applications would be facilitated by real datasets.

AIER, an ANN developed by Wu et al., predicts the least likelihood of congestion among all path configurations. Given the loads on all data flows and every possible path configuration, the network is educated to anticipate congestion.

Three sets of weather-related factors are merged with ML algorithms by Rahman, F. I.to better predict Traffic Flow. KNN, SVM, and ANN are three ML techniques used in this work. It can be challenging to select the optimum ML TFP model for a particular set of data, though. This study demonstrates the impact of choosing one out of three ML algorithms' main components on prediction accuracy. Weather training uses historical TF data from five months. Then, the TF for a month is predicted. In one-hour TFP, KNN performs better than SVM and ANN.

Jia, T., et al. propose a spatiotemporal neural network model based on deep learning. to thoroughly examine TF patterns in order to precisely determine citywide TF for each road segment. Recurrent convolutional networks are used to learn temporal dependences whereas densely linked convolutional networks are used to learn spatial dependences and handle spatial sparsity. By using various weights in their model, they want to combine the results of those hybrid networks. This goal is supported by outside data, such as the day of the week. China's Wuhan taxicab trajectory data was used to train and evaluate the model.

A model for predicting traffic volume is provided by Kumar, B. R., et al. and can be used in transportation planning, management, and assessment. Many TFP techniques, including historical, real-time, and time-series analysis, are put forth; nonetheless, the accuracy and effectiveness of time in forecasting are problematic and provide contradictions. An effective traffic prediction is made using real-time ANN and SVR technology. This study builds a model for traffic volume prediction along Nizampet Road in Nizampet using data from observations made in Hyderabad.

The traffic data time singularity ratio in the dropout module is defined by Liu B et al, who also suggest a combination prediction approach based on the enhanced long short-term memory neural network and the time series autoregressive integrated moving average model (SDLSTM-ARIMA), which is descended from the recurrent neural networks (RNN) model. To produce an accurate prediction of traffic flow data, it compares the temporal singularity of the traffic data with the probability value in the dropout module and then combines them at different time intervals. The experimental results show that the SDLSTM-ARIMA model-based method is more accurate than the corresponding method that solely uses autoregressive integrated moving average or autoregressive.

The SDN controller can avoid traffic congestion, enhance QoS, and proactively provision the network thanks to traffic prediction. A dynamic optimal routing meta-heuristic approach is suggested in the literature for dynamic optical routing. The three stages of these meta-heuristic algorithms are offline scheduling, online routing and offline planning. A neural network is used to forecast network traffic load for the best resource allocation during the offline scheduling phase. Online routing decisions are based on the least expensive routing path. For the purpose of path load optimization, a load balancing technique is suggested. IN order to estimate the path load using a neural network model, SDN controllers use four features: packet loss rate, transmission hop, transmission latency and bandwidth utilization ratio. For the new traffic flows, the path with the least amount of load is chosen. The research proposes the NeuTM LSTM-based framework to forecast the network traffic matrix. Real traffic data from the GEANT backbone network [28] are utilized to train the LSTM model. The performance of LSTM prediction is good for route optimization, according to results from the simulation environment. For evaluating network performance, researchers employ QoS metrics like throughput, latency, loss rate, and jitter. User perception and satisfaction levels are crucial for service providers and network operators. Utilizing user-focused measurements, QoE is utilized to evaluate the network performance. In order to deliver network services to consumers with high customer satisfaction, service providers utilize prediction algorithms to forecast QoS and QoE. For the purpose of predicting QoS and QoE, machine learning methods are used to data and statistics acquired by SDN controllers .

By forecasting QoS parameters based on key performance indicators, QoS management can be enhanced (KPIs). Because the values of the QoS metrics are continuous, predicting their values is seen as a regression task. For the prediction of QoS parameters, supervised machine learning-based algorithms such random forest, support vector regression, and ANN-based regression are utilized. Mean opinion score (MOS) is one example of a subjective indicator that QoE identifies across the network [.

ISBN: 978-81-965283-7-9

Table 1 gives the merits and demerits of certain algorithms for traffic prediction.

Author	Techniques	Objective	Merit	Demerit
Azzouni et al	**Artificial Neural Network (ANN)**	To find max throughput and minimum cost	✓ Less Execution time ✓ Min cost and Max throughput obtained	Not yet tested in large networks and datasets.
Azzouni et al	**LSTM –RNN**	Traffic Matrix prediction	✓ Best suited for sequence modeling and sequence labeling tasks	Non linear models such as ARMA, ARAR and HW not able to get high accuracy.
Trio et al	**Logistic regression**	Traffic matrix prediction and optimized routing	✓ Shortest path algorithm is improved ✓ Reduced network congestion dynamically	No real datasets are used
Wu et al	**ANN**	Predicting Minimum congestion probability	Average throughput, packet delay and loss ratio improved	Not scalable

Table 1 Comparison of Traffic prediction Algorithm

2. CONCLUSION AND FUTURE WORK

In this paper, we discussed the machine learning techniques that are currently used in SDN for traffic classification and traffic prediction. The current architecture is insufficient for satisfying the computational requirements of the features, according to research conducted on the papers on SDN. The current architecture also takes up a lot of time because it introduces numerous defects and errors, leaving no time for creating a useful architecture. Although the SDN architecture also has many difficulties, it is still far superior to the current one. Network virtualization, operation, and mobility management present difficulties. While the majority of SDN research focuses on the scalability of solutions, the control plane and data plane, and distributed versus centralized control plane, there is surprisingly little attention paid to the problems. A proper understanding of this emerging area is required if the person wants to address multiple challenges that will involve the software-defined networking process. From the review of various traffic prediction algorithms, deep learning algorithms can perform better. In our future work, we planned to propose a novel method that solves SDN issues using nature inspired optimization algorithms for better improvement.

REFERENCES

1. Kumar, S. D., Raihan, U. and Mahbubur, R. (2020).Performance Analysis of SDN-Based Intrusion Detection Model with Feature Selection Approach. International Joint Conference on Computational Intelligence, Algorithms for Intelligent.pp.483 494.
2. Ramkumar, M. P., Emil, S. and Bavani, K. (2020).Statistical Approach Based Detection of Distributed Denial of Service Attack in a Software Defined.International Conference on Advanced Computing & Communication Systems (ICACCS).No. 6.pp. 380 -385.
3. Jankowski, D. and Marek, A. (2016). On Efficiency of Selected Machine Learning Algorithms for Intrusion Detection in Software Defined Networks. International Journal of Electronics and Telecommunications. VOL. 62, No. 3, pp. 247-252. DOI: 10.1515/eletel-2016-0033.
4. Xu, G.; Mu, Y.; Liu, J. Inclusion of artificial intelligence in communication networks and services. ITU J. ICT Discov. Spec. **2017**, 1, 1–6.
5. Amaral, P.; Dinis, J.; Pinto, P.; Bernardo, L.; Tavares, J.; Mamede, H.S. Machine learning in software defined networks: Data collection and traffic classification. In Proceedings of the 2016 IEEE 24th International Conference on Network Protocols (ICNP), Singapore, 8–11 November 2016.
6. Xiao, P.; Qu,W.; Qi, H.; Xu, Y.; Li, Z. NAn efficient elephant flow detection with cost-sensitive in SDN. In Proceedings of the IEEE INISCom'15, Tokyo, Japan, 16 July 2015; pp. 24–28.
7. Wang, P.; Hao,W.; Jin, Y. Fine-grained traffic flow prediction of various vehicle types via fusison of multisource data and deep learning approaches. IEEE Trans. Intell. Transp. Syst. **2020**.

8. Uddin, M.; Nadeem, T. rafficVision: A case for pushing software defined networks to wireless edges. In Proceedings of the 2016 IEEE 13th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), Brasilia, Brazil, 10–13 October 2016.
9. Wang, P.; Lin, S.-C.; Luo, M. A framework for QoS-aware traffic classification using semi-supervised machine learning in SDNs. In Proceedings of the IEEE SCC'16, San Francisco, CA, USA, 1 September 2016; pp. 760–765.
10. Alqahtani, J.; Alanazi, S.; Hamdaoui, B. Traffic Behavior in Cloud Data Centers: A Survey. In Proceedings of the 2020 International Wireless Communications and Mobile Computing (IWCMC), Limassol, Cyprus, 27 July 2020; pp. 2106–2111.
11. Glick, M.; Rastegarfar, H. Scheduling and control in hybrid data centers. In Proceedings of the 2017 IEEE Photonics Society Summer Topical Meeting Series (SUM), San Juan, PR, USA, 21 August 2017.
12. Owusu, A.I.; Nayak, A. An Intelligent Traffic Classification in SDN-IoT: A Machine Learning Approach. In Proceedings of the 2020 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), Odessa, Ukraine, 26–29 May 2020; pp. 1–6.
13. A. Azzouni, R. Boutaba, and G. Pujolle, ``NeuRoute: Predictive dynamic routing for software-defined networks,'' in *Proc. 13th Int. Conf. Netw Service Manage. (CNSM)*, Nov. 2017, pp. 1_6.
14. A. Azzouni and G. Pujolle, ``NeuTM: A neural network-based framework for traf_c matrix prediction in SDN,'' in *Proc. IEEE/IFIP Netw. Oper.Manage. Symp. (NOMS)*, Apr. 2018, pp. 1_5.
15. S. Troia, A. Rodriguez, I. Martín, J. A. Hernández, O. G. De Dios, R. Alvizu, F. Musumeci, and G. Maier, ``Machine-learning-assisted routing in SDN-based optical networks,'' in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, 2018, pp. 1_3.
16. Y.-J. Wu, P.-C. Hwang, W.-S. Hwang, and M.-H. Cheng, ``Artificial intelligence enabled routing in software de_ned networking,'' *Appl. Sci.*,vol. 10, no. 18, p. 6564, Sep. 2020, doi: 10.3390/app10186564.
17. Rahman, F. I. (2020). Short-Term Tfp Using Machine Learning-Knn, Svm, And Ann With Weather Information. *International Journal for Traffic & Transport Engineering, 10*(3).
18. Jia, T., & Yan, P. (2020). Predicting citywide road traffic flow using deep spatiotemporal neural networks. *IEEE Transactions on Intelligent Transportation Systems, 22*(5), 3101-3111.
19. Kumar, B. R., Chikkakrishna, N. K., & Tallam, T. (2020). Short Term Predictions of Traffic Flow Characteristics using ML Techniques. *In 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA) 4*(1), 1504-1508.
20. Boyi Liu, Xiangyan Tang, Jieren Cheng, Pengchao Shi, (2018) 'Traffic Flow Combination Forecasting Method Based on Improved LSTM and ARIMA', *Int. J. Embedded Systems*, Vol. X, No. Y4, pp.000–000.
21. Azzouni, A.; Boutaba, R.; Pujolle, G. NeuRoute: Predictive dynamic routing for software-defined networks. IEEE Access. **2020**, 8, 46193–46205.
22. Jamil, F.; Iqbal, N.; Imran; Ahmad, S.; Kim, D. Peer-to-Peer Energy Trading Mechanism based on Blockchain and Machine Learning for Sustainable Electrical Power Supply in Smart Grid. IEEE Access. **2021**, 9, 39193–39217.

ID: 03

IMPROVING HEART DISEASE PREDICTION IN HEALTHCARE USING DEEP LEARNING HIERARCHY CONSTRUCTION

Ms.I.AJITHA
Assistant Professor
Department of Computer Science
Dr.SNS Rajalakshmi College of Arts & Science,
Coimbatore, Tamil Nadu, India.

Dr.A.DEVI
Assistant Professor
Department of Computer Science
Dr.SNS Rajalakshmi College of Arts & Science,
Coimbatore, Tamil Nadu, India.

ABSTRACT:
The healthcare sector is a critical area that requires focused attention due to the prevalence of various diseases in modern life. With patient expectations for the highest level of care and services, healthcare providers face the challenge of efficiently extracting knowledge from vast amounts of complex data. Artificial intelligence is a promising technique that can assist doctors in their work, particularly in predicting diseases such as heart disease. This paper examines heart disease-related issues and proposes a Deep Learning approach using Category Hierarchy Construction algorithm to extract valuable information from medical big data. The proposed hierarchical model enables automatic feature extraction and model construction for efficient disease prediction, achieving an accuracy of 95.6%.

KEYWORDS: *Prediction, Deep learning, Hierarchy Construction Algorithm, SVM and DNN.*

I. INTRODUCTION

Heart disease is a major contributor to mortality rates worldwide, affecting both men and women. It encompasses a range of conditions that impact the heart and blood vessels, leading to serious consequences for those affected. The diagnosis of heart disease often involves analyzing a large amount of clinical and pathological data, making it a complex task that requires expertise and experience. While detecting and diagnosing cardiovascular disease is an ongoing challenge, qualified professionals can use their knowledge and skills to help identify and treat this condition.

Heart disease is influenced by both modifiable and non-modifiable risk factors, such as blood pressure, cholesterol levels, age, sex, and medical history. Despite the wealth of information contained in a patient's medical records, there is a significant challenge in using this data effectively to inform clinical decisions. However, this data can be utilized to accurately predict the onset of heart disease in its early stages [2]. While invasive cardiology techniques such as angiography are used for diagnosis, they require significant technical expertise and rely on the patient's medical history and symptoms. Therefore, predicting heart disease is crucial to enable individuals to take proactive steps to safeguard their health when symptoms arise [3].

Heart disease is a complex condition that can be challenging to diagnose due to various underlying risk factors, including diabetes, high blood pressure, high cholesterol, and abnormal heart rate. To assess the severity of heart disease, researchers have employed data mining and neural network techniques. Deep learning has also been found to be effective in helping healthcare professionals analyze the vast amount of data generated by the healthcare industry to make informed decisions and predictions related to heart disease [4].

Deep learning algorithms can be used for heart disease prediction systems. Applying Deep learning is a key approach to utilize large volumes of available Heart-related data. Deep learning is of great concern when it comes to diagnosis, management and other related clinical administration aspects. Various deep learning techniques of classifiers can be used in improving prediction accuracy [5,6]. Deep learning techniques helps in identifying the data and automatically makes the predictions. In this work we are predicting the heart disease occurrence in a patient based on some important characteristics which are best suited based on our data set that we have collected. In this paper find the heart disease related issues and provides prediction of heart disease using Deep learning approach, by using the Category Hierarchy Construction algorithm using SVM and DNN. This Hierarchy Construction algorithm is used to predict this disease by plotting the train dataset. Here SVM and DNN have been widely employed to predict heart diseases, where various accuracies were obtained. Hence, only a marginal success is achieved in the creation of such predictive models for heart disease patients therefore, there is need for more complex models that incorporate multiple geographically diverse data sources to increase the accuracy of predicting the early onset of the disease.

II. LITERATURE REVIEW

Syed Arslan Ali et al [7], focus on every aspect that may have an influence on the final performance of the system, i.e., to avoid overfitting and under-fitting problems or to solve network configuration issues and optimization problems and introduce an optimally configured and improved deep belief network named OCI-DBN to solve these problems and improve the performance of the system. We used the Ruzzo-Tompa approach to remove those features that are not contributing enough to improve system performance. To find an optimal network configuration, we proposed a stacked genetic algorithm that stacks two genetic algorithms to give an optimally configured DBN. An analysis of a RBM and DBN trained is performed to give an insight how the system works. Six metrics were used to evaluate the proposed method, including accuracy, sensitivity, specificity, precision, F1 score, and Matthew's correlation coefficient.

Jian Ping Li et al [8] proposed an efficient and accurate system to diagnosis heart disease and the system is based on machine learning techniques and also proposed novel fast conditional mutual information feature selection algorithm to solve feature selection problem. The features selection algorithms are used for features selection to increase the classification accuracy and reduce the execution time of classification system. Furthermore, the leave one subject out cross-validation method has been used for learning the best practices of model assessment and for hyper-parameter tuning. The performance measuring metrics are used for assessment of the performances of the classifiers. The performances of the classifiers have been checked on the selected features as selected by features selection algorithms. The experimental results show that the proposed feature selection algorithm (FCMIM) is feasible with classifier support vector machine for designing a high-level intelligent system to identify heart disease.

Reddy, P. K., T. S. Reddy et al [9], Heart disease is a common problem which can be very severe in old ages and also in people not having a healthy lifestyle. With regular check-up and diagnosis in addition to maintaining a decent eating habit can prevent it to some extent. In this paper we have tried to implement the most sought after and important machine learning algorithm to predict the heart disease in a patient. The decision tree classifier is implemented based on the symptoms which are specifically the attributes required for the purpose of prediction. Using the decision tree algorithm, we will be able to identify those attributes which are the best one that will lead us to a better prediction of the datasets. The decision tree algorithm works in a way where it tries to solve the problem by the help of tree representation. Here each internal node of the tree represents an attribute, and each leaf node corresponds to a class label. The support vector machine algorithm helps us to classify the datasets on the basis of kernel and it also groups the dataset using hyperplane.

Khourdifi, Youness, and Mohamed Bahaj [10] prediction of heart disease is one of the areas where machine learning can be implemented. Optimization algorithms have the advantage of dealing with complex non-linear problems with a good flexibility and adaptability. In this paper, we exploited the Fast Correlation-Based Feature Selection (FCBF) method to filter redundant features in order to improve the quality of heart disease classification. Then, we perform a classification based on different classification algorithms such as K-Nearest Neighbour, Support Vector Machine, Naïve Bayes, Random Forest and a Multilayer Perception | Artificial Neural Network optimized by Particle Swarm Optimization (PSO) combined with Ant Colony Optimization (ACO) approaches. The proposed mixed approach is applied to heart disease dataset; the results demonstrate the efficacy and robustness of the proposed hybrid method in processing various types of data for heart disease classification.

Rubini P E, Deeksha G S et al [11], analyse the Heart Diseases thus an application was developed which can predict the vulnerability of heart disease, given basic symptoms like age, gender, pulse rate, resting blood pressure, cholesterol ,fasting blood sugar, resting electrocardiographic results, exercise induced angina, ST depression ST segment the slope at peak exercise, number of major vessels coloured by fluoroscopy and maximum heart rate achieved .This can be used by doctors to re heck and confirm on their patients condition. In the existing surveys they have considered only 10 features for prediction, but in this proposed research work 14 necessary features were taken into consideration.

Also, this paper presents a comparative analysis of machine learning techniques like Random Forest (RF), Logistic Regression, Support Vector Machine (SVM), and Naïve Bayes in the classification of cardiovascular disease. By the comparative analysis, machine learning algorithm Random Forest has proven to be the most accurate and reliable algorithm.

III. PROBLEM STATEMENT

Big data classification is considered as a critical and challenging problem to be addressed in big data analytics especially in health care industry. In large dataset, classification problems are rarely seen when the list of labels is infinite. In existing work, the hybrid method of Random forest and linear method is used for predicting heart disease and has recorded 88.7% accuracy only. A key issue here is overfitting, which is the central problem in the field of classification model performs effectively using a training dataset, but the model may perform less well with a test dataset. However, in most cases, training data size is often not large enough.

Therefore, overfitting can result in a high diagnostic accuracy based on the training dataset but a relatively less high diagnostic accuracy with the testing dataset when presented with novel cases [12]. Another drawback is weak correlation of data hence it damages all the work process and requires time consuming operations. It leads to more time to utilize the data.

IV. PROPOSED WORK

Prediction of heart disease using Deep Learning (DL) techniques has been an ongoing effort for the past two decades. It is hard to predict for the medical practitioners as it is a difficult task which demands expertise and higher knowledge for prediction. As previous studies suggested the common reasons behind heart disease can be unhealthy food, tobacco, excessive sugar, overweight or extra body fat [13]. Whereas the common symptoms can be pain in arms and chest.

Noticeably, these reasons are independent from each other; proper analysis on this kind of dataset can improve the process of diagnosing and can assist the heart surgeons as well. This paper addresses the issue of prediction of heart disease according to input attributes based on deep learning approaches. In this paper we proposed a Hierarchy Construction algorithm (HCA) to predict the heart disease early. HCA is used with DL to improve the performance measure by applying SVM and DNN.

SVM have attracted a great deal of attention in the last decade and actively tested to various domains applications. SVMs are mostly used for learning classification, regression or ranking function. SVM are based on statistical learning theory and structural risk minimization principal and have the intent of determining the location of decision boundaries also known as hyperplane that produce the optimal separation of classes [14]. SVM is the most robust and exact classification technique.

Deep Neural Network (DNN) have recently emerged as a powerful data-driven technique, effective at learning latent features in data. We therefore formulate heart rate estimation as a regression problem, and use for a prediction of heart disease [15]. Here developed DNN model contains classification and prediction models, which are based on a deep multilayer perceptron with linear and non-linear transfer functions, regularization and dropout, and a binary sigmoid classification using deep learning technologies, thereby creating strong and enhanced classification and prediction models.

a. Data Collection

This will involve collection of medical information from various hospitals. The dataset contains patient's information related to heart disease. The data is collected from the UCI machine learning repository. The UCI machine learning repository contains a vast and varied amount of datasets which include datasets from various domains. The data set is named Heart disease dataset. Brief discussion about dataset is explained in section V. The dataset we obtained is not completely accurate and error free. Hence, we will go to the pre-processing method.

b. Data Pre Processing

Pre-processing is applied on dataset which will remove all the unnecessary data and extract important features from data [16]. Data pre-processing is an essential step use to clean the data and make it useful for analyse the result. It also carried out to remove inconsistencies and missing values. Heart disease data is pre-processed after collection of various records. The dataset contains a total of 212 patient records. Pre-processing is done by using A Min-Max Algorithm

$$X \frac{X-X_{min}}{X_{Max}-X_{Min}} \qquad (1)$$

It is mainly used for normalization.

$$X_{Max} = \arg\max(x_i^k), \qquad X_{Min} = \arg\min(x_i^k) \quad (2)$$

After pre-processing some missing values are get reduced and the record is up to 400. Those 20 records have been removed from the dataset and the remaining 220 patient records are used in pre-processing.

c. Feature Selection

Feature selection is an important requirement before performing classification. It is a process of hauling out useful and relevant features from datasets by eliminating the redundant, noisy, and irrelevant features. The heart disease dataset has seven to ten attributes [17]. Since a large number of irrelevant and redundant attributes are involved in these expression data, the heart disease classification task is made more complex. If complete data are used to perform heart disease classification, accuracy will not be as accurate, and calculation time and costs will be high. Therefore, the feature selection, as a pre-treatment step to machine learning, reduces sizing, eliminates unresolved data, increases learning accuracy, and improves understanding of results. The recent increase in the dimensionality of the data poses a serious problem to the methods of selecting characteristics with regard to efficiency and effectiveness.

d. Hierarchy Construction Algorithm

Heart disease describes a range of condition that affects the heart. So Prediction this heart disease is considered as one of the most important subject for the data analytics. The aim of this paper is to classifies the heart disease in the best to provide optimal results. here we proposed a hierarchy construction algorithm by using SVM and DNN of deep learning approach [18, 19]. The HCA method combining SVM and DNN is introduced which is the main novel contribution of this paper. The proposed method effectively reduced the set of critical attributes. The remaining attributes are input for DNN subsequently. The heart disease datasets are used to demonstrate the efficacy of the development of the hybrid approach.

Let the training samples having dataset Data=$\{T_i, S_i\}$;i=1,2,…,n where $S_i \in$ Rn represent the i^{th} vector and $T_i \in R_n$ represent the target item. The linear SVM finds the optimal hyperplane of the form $f(x)=w^T s+b$ where w is a dimensional coefficient vector and b is an offset. This is done by solving the subsequent optimization problem.

Algorithm

Require: Input: Datasets with partition – p1, p2, p3,. . . , pn
FOR ∀apply the rules do
On the dataset R(p1,p2,p3,. . . . . . pn)
END FOR
Dataset Classify based on rules C(R(d1),R(d2) . . .R(dn))
Output: Classified datasets C(R(d1),R(d2) . . .R(dn))

High accurate results are generated using Deep neural network. The proposed system increases the classification accuracy. The dataset is divided into the testing data and training dataset. The training dataset was given to the neural network. DNN are set of algorithms that are used to recognize patterns. The layers in the DNN are made up of activation function. The training features are provided to the network through the input layer. The features are given to hidden layer where actual processing happens with the help of weighted connection. Output layer of the network is attached with the hidden layer. Generation of hypothesis through deep learning models was the aim of predictive model. Hypothesis is the relationship between data which can be tested by collecting data and making observation. We can generate the hypothesis by minimising the error in the training instances. The performance of the network is dependent on the number of guidelines used which decide the behaviour of the network. Deep neural network contains more than one hidden layer. The activation of neurons presents at the output layer.

$$f(x) = \frac{1}{1+e^{-x}} \tag{3}$$

Performance quality of the deep neural network prediction (or diagnostic) model is heavily dependent on the DNN classification model during training. In this research, final weights of the deep neural network prediction model were loaded from the deep learning training model subsystem after the training processing was completed. Generally, the DNN prediction model with (L-1) hidden layers has an output function

$$Y = \Phi_L((\cdots \Phi_3(\Phi_2(\Phi_1(XW_1 + B_1)W_2 + B_2)W_3 + B_3)\cdots)W_L + B_L) \tag{4}$$

Where input matrix data X is fed into the layer for input; Wn and Bn, n = 1, 2, ..., L, are weight matrix and bias vectors, respectively, for one of the n hidden layers; the transfer function Φn, n = 1, 2, ..., L, is either linear or nonlinear. The last layer at n = L is known as an output layer, and the other layers are hidden layers in the DNN system and architecture. As a result, the DNN prediction model in Eq. (3) can be used to detect heart disease in future patients with novel clinical data during a diagnostic process. SVM algorithm presented higher accuracy and better performance with DNN. SVM to predict patterns on large dataset in a time efficient manner.

V. EXPERIMENTAL RESULT

This section examines the efficacy of all classifiers by measuring the time required to build the model, the number of correctly classified and incorrectly classified instances, and the accuracy achieved. The classification experiment in this paper was conducted using MATLAB software.

Dataset:

The dataset used in this research was sourced from the Kaggle platform. Although the dataset contains multiple attributes, only seven to ten attributes were selected for this experiment based on suggestions from various scholars as the most useful for predicting heart disease in patients. The database file includes records of 212 patients. Notably, the dataset is open, enabling other researchers to access and analyze the data for future studies.

Evaluation Metrics

Accuracy, precision, Recall, F-measure and specificity are the various assessing methods which are used for assessing the execution of the developed predictive model. Exactness is the rate of accurately classified subject in the testing data.

Specificity = True Negative/ (True Negative +False Positive)

Accuracy = (True Positive + True Negative) / (True Positive + True Negative+ False Negative+ False Positive)

From the different classifiers we got best accuracy results. We can see that the best results are those generated by Classifiers optimized by HRFLA and GA. Comparison table is shown in table 1

Table 1: Comparison table

HCA	HRFLA	GA
Accuracy 95.6044	Accuracy 94.5055	Accuracy 90.1099
Precision 92.3077	Precision 90	Precision 85
Recall 97.2973	Recall 97.2973	Recall 91.8919
F-Measure 94.7368	F-Measure 93.5065	F-Measure 88.3117
Specificity 97.2973	Specificity 97.2973	Specificity 91.8919

Accuracy results is provided here, once the predictive model is built, we can check how efficient it is. For that, we compare the accuracy measures based on precision, recall, TP rate and FP rate values.

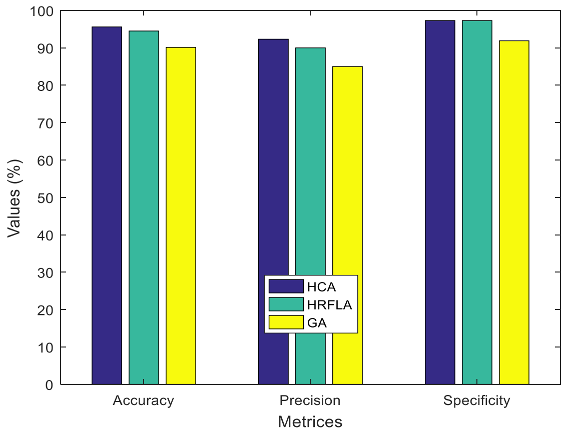


Chart 1: Comparison Matrices of Accuracy, Precision and Specificity

Accuracy in the current context would mean the percentage of instances correctly predicting from among all the available instance. Precision is defined as the percentage of corrective prediction in the positive class of the instances. To identify the significant features of heart disease, five performance metrics are used. HCA technique focuses on the best performing model compared to the existing models.

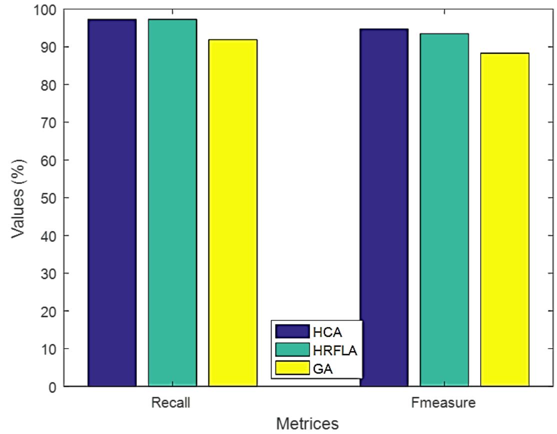


Chart 2: Comparison Matrices of Recall and F-Measure

The results are shown in chart 1 and 2 with the HCA approach. We got better performance when compare with existing Hybrid random forest with a linear model (HRFLA) and Genetic algorithm (GA).

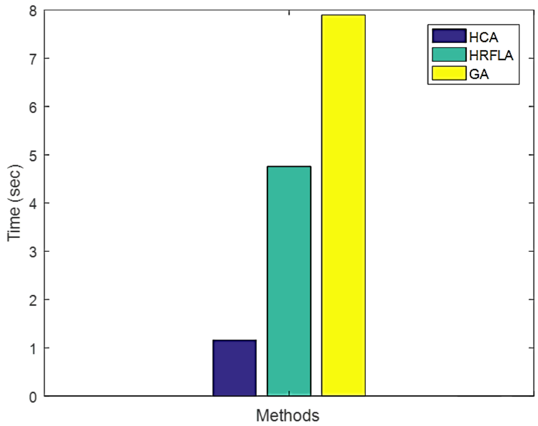


Chart 3: Time Comparison

For an implementation it requires time consuming operations. With the approach of HRFLA and GA it leads to more time to utilize the data, but oru approach requires less time which is shown in chart 3.

VI. CONCLUSION

Deep learning is a highly effective approach for reducing the risk of heart disease and assisting doctors in treating heart patients more efficiently. This paper applies deep learning techniques to predict heart disease at an early stage. SVM and deep neural network classification and prediction models were developed and evaluated based on their diagnostic performance in patients, using metrics such as Accuracy, Recall, Specificity, Precision, and F-score. The deep neural network models were built using a multilayer perceptron approach to enhance classification and prediction accuracy. These models were trained and tested using the HCA method and clinical dataset from patients, resulting in a 95.6% accuracy in predicting heart disease. Our approach significantly improved classifier performance in terms of accuracy and execution time.

VII. REFERENCES

1. Guo, Chunyan, Jiabing Zhang, Yang Liu, Yaying Xie, Zhiqiang Han, and Jianshe Yu. "Recursion Enhanced Random Forest With an Improved Linear Model (RERF-ILM) for Heart Disease Detection on the Internet of Medical Things Platform." IEEE Access 8 (2020): 59247-59256.
2. M. G. Feshki and O. S. Shijani, "Improving the heart disease diagnosis by evolutionary algorithm of PSO and feed forward neural network", Proc. Artif.

Intell. Robot. (IRANOPEN), pp. 48-53, Apr. 2016.
3. Al'Aref, Subhi J., Khalil Anchouche, Gurpreet Singh, Piotr J. Slomka, Kranthi K. Kolli, Amit Kumar, Mohit Pandey et al. "Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging." European heart journal 40, no. 24 (2019): 1975-1986.
4. H. Li, K. Ota, and M. Dong, "Learning IoT in edge: Deep learning for the Internet of Things with edge computing," IEEE Netw., vol. 32, no. 1, pp. 96–101, Jan./Feb. 2018.
5. P. Kelwade and S. S. Salankar, "Radial basis function neural network for prediction of cardiac arrhythmias based on heart rate time series," in Proc. IEEE, Jan. 2016, pp. 454–458.
6. Sowmiya and P. Sumitra, "Analytical study of heart disease diagnosis using classification techniques," in Proc. IEEE, Mar. 2017, pp. 1–5
7. Ali, S.A., Raza, B., Malik, A.K., Shahid, A.R., Faheem, M., Alquhayz, H. and Kumar, Y.J.,. An Optimally Configured and Improved Deep Belief Network (OCI-DBN) Approach for Heart Disease Prediction Based on Ruzzo–Tompa and Stacked Genetic Algorithm. IEEE Access, 2020, 8, pp.65947-65958.
8. Li, Jian Ping, Amin Ul Haq, Salah Ud Din, Jalaluddin Khan, Asif Khan, and Abdus Saboor. "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare." IEEE Access 8 (2020): 107562-107582.
9. Reddy, P. K., T. S. Reddy, S. Balakrishnan, S. M. Basha, and R. K. Poluru. "Heart disease prediction using machine learning algorithm." International Journal of Innovative Technology and Exploring Engineering 8, no. 10 (2019): 2603-2606.
10. Khourdifi, Youness, and Mohamed Bahaj. "Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization." Int. J. Intell. Eng. Syst. 12, no. 1 (2019): 242-252.
11. Rubini P E, Deeksha G S et al, "A Cardiovascular Disease Prediction using Machine Learning Algorithms", IJEAT, 2019
12. 12 Mohan, Senthilkumar, Chandrasegar Thirumalai, and Gautam Srivastava. "Effective heart disease prediction using hybrid machine learning techniques." IEEE Access 7 (2019): 81542-81554.
13. Min Chen, Yixue Hao, Kai Hwang, Lu Wang, and Lin Wang, "Disease Prediction by Machine Learning over Big Data from Healthcare Communities", IEEE transaction, 2017, pp 8869- 8879
14. Dolatabadi, S. E. Khadem, B. M. Asl, "Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM," Computer Methods and Programs in Biomedicine, Vol. 138, Issue C, pp. 117-126, 2017.
15. P. Tran and A. A. Al-Jumaily, "Non-contact Doppler radar based prediction of nocturnal body orientations using deep neural network for chronic heart failure patients," in Proc. Int. Conf. Elect. Comput. Technol. Appl. (ICECTA), Nov. 2017, pp. 1–5.
16. Uyar and A. Ilhan, "Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks," Procedia Comput. Sci., vol. 120, pp. 588–593, 2017.
17. J. Schmidhuber, "Deep learning in neural networks: An Overview." Neural Networks, Vol. 61, pp. 85-117, 2015.
18. Ramalingam, V. V., Ayantan Dandapath, and M. Karthik Raja. "Heart disease prediction using machine learning techniques: a survey." *International Journal of Engineering & Technology* 7, no. 2.8 (2018): 684-687..
19. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," Journal of Machine Learning Research, Vol. 15, pp. 1929- 1958, 2014.

ID: 04

SURVEY ON BIG DATA ANALYTICS AND ITS APPLICATIONS

Mrs.M.USHA DEVI
Assistant Professor,
Department of Computer Science,
Rathinam Arts and Science College,
Coimbatore, Tamil Nadu, India.

Ms.M.DEEPADHARSHANA
Assistant Professor,
Department of Computer Science,
Rathinam Arts and Science College,
Coimbatore, Tamil Nadu, India.

ABSTRACT
Big Data is data whose scale, diversity, and complexity require new architecture, techniques, algorithms, and analytics to manage it and extract value and hidden knowledge from it. Big data analytics uses advanced analytics on large structured and unstructured data collections to produce valuable business insights. It is used widely across industries as varied as health care, education, insurance, artificial intelligence and retail.

KEYWORDS: *Big data, Big Data Analytics, Supply chain and manufacturing, Telecom Industries.*

I. Introduction

Big Data Analytics is a collection of enormous data sets and volume of that data include large quantities of data, Management of the data, Analytics in social media and real time data. The big data analytics is the process of analysing huge volume of data. Data are measured by various terms such as terabytes and peta bytes etc. It was known as big data. After the process examining the huge quantity of data. Big data then it will involve into the big data analytics. The huge amount of data can be able to produce structured, unstructured and semi-structured from various resources.

Figure 1. Big Data

Benefits of the Technology

This technology helps organizations and companies to collect the data of their competitors. Also, it guides to perform best business moves. Moreover, this can increase the profit and will be more convenient. The available of big data and big data analytics is more. The main need of analysis of huge amount of data are in unstructured format. This type of data's is hard to analyse or process millions of records that's from social media because of these big data analytics take a huge part for these types of problems. In Generating the meaningful information in achieved by using the big data analytics and also every day retrieving the information speed is increasing. The big data can accommodate various fields that include Medical, Education, and Entertainment and so on.

II. RELATED WORKS

Big data is defined as togetherness of structured, semi structured and unstructured data and big data used to convert that into a proper information among huge amount of data. Which is useful for all sectors[1]. In systematic view all the system are used to process the large amount of data. It concludes the familiar concept of data management. Big data analytics is a next forward step of big data. In analytics part also can done some manipulation among the huge data and data sets[2]. The big data analytics is the process of making decisions faster and these decisions are based on big data. These analytics can be able to give future predictions[3].

The major component of big data analytics is dramatic improvement is business intelligence. The analytics of big data can produce the appropriate output from variety of various sources and it was known as definition for big data analytics[4].

Big data analytics is the use of advanced analytic techniques against very large, diverse data sets that include structured, semi-structured and unstructured data, from different sources, and in different sizes from terabytes to zetta bytes [5]. Analysis of big data allows analysts, researchers and business users to make better and faster decisions using data that was previously inaccessible or unusable[6]. Businesses can use advanced analytics techniques such as text analytics, machine learning, predictive analytics, data mining, statistics and natural language processing to gain new insights from previously untapped data sources independently or together with existing [7]. Big data analytics is the process of analyzing large, complex data sources to uncover trends, patterns, customer behaviors, and market preferences to inform better business decisions [8].

The complexity of analyzing big data requires various methods, including predictive analytics, machine learning, streaming analytics, and techniques like in-database and in-cluster analysis [9]. The rise of structured and unstructured data known as big data has radically transformed the function of business intelligence (BI) by converting data into action and adding value to the business. While big data analytics has increased opportunities to uncover valuable insights across the business, it has also presented new challenges in capturing, storing, and accessing information[10]. In the era of big data analytics, BI challenges have grown due to an exponential growth in the volume of data, the variety of data, and the velocity of data accumulation and change. This shift has placed significant new demands on data storage and analytics software, posing new challenges for businesses[11]. Using big data analytics, organizations can find interesting new opportunities to build business today and tomorrow. It can take data collection and data discovery to an entirely new level. Big data analytics combines data at rest (traditional structured data) with data in motion (unstructured data in the moment) to identify opportunities and take advantage of them in real time[12].

Table.1 Difference Between Big Data and Big Data Analytics

S.No	Big Data Analytics	Big Data
1	Extract the common data from minimum structured data's also	Extract the relevant insights to gain a new ideas with huge data
2	Data analytics can process the identify patterns and give the answers for the queries and those queries can compared with the historical data.	Big data is the process of collecting large amount of data to help to make an better future decision to construct strong big data platform.
3	Big Data Analytics tools and languages are Tableau public, Apache, Spark and R	Big Data tools and languages are Hive, NoSQL and Hadoop
4	Performed as Data Analyst	Big Data is applicable to handle big data professionals
5	It refers to analysing process of raw data to give some conclusion	It refers to huge quantity of data and speed of extraction.

- **Data Collection**

  Data collection is a process of gathering data for various purpose such as Research, Business, Decision Making and Inventing new strategies etc.
- **Data Storage**

Data Storage is used to keep and store the large amount of data and files and used for easy retrieving.

- **Data Consumption**

  Data Consumption deals with how much data absorption will take place during big data analytics.

- **Data cleaning**

  Arranging and abolish the incorrect data and to find and remove the null values from the huge or small quantity of data.

Figure3. Big Data Analytics

- **Data analytics**

  Data analytics can convert and analyse raw data into useful information or used for future predictions

- **Data Mining**

  Data Mining is the process of converting a useful information from raw data for improving various sources.

Figure 2. Big Data Use Cases

III. Supply chain and manufacturing

- **Better product design**

Big data analytics be as a part creating and implementing for new product by getting feedback from customer, reach of that particular product, range of product order, stock level of the product etc.

- **Predictive maintenance**

It can be one of the big data applications, it was very helpful to the manufacturers and goods prevention. The manufacturing products data are collected from sensors and resources. It helps to protect the goods. Schedule the maintenance.

- **Inventory management**

Big data analytics can play an important part in inventory management. These data analytics was providing great suggestions and strategies for managing the inventories.

IV. Telecom Industries

- **Marketing and sales**

In marketing and sales field is one of the dependent sectors for the big data analytics. It gives dramatic development in the sales field. Analysing data on usage behaviour, device specification, call pattern etc.

- **Network optimization**

The term network optimization of telecom industries is the back bone for the customer services. Here, the big data analytics take part in network traffic while customer usage, network usage capacity, it helps to reduce interruption and so on.

- **Fraud Prevention and Detection**

In telecom companies' big data analytics used to detect and prevent so many incorrect ways of usages like profile hacking, un-authorization persons.

V. INDUSTRY-SPECIFIC BIG DATA CHALLENGES

The healthcare sector has access to huge amounts of data but has been plagued by failures in utilizing the data to curb the cost of rising healthcare and by inefficient systems that stifle faster and better healthcare benefits across the board. This is mainly because electronic data is unavailable, inadequate, or unusable. Additionally, the healthcare databases that hold health-related information have made it difficult to link data that can show patterns useful in the medical field.

VI. APPLICATIONS OF BIG DATA IN THE HEALTHCARE SECTOR

Some hospitals, like Beth Israel, are using data collected from a cell phone app, from millions of patients, to allow doctors to use evidence-based medicine as opposed to administering several medical/lab tests to all patients who go to the hospital. A battery

ISBN: 978-81-965283-7-9

of tests can be efficient, but it can also be expensive and usually ineffective.

Free public health data and Google Maps have been used by the University of Florida to create visual data that allows for faster identification and efficient analysis of healthcare information, used in tracking the spread of chronic disease. Obama care has also utilized Big Data in a variety of ways. Big Data Providers in this industry include Recombinant Data, Humedica, Explorys, and Cerner.

VII. APPLICATIONS OF BIG DATA IN EDUCATION

Big data is used quite significantly in higher education. For example, the University of Tasmania. An Australian university with over 26000 students has deployed a Learning and Management System that tracks, among other things, when a student logs onto the system, how much time is spent on different pages in the system, as well as the overall progress of a student over time.

In a different use case of the use of Big Data in education, it is also used to measure teacher's effectiveness to ensure a pleasant experience for both students and teachers. Teacher's performance can be fine-tuned and measured against student numbers, subject matter, student demographics, student aspirations, behavioral classification, and several other variables. On a governmental level, the Office of Educational Technology in the U. S. Department of Education is using Big Data to develop analytics to help correct course students who are going astray while using online Big Data certification courses. Click patterns are also being used to detect boredom.

VIII. APPLICATIONS OF BIG DATA IN GOVERNMENT

In public services, Big Data has an extensive range of applications, including energy exploration, financial market analysis, fraud detection, health-related research, and environmental protection Big data is being used in the analysis of large amounts of social disability claims made to the Social Security Administration (SSA) that arrive in the form of unstructured data.

Figure.3 Applications of Big Data

The analytics are used to process medical information rapidly and efficiently for faster decision making and to detect suspicious or fraudulent claims. The Food and Drug Administration (FDA) is using Big Data to detect and study patterns of food-related illnesses and diseases. This allows for a faster response, which has led to more rapid treatment and less death. The Department of Homeland Security uses Big Data for several different use cases. Big data is analyzed from various government agencies and is used to protect the country.

IX. REFERENCES

1. Grady, Nancy W., et al. "Big data: Challenges, practices and technologies: NIST big data public working group workshop at IEEE big data 2014." 2014 IEEE International Conference on Big Data (Big Data). IEEE, 2014.
2. Sagiroglu, Seref, and Duygu Sinan. "Big data: A review." 2013 international conference on collaboration technologies and systems (CTS). IEEE, 2013.
3. Hu, Han, et al. "Toward scalable systems for big data analytics: A technology tutorial." IEEE access 2 (2014): 652-687.
4. Demchenko, Yuri, Cees De Laat, and Peter Embrey "Defining architecture components of the Big Data Ecosystem." 2014 International conference on collaboration technologies and systems (CTS). IEEE, 2014.
5. Ranjan, Rajiv, et al. "Orchestrating Bigdata analysis workflows." IEEE Cloud Computing 4.3 (2017): 20-28.
6. Yue, Peng, et al. "An SDI approach for big data analytics: The case on sensor web event detection and geoprocessing workflow." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8.10 (2015): 4720-4728.

7. Chih-Lin, I., et al. "The big-data-driven intelligent wireless network: architecture, use cases, solutions, and future trends." IEEE vehicular technology magazine 12.4 (2017): 20-29.
8. Marjani, Mohsen, et al. "Big IoT data analytics: architecture, opportunities, and open research challenges." ire access 5 (2017): 5247-5261.
9. Saranya, P., and P. Asha. "Survey on big data analytics in health care." 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE, 2019.
10. Sanghvi, Harshal A., et al. "Data science for E-Healthcare, entertainment and finance." 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA). IEEE, 2021.
11. Zahid, Hira, et al. "Big data analytics in telecommunications: literature review and architecture recommendations." IEEE/CAA Journal of AutomaticSinical 7.1 (2019): 18-38.
12. Gahi, Youssef, Mouhcine Guenon, and Hussein T. Muftah. "Big data analytics: Security and privacy challenges." 2016 IEEE Symposium on Computers and Communication (ISCC). IEEE, 2016.

ID: 05

DATA MINING FOR SECURITY APPLICATIONS

Mr.R.CHENNAPPAN
Assistant Professor,
Department of Computer Science,
Karpagam Academy of Higher Education,
Coimbatore, Tamil Nadu, India.

Ms.J.GOKULAPRIYA
Assistant Professor,
Department of Computer Science,
Rathinam College of Arts and Science (Autonomous),
Coimbatore, Tamil Nadu, India.

ABSTRACT

In this paper we tend to discuss numerous data processing techniques that we've got with success applied for cyber security. These applications embody however aren't restricted to malicious code detection by mining binary executable, network intrusion detection by mining network traffic, anomaly detection, and knowledge stream mining. We tend to summarize our achievements and current works at the University of Lone-Star State at urban centre on intrusion detection, and cyber-security analysis.

I. INTRODUCTION

Ensuring the integrity of laptop networks, each in relevance security and with relevancy the institutional lifetime of the state normally, could be a growing concern. Security and defence networks, proprietary analysis, material possession, and knowledge based mostly market mechanisms that dependupon unobstructed and artless access, will all be severely compromised by malicious intrusions. We'd like to search out the most effective thanks to defend these systems. Additionally we'd like techniques to find security breaches. Data mining has several applications in security as well as in national security (e.g., surveillance) also as in cyber security (e.g., virus detection).

The threats to national security embrace assaultive buildings and destroying vital infrastructures like power grids and telecommunication systems. Data processing techniques square measure getting used to spot suspicious people and teams, and to get that people and teams square measure capable of ending terrorist activities.

Cybersecurity cares with protective laptop and network systems from corruption because of malicious package as well as Trojan horses and viruses. Data processing is additionally being applied to produce solutions like intrusion detection and auditing. During this paper we'll focus primarily on data processing for cyber security applications.

To understand the mechanisms to be applied to safeguard the nation's computers and networks, we need to perceive the categories of threats. In [1] we tend to delineate time period threats yet as noontime period threats. A time period threat may be a threat that has got to be acted upon among a restricted time to stop some harmful scenario. Note that non time period threats will become time period threats as new info is uncovered. For instance, one may suspect that a bunch of terrorists can eventually perform some act of act of terrorism. However, if later intelligence reveals that this act can possible occur before July 1, 2008, then it becomes a time period threat and that we need to take actions straight off.

If the time bounds square measure tighter like "an attack can occur among 2 days" then we tend to cannot afford to create any mistakes in our response.

There has been a great deal of labor on applying data processing for each national security and cyber security. Abundant of the main target of our previous paper was on applying data processing for national security [1].

During this a part of the paper we'll discuss data processing for cyber security. In section two we'll discuss data processing for cyber security applications. Specially, we'll discuss threats to computers and networks and describe

ISBN: 978-81-965283-7-9

applications of information mining to observe such threats and attacks. a number of our current analysis at the University of American state at city are going to be mentioned in section three. The paper is summarized in section four.

II. DATA MINING FOR CYBER SECURITY

2.1 Overview

This section discusses data connected coercion. By data connected { terrorism | act of coercion | terrorist act | coercion } we tend to mean cyber- terrorism also as security violationsthroughaccessmanagement and alternative means that. Malicious software system like Trojan horses and viruses also are data connected security violations that we tendto cluster into data connected coercion activities. In the next few subsections we tend to discuss numerous data connected terrorist attacks. In section a pair of. 2 we tend to provide an outline of terrorist act so discuss business executive threats and external attacks. Malicious intrusions are the topic of section a pair of. 3. Master card and fraud are mentioned in section a pair of. 4. Attackson crucial infrastructures are mentioned in section a pair of.5. Data processing for cyber security is mentioned in section a pair of.6.

2.2 Cyber-terrorism, Insider Threats, and External Attacks

Cyber-terrorism is one in every of the key terrorist threats display to our nation these days. As we've mentioned earlier, this threat is exacerbated by the large quantities of knowledge currently obtainable electronically and on the net. Attacks on our computers, networks, databases and also the web infra-structure may well be devastating to businesses. It's calculable that terrorist act may cause billions of bucks to businesses.

A classic example is that of a banking system. If terrorists attack such a system and eat up accounts of funds, then the bank may lose millions and maybe billions of bucks. By incapacitating the pc system voluminous hours of productivity may well be lost, that is ultimately reminiscent of direct financial loss. Even a straightforward breakdown at run through some accident may cause many hours of productivity loss and as a result a serious loss. Thus it's important that our info systems be secure.

We have a tendency to discuss varied sorts of software engineer attacks. One is that the propagation of malicious mobile code that may harm or leak sensitive files or different data; another is intrusions upon pc networks.

Threats will occur from outside or from the within of

an organization. Outside attacks are attacks on computers from somebody outside the organization.

We have a tendency to hear of hackers breaking into pc systems and inflicting mayhem inside a corporation. Some hackers unfold viruses that harm files in varied pc systems. However an additional sinister downside is that of the business executive threat. Business executive threats are comparatively well understood within the context of non-information connected attacks, however info connected business executive threats are typically UN noted or underestimated.

Individuals within a corporation United Nations agency have studied the business' practices and procedures have a massive advantage once developing schemes to cripple the organization's info assets. These individuals may well be regular workers or maybe those acting at pc centres. The matter is kind of serious as someone could also be masquerading as some other person and inflicting every kind of injury. Within the next few sections we'll examine however data processing will be leveraged to notice and maybe stop such attacks.

2.3. Malicious Intrusions

Targets of malicious intrusions embrace networks, net purchasers and servers, databases, and operative systems. Several cyber war attacks are thanks to malicious intrusions. We tend to hear abundant concerning of net-work intrusions. What happens here is that intruders try and faucet into the networks and acquire the knowledge that's being transmitted. These intruders is also human intruders or automatic malicious software package originated by humans. Intrusions also can target files rather than network communications. For instance, associate degree offender will masquerade as a legitimate user and use their credentials to log in and access restricted files.

Intrusions also can occur on databases. During this case the taken credentials modify the offender to create queries like SQL queries and access restricted information. Essentially cyber war includes malicious intrusions also as sabotage through malicious intrusions or otherwise. Cyber security consists of security mechanisms that decide to give solutions to cyber-attacks or cyber coercion.

Once discussing malicious intrusions or cyber-attacks it's typically useful to draw analogies from the non-cyber world—that is, no data connected terrorism—and then translate those attacks to attacks on computers and networks. For instance, a stealer might enter a building through a door. Within the same method, a laptop entrant might enter the pc or network through some style of a door that has been advisedly} engineered by a malicious business executive and left unattended maybe through careless design. Another example could be a thief's use of a taken uniform to pass as a guard.

The associate degree logy here is an entrant masquerading as some-one else, lawfully coming into the system and taking all the knowledge assets. Cash within the world would translate to data assets within the cyber world. Thus, there are several parallels between non- data connected attacks and data connected attacks. We will proceed to develop counter-measures for each kinds of attacks.

2.4 Credit Card Fraud and Identity Theft

We are hearing heaps lately concerning MasterCard fraud and fraud. Within the case of MasterCard fraud, associate degree wrongdoer obtains a person's MasterCard and uses it to create unauthorized purchases. By the time the owner of the cardboard becomes tuned in to the fraud, it's going to be too late to reverse the injury or apprehend the offender. An identical downside happens with phone phone occupation cards. In truth this kind of attack is going on to Maine in person. Maybe whereas I used to be creating phone calls victimization my occupation card at airports somebody noticed the dial tones and reproduced them to create free calls. This was my company occupation card. As luck would have it our telephone service detected the matter and educated my company. The matter was prohibited at once.

A additional serious thievery is fraud. Here one assumes the identity of another person by getting key personal info like social insurance variety, and uses that info to hold out transactions underneath the opposite person's name. Even one such dealing, like marketing a house and depositing the financial gain in a very deceitful checking account, will have devastating consequences for the victim. By the time the owner finds out it'll be so much too late.

It's terribly probably that the owner could have lost countless bucks because of the fraud. We need to explore the employment of information mining each for master card fraud detection also as for fraud. There are some efforts on sleuthing master card fraud (see [2]). We want to begin operating actively on sleuthing and preventing identity thefts.

2.5 Attacks on Critical Infrastructures

Attacks on vital infrastructures may cripple a nation and its economy. Infrastructure attacks embrace assaultive the telecommunication lines, the electrical, power, gas, reservoirs and water sup-plies, food provides and alternative basic entities that area unit vital for the operation of a nation.

Attackson vital infrastructures may occur throughout any style of attack whether or not theyareaunit non- data connected, data connected or bio- terrorist act attacks. For instance, one may attack the code that runs the telecommunications business and shut down all the telecommunication lines. Similarly, code that runs the facility and gas provides can be attacked. Attacks may additionally occur through bombs and explosives. That is, the telecommunication lines can be physically attacked. Assaultive transportation lines like highways and railway tracks are attacks on infrastructures.

Infrastructures may even be attacked by natural disaster like hurricanes and earth quakes. Our main interest here is that the attacks on infrastructures through malicious attacks, each data connected and non- data connected. Our goal is to look at data processing and connected knowledge management technologies to find and stop such infrastructure attacks.

2.6 Data Mining for Cyber Security

Data mining is being applied to issues like intrusion detection and auditing. As an example, anomaly sight ion techniques may well be accustomed detect uncommon patterns and behaviors. Link analysis is also accustomed trace self-propagating malicious code to its authors. Classification is also accustomed cluster numerous cyber-attacks so use the profiles to sight Associate in nursing attack once it happens. Prediction is also accustomed confirm potential future attacks relying in an exceedingly means on info learnt concerning terrorists through email and phone conversations. Also, for a few threats non real- time data processing might answer whereas certainly different threats like for network intrusions we have a tendency to may have time period data processing.

Several researchers are investigation the utilization of information mining for intrusion detection. Whereas we'd like some kind of time period data processing, that is, the results have to be compelled to be generated in real- time, we have a tendency to additionally get to build models in time period. As an example, MasterCard fraud detection may be a kind of real- time process. However, here models are sometimes engineered previous time. Building models in time period remains a challenge. Data processing also can be used for analyzing net logs likewise as analyzing the audit trails.

Supported the results of the information mining tool, one will then confirm whether or not any unauthorized intrusions have occurred and/or whether or not any unauthorized queries are expose.

Other applications of information mining for cyber security include analyzing the audit information. One might build a repository or a warehouse containing the audit information so conduct Associate in Nursing analysis mistreatment numerous data processing tools to visualize if there ar potential anomalies. As an example, there may well be a scenario wherever an exact user cluster might access the information between three and 5am within the morning.

It may well be that this cluster is functioning the night shift during which case there is also a sound rationalization. But if this cluster is functioning between say 9am and 5pm, and this could be Associate in nursing uncommon incidence. Another example is once someone accesses the information's invariably between one and two pm; except for the last 2 days he has been accessing the database between one and 2am. This might then be flagged as Associate in nursing uncommon pattern that will want any investigation. Corporate executive threat analysis is additionally a haul each from a national security likewise from a cyber-security perspective. That is, those operating in an exceedingly corporation WHO are thought-about to be trusty might commit undercover work. Equally those with correct access to the pc system might plant Trojan horses and viruses. Catching such terrorists is much harder than catching terrorists outside of a corporation.

One may have to observe the access patterns of all the people of a company even though {they are | they're} system directors to visualize whether or not they are completing cyber war activities [3], [4].

While data processing will be accustomed sight and prevent cyber-attacks, data processing additionally exacerbates some security issues like illation and privacy. With data processing techniques one might infer sensitive associations from the legitimate responses. For a lot of details on privacy we have a tendency to check with [5], [6].

III. OUR CURRENT RESEARCH AND DEVELOPMENT

3.1 Data Mining for Intrusion and MaliciousCode Detection

We are developing variety of tools that use data processing for cyber security applications at the University of TX at city, together with tools for intrusion detection, malicious code detection, and botnet detection. Associate intrusion may be outlined as any set of actions that makes an attempt to compromise the integrity, confidentiality, or handiness of a resource. As systems become a lot of advanced, there are forever exploitable weaknesses thanks to style and programming errors, or through the employment of varied "socially engineered" penetration techniques.

Pc attacks are split into 2 classes, host-based mostly} attacks and network based attacks. Host-based attacks target a machine and take a look at to achieve access to privileged services or resources on it machine. Host- primarily based detection sometimes uses routines to get call knowledge from associate audit-process that tracks all system calls created by every user-process.

Network-based attacks create it tough for legitimate users to access numerous network services by designedly occupying or sabotaging network resources and services. This could be done by causing giant amounts of network traffic, exploiting well-known faults in networking services, overloading network hosts, etc. Network-based attack detection uses network traffic knowledge (i.e., tcpdump) to seem at traffic self-addressed to the machines being monitored. Intrusion detection systems are split into 2 groups: anomaly detection systems and misuse detection systems. Anomaly detection is that the arrange to determine malicious traffic supported deviations from established traditional network traffic patterns. Misuse detection is that the ability to spot intrusions supported a celebrated pattern for the malicious activity. These celebrated patterns are said as signatures. Anomaly detection is capable of catching new attacks. However, new legitimate behaviour also can be incorrectly known as associate attack, leading to a false positive. The main target with the present state of the art is to cut back false negative and false positive rate.

We have used multiple models like support vector machines (SVM).

But we've got improved SVM a good deal by combining it with a unique algorithm that we've got developed. We are going to describe this novel formula similarly as our approach to combining it with SVM. Additionally we are going to conjointly discuss our experimental results. For a lot of details of our analysis we have a tendency to seek advice from [7].

Our different tools embody those for email worm detection, malicious code detection, buffer overflow detection, botnet detection, and analysis of firewall policy rules. For email worm detection we have a tendency to ex-amine emails and extract options like "number of attachments" and therefore the train an information mining tools with techniques like SVM and Naïve theorem classifiers to develop a model.

Then we have a tendency to check the model to work out whether or not the e-mail incorporates a virus/worm. We have a tendency to use coaching and testing knowledge sets announce on numerous internet sites [8]. For firewall policy rule analysis we have a tendency to use association rule mining techniques to work out whether or not there are any anomalies within the policy rule set [9].

Similarly, for malicious code detection we have a tendency to extract n-gram options each with assembly code and code. We have a tendency to train the information mining tool with SVM and so check the model. The classifier then predicts whether or not the code is malicious. For buffer overflow detection we have a tendency to assume that malicious messages contain code whereas traditional messages contain knowledge. Characteristic code from knowledge is tough on several computing architectures like Windows x86 architectures attributable to variable-length instruction encodings, mixtures of code and knowledge in every section of the binary, and encrypted or compressed code segments. Whereas these obstacles have obstructed commonplace disassembly-based static analyses, we've got found success victimization SVM coaching and testing [10].

3.2 Data Mining for Botnet Detection

Our current analysis with the University of Illinois town Champaign is focusing in applying data processing techniques for botnet detection. The term "bot" comes from the word automaton. A larva is often autonomous code capable of playacting sure functions. A botnet may be a network of bots that area unit utilized by a personality's operator or bot master to hold out malicious actions. Botnets area unit one in all the foremost powerful tools utilized in cyber-crime nowadays, being capable of effecting distributed denial-of-service attacks, phishing, spamming, and eavesdropping on remote computers. Usually businesses, governments, and people face million-dollar damages caused by hackers with the assistance of botnets. It's a serious challenge to the cyber-security analysis community to combat this threat.

Botnets have totally different topologies and protocols. The foremost prevailing botnets use communications supported net Relay Chat (IRC), and have a centralized design. There are a unit several approaches accessible to find and dismantle these IRC botnets. On the opposite hand, Peer-to-Peer (P2P) networks area unit a comparatively new technology utilized in botnets. P2P botnets use localised P2P protocols to speak among the bots and also the bot master. These botnets area unit distributed, having no central purpose of failure. As a result, these botnets area unit harder to find and destroy than the IRC botnets. Moreover, most of the present analysis associated with P2P botnets area unit within the analysis p The most goal of our project is to plan associate degree economical technique to find P2P botnets. We tend to approach this drawback from a knowledge mining perspective. We tend to area unit developing techniques to mine net-work traffic for sleuthing P2P botnet traffic.

Our analysis on the botnet drawback follows from the vital observation that network traffic (as well as botnet traffic) may be a continuous flow of information stream. Typical data processing techniques don't seem to be directly applicable to stream information thanks to thought drift and infinite-length. We tend to propose a way that may expeditiously handle each issues. Our main focus is to adapt 3 major data processing techniques: classification, clustering, and outlier detection to handle stream information. Our preliminary study on the event of latest stream classification techniques for P2P botnet detection has encouraging results. [11]art.

IV. SUMMARY AND DIRECTIONS

This paper has mentioned data processing for security applications. We have a tendency to initial started with a discussion of knowledge mining for cyber security applications and so provided a quick summary of the tools we have a tendency to area unit developing. Data processing for national security also as for cyber security could be a terribly active analysis space. Varied data processing techniques together with link analysis and association rule mining area unit being explored to sight abnormal patterns. Thanks to data processing, users will currently create all types of correlations. This additionally raises privacy issues.

One of the area unites we have a tendency to are exploring for future analysis is active defense. Here we have a tendency to area unit work ways that to observe the adversaries. For such watching to be effective, the monitor should avoid detection by the static and dynamic analyses utilized by commonplace anti-malware packages. We have a tendency to area unit so developing techniques which will dynamically adapt to new detection ways and still monitor the antagonist. We have a tendency to area unit exploring the utilization of reconciling machine learning techniques for this purpose. Additionally, we have a tendency to area unit enhancing the techniques we've got developed to scale back false positive and false negatives. Moreover, we have a tendency to area unit exploring the pertinence of our techniques to distributed and pervasive environments.

V. REFERENCES

1. Thuraisingham, B., “Web Data Mining Technologies and Their Applications in Business Intelligence and Counter-terrorism”, *CRC Press*, FL, 2003.
2. Chan, P, et al, “Distributed Data Mining in Credit Card Fraud Detection”, *IEEE Intelligent Systems*, 14 (6), 1999.
3. Lazarevic, A., et al., “Data Mining for Computer Security Applications”, Tutorial *Proc. IEEE Data Mining Conference*, 2003.
4. Thuraisingham, B., “Managing Threats to Web Databases and Cyber Systems, Issues, Solutions and Challenges”, *Kluwer*, MA 2004 (Editors: V. Kumar et al).
5. Thuraisingham B., “Database and Applications Security”,
6. Thuraisingham B., “Data Miming, Privacy, Civil Liberties and National Security”, *SIGKDD Explorations*, 2002.
7. Dr. P.Logeswari “Extraction of Subset-Want in Data Stream using EMDMICA Algorithm “ Volume 7 Issue VI, June 2019.
8. Dr. P.Logeswari, J.Gokulapriya “A Literature Review on Data Mining Techniques “in July Volume -7 Issue -7.
9. Dr. P.Logeswari, J.Gokulapriya “Literature Survey on Big Data mining And Its Algorithmic Techniques “in July Volume -8 Issue7.
10. Dr. P.Logeswari, G.Banupriya “A Survey on Implementations Solutions for Attack Prevention Cryptography Technique’s in WSN UsingNS2” Volume 7,Issue 6 June 2021.
11. Dr. P.Logeswari, G.Banupriya “Review on Cryptography Techniques in WSN for Attack Prevention” volume 8, Issue 8.
12. Dr. P.Logeswari, J.Gokulapriya “Data Mining Approaches and Applications”.
13. Dr. P.Logeswari, G.Banupriya “ Image processing and its Application”
14. Dr. P.Logeswari, G.Banupriya “Cryptography Techniques and its analysis”
15. Dr. P.Logeswari, S.Sudha “Survey on Privacy Preserving Secure Mining”

ID: 06

INTRODUCTION TO DATA MINING TECHNIQUES: UNVEILING THE HIDDEN GEMS OF BIG DATA

Ms.B.KANNAKI DEVI
Ph.D. Research Scholar,
Department of Computer Science,
Bharathiar University,
Coimbatore, Tamil Nadu, India.

Dr.D.RAMYACHITRA
Associate Professor,
Department of Computer Science,
Bharathiar University,
Coimbatore, Tamil Nadu, India.

ABSTRACT:
Data mining techniques have become increasingly prevalent in recent years due to the explosive growth of data and the need to extract valuable insights from it. As the volume and complexity of data continue to escalate, data mining has emerged as a powerful approach to analyze and discover hidden patterns, associations, and trends within large datasets. As the volume of data continues to grow exponentially, data mining plays a critical role in transforming raw information into actionable knowledge. This chapter provides a concise overview of key data mining techniques, their applications, and the significance of data mining in various fields. [1][2][3]

KEYWORDS: *Data mining, Techniques, Data analysis, Pattern recognition, Machine learning, Big data, Classification, Clustering, Association rule mining, Regression, Predictive modeling.*

1. INTRODUCTION

The rapid advancement of technology has ushered in an era of unprecedented data generation, leading to the accumulation of vast and complex datasets. Within this sea of information lies a treasure trove of valuable insights, waiting to be unearthed. Data mining techniques offer a systematic and powerful approach to discover patterns, relationships, and knowledge hidden within these colossal datasets. In this book chapter, we embark on a journey into the realm of data mining, exploring its significance, methodologies, and applications in today's data-driven world. . Using data mining techniques in big data environments can yield valuable insights and patterns from vast and complex datasets. Data mining involves the process of discovering hidden patterns, trends, correlations, and relationships within the data that may not be immediately apparent. When applied to big data, data mining can help organizations make informed decisions, optimize processes, and gain a competitive edge. [1][4]

1.1 Significance of Data Mining

In today's digital landscape, data has become a strategic asset for organizations across various domains. However, the sheer volume and complexity of data present significant challenges in extracting meaningful knowledge. Data mining extracts valuable insights from vast big data, uncovering hidden patterns, trends, and relationships. By revealing hidden patterns and trends, data mining plays a pivotal role in shaping the future of data-driven decision-making. It empowers informed decision-making, enhances business strategies, and optimizes processes, leading to competitive advantages and innovation. [2][5]

2. METHODOLOGIES OF DATA MINING

The methodologies of data mining refer to the systematic approach and processes used to extract valuable patterns and insights from large datasets. These methodologies involve a series of steps that guide the data mining process, enabling analysts and data scientists to transform raw data into actionable knowledge. Here are the key methodologies of data mining:

I. Problem Definition: In the realm of big data, data mining addresses the challenge of extracting meaningful knowledge and patterns from massive and complex datasets. It involves developing scalable algorithms to identify correlations, trends, and anomalies, facilitating decision-making, predictive analysis, and process optimization on an unprecedented scale.[7]

II. Data Collection: Data mining requires access to relevant and high-quality data. In this step, data is collected from various sources, such as databases, data warehouses, or external data repositories. Data collection involves identifying the variables of interest and acquiring the necessary permissions and data access rights.

III. Data Preprocessing: Data preprocessing is a critical step that involves cleaning and transforming the data to ensure its quality and suitability for analysis. This includes handling missing values, outlier detection and treatment, data normalization, and feature engineering to create relevant and informative features.[8]

IV. Data Exploration: Data exploration aims to gain an initial understanding of the data and identify potential patterns and relationships. Exploratory data analysis (EDA) techniques, such as data visualization and summary statistics, help in discovering interesting trends and correlations.

V. Model Selection: Based on the problem definition and the nature of the data, appropriate data mining models and algorithms are selected. This includes choosing between classifications, clustering, regression, or association rule mining techniques, among others.

VI. Model Building: In this step, the selected data mining algorithms are applied to the preprocessed data to build models. Model building involves training the algorithms on the labeled or unlabeled data and optimizing their parameters to achieve the best performance.[1]

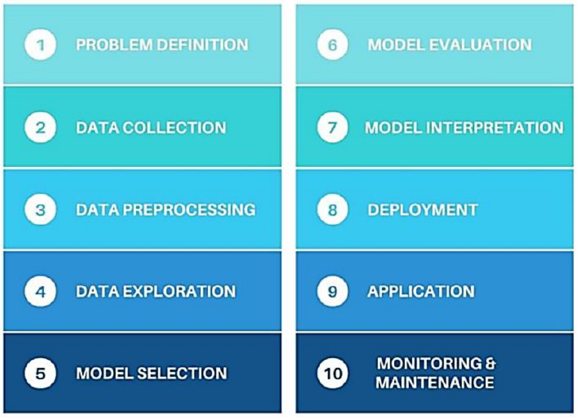


Fig.1: Data Mining Methods

Fig 1: Represents the data mining steps and methods which we follow during a research work

VII. Model Evaluation: Once the models are built, they need to be evaluated to assess their performance and generalization capability. Evaluation metrics such as accuracy, precision, recall, and F1-score are used to measure the model's effectiveness.

VIII. Model Interpretation: Model interpretation involves understanding the insights provided by the data mining models and explaining their predictions or findings in a human-interpretable manner. Interpretability is crucial in gaining trust and acceptance of the results, especially in critical applications.[2]

IX. Deployment and Application: After model evaluation and interpretation, the successful data mining models are deployed into the operational environment for practical use. The models are applied to new data to make predictions, generate insights, or support decision-making processes.

X. Monitoring and Maintenance: Data mining models may require periodic monitoring and maintenance to ensure their ongoing effectiveness. Changes in data patterns or business conditions may necessitate retraining or updating the models to maintain their relevance and accuracy.

These methodologies of data mining provide a structured framework for extracting valuable knowledge from data while ensuring the accuracy, interpretability, and practicality of the results. Data mining is a dynamic process that involves iteration and continuous improvement, allowing organizations to harness the full potential of their data for informed decision-making and enhanced business outcomes. [3]

3. DATA MINING TECHNIQUES

I. Classification: Classification is the process of categorizing data instances into predefined classes or groups based on their features or attributes. It is a supervised learning technique where the algorithm learns from labeled training data to make predictions on new, unlabeled data. Examples of classification applications include email spam filtering, sentiment analysis, and medical diagnosis.[4]

II. Clustering: Clustering is an unsupervised learning technique that group's similar data instances together based on their similarities. Unlike classification, clustering does not have predefined classes; it discovers natural groupings in the data. Clustering is widely used in market segmentation, anomaly detection, and image segmentation.[3]

III. Association Rule Mining: Association rule mining identifies interesting relationships and associations between items in large transactional databases. The goal is to discover frequently occurring patterns, such as "if A, then B" relationships. It is often used in market basket analysis, where retailers analyze customers' purchase patterns to optimize product placement and promotions.[5]

IV. Regression: Regression analysis is a data mining technique used to predict numerical values or continuous variables based on input features. It aims to find the best-fit line or function that represents the relationship between the variables. Regression is commonly used in sales forecasting, stock market analysis, and demand prediction.[6]

V. Clustering and Outlier Analysis: This technique focuses on identifying anomalies or outliers in the data, which deviate significantly from the norm. Outlier analysis is vital in fraud detection, network intrusion detection, and identifying potential errors in data.[3]

VI. Text Mining: Text mining involves extracting useful information, patterns, and relationships from unstructured text data. It uses natural language processing (NLP) techniques to analyze large volumes of text from sources such as social media, customer reviews, and news articles. Applications include sentiment analysis, topic modeling, and document categorization.[7]

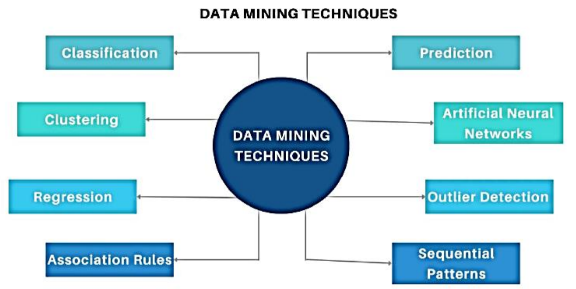


Fig: 2: Data Mining Techniques

Fig 1: Represents all the techniques which are used in the process of mining a data

VII. Time Series Analysis: Time series analysis focuses on understanding and predicting data that is collected over time. It involves techniques like moving averages, auto regression, and exponential smoothing. Time series analysis finds applications in financial forecasting, stock market analysis, and weather prediction. [8]

VIII. Feature Selection and Dimensionality Reduction: Feature selection techniques aim to identify the most relevant and significant features in a dataset, eliminating irrelevant or redundant attributes. Dimensionality reduction techniques, such as Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE), reduce the number of features while preserving the most essential information, making it easier to visualize and analyze the data.[2]

IX. Ensemble Methods: Ensemble methods combine multiple machine learning models to improve prediction accuracy and reduce overfitting. Techniques like bagging, boosting, and random forests are examples of ensemble methods commonly used in data mining.

These data mining techniques are essential tools in various industries, including finance, healthcare, marketing, and scientific research, as they enable organizations to uncover valuable insights and make data-driven decisions. [8]

4. DATA MINING APPLICATIONS

The applications of data mining techniques are far-reaching, spanning diverse industries and research fields. In finance, data mining aids in fraud detection, credit risk assessment, and algorithmic trading strategies. Healthcare professionals leverage data mining to predict disease outcomes, identify potential drug interactions, and personalize treatment plans. Marketing and retail sectors utilize data mining for customer segmentation, churn prediction, and targeted advertising campaigns. Moreover, data mining plays an essential role in scientific research, uncovering insights in areas such as genomics, climate modeling, and social network analysis.

I. Business and Marketing: Data mining plays a pivotal role in harnessing the potential of big data for business and marketing purposes. Data mining is extensively used in business and marketing to analyze customer behavior, preferences, and trends. It helps businesses identify target customer segments, personalize marketing campaigns, optimize pricing strategies, and forecast demand for products and services. [9]

II. Healthcare: Data mining has a profound impact on the healthcare industry by leveraging big data to extract valuable insights, improve patient care, optimize operations, and advance medical research. In the healthcare industry, data mining is applied to predict disease outcomes, support clinical decision-making, and identify potential drug interactions. It aids in medical diagnosis, patient risk stratification, and early detection of medical conditions. [9]

III. Finance and Banking: Data mining in the context of big data has significant applications in the finance and banking sectors, providing insights that improve risk management, customer service, fraud detection, and more. Data mining is used in the financial sector for credit risk assessment, fraud detection, and customer churn prediction. It helps financial institutions identify potentially fraudulent activities and improve customer retention strategies. [9]

IV. Retail and E-commerce: Data mining has numerous applications in the retail and e-commerce sectors, driving customer engagement, sales optimization, and operational efficiency. Retailers leverage data mining to perform market basket analysis, recommending products to customers based on their purchase history, and optimizing inventory management. [10]

V. Manufacturing and Supply Chain Management: Big data mining has significant applications in the manufacturing and supply chain management sectors, enabling companies to optimize processes, improve efficiency, and make data-driven decisions. Data mining is utilized to optimize production processes, predict equipment failures, and improve supply chain efficiency by identifying patterns and bottlenecks in the manufacturing and distribution processes. [10]

VI. Telecommunications: Big data mining has numerous applications in the telecommunications industry, enabling companies to extract insights, improve customer experiences, enhance network performance, and make data-driven decisions. Telecommunication companies use data mining to analyze call records, customer usage patterns, and network performance to enhance service quality and customer experience. [10]

VII. Social Media and Web Analytics: Big data mining techniques are employed to analyze user behavior on social media platforms and websites. It aids in understanding customer sentiments, identifying trending topics, and optimizing content recommendations.[10]

VIII. Scientific Research: Big Data mining plays a crucial role in scientific research, particularly in fields like genomics, bioinformatics, climate modeling, and astronomy. It helps scientists identify patterns and correlations in complex datasets.[11]

IX. Education: In the education sector, big data mining is used to analyze student performance data, identify at-risk students, and improve educational outcomes through personalized learning approaches.[11]

X. Crime Detection and Prevention: Big data mining plays a crucial role in crime detection and prevention by analyzing large volumes of data to identify patterns, trends, and anomalies that can help law enforcement agencies take proactive measures. Law enforcement agencies use data mining to analyze crime patterns, identify crime hotspots, and predict potential criminal activities to improve public safety. [11]

XI. Environmental Analysis: Big data mining has significant applications in environmental analysis by processing and analyzing large volumes of data to gain insights into environmental trends, natural phenomena, and sustainability. Data mining techniques aid in environmental analysis by identifying patterns and trends in climate data, natural disaster prediction, and ecological modeling.[12]

XII. Transportation and Logistics: Big data mining has transformative applications in the transportation and logistics industry, enabling organizations to enhance efficiency, optimize routes, improve supply chain management, and provide better services. Data mining is applied in transportation and logistics to optimize route planning, predict maintenance needs, and improve overall efficiency in the transportation network.[12]

5. CONCLUSION

In the contemporary landscape, the utilization of big data has evolved into a transformative force across various industries, reshaping the way businesses operate, governments make decisions, and society functions as a whole. The exponential growth of digital information has opened doors to unprecedented opportunities, and the strategic mining of this massive data reservoir holds immense potential for innovation and progress. Data mining techniques have become indispensable in today's data-driven world. They offer a systematic and effective approach to analyze vast amounts of data, transforming raw information into valuable knowledge. From classification and clustering to association rule mining and predictive modeling, data mining techniques continue to drive innovation and provide meaningful insights that enable informed decisions across various industries and research domains. As data continues to grow in complexity and size, data mining will remain a critical tool for understanding and extracting value from this wealth of information. [13]

6. REFERENCES

1. Rakesh Agrawal, Ramakrishnan Srikant. (1994) "Fast Algorithms for Mining Association Rules." Proceedings of the 20th International Conference on Very Large Data Bases (VLDB).
2. Vanjimalar, S., Ramyachitra, D. & Manikandan, P. “A review on feature selection techniques for gene expression data.” IEEE Int. Conf. Comput. Intell. Comput. Res. 18869915. https://doi.org/10.1109/ICCIC.2018.8782294 (2018).
3. R.Nithya , P.Manikandan , Dr.D.Ramyachitra, “Analysis of clustering technique for the diabetes dataset using the training set parameter” International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 9, September 2015.
4. Priya, V. & Duraisamy, Ramyachitra. (2019). Modified Genetic Algorithm (MGA) based feature selection with Mean Weighted Least Squares Twin Support Vector Machine (MW-LSTSVM) approach for vegetation classification. Cluster Computing. 22. 10.1007/s10586-018-2003-8.
5. Jiawei Han, Micheline Kamber, Jian Pei. (2011) "Data Mining: Concepts and Techniques." Morgan Kaufmann Publishers.
6. Leo Breiman. (2001) "Random Forests." Machine Learning, vol. 45, no. 1, pp. 5-32.
7. Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman. (2014) "Mining of Massive Datasets." Cambridge University Press.
8. Pedro Domingos, Geoff Hulten. (2003) "Mining High-Speed Data Streams." Proceedings of the 2003 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
9. Jiawei Han, Jian Pei, Yiwen Yin. (2000) "Mining Frequent Patterns without Candidate Generation." Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data.
10. Ron Kohavi, Foster Provost. (1998) "Glossary of Terms." Machine Learning, vol. 30, no. 2-3, pp. 271-274.
11. Andrew Kachites McCallum, Kamal Nigam. (1998) "A Comparison of Event Models for Naive Bayes Text Classification." Proceedings of the AAAI/ICML-98 Workshop on Learning for Text Categorization.
12. Jiawei Han, Micheline Kamber, Jian Pei. (2012) "Data Mining: Concepts and Techniques" (3rd Edition). Morgan Kaufmann Publishers.

13. Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth. (1996) "From Data Mining to Knowledge Discovery: An Overview." Advances in Knowledge Discovery and Data Mining, pp. 1-34.

ID: 07

INFORMATION SECURITY ALGORITHMS FOR BIOMETRIC IMAGES

Ms.S.SELVARANI
Assistant Professor,
Department of Computer Applications,
Fatima College (Autonomous),
Madurai, Tamil Nadu, India.

ABSTRACT

Biometric authentication, a technology that leverages unique physiological and behavioural traits for identity verification, has gained prominence across various domains due to its convenience and security. The rapid advancement of biometric technologies has brought about remarkable changes in the field of security and privacy. Biometric methods, which utilize unique physical and behavioural traits for identification and authentication, offer promising solutions for various applications. However, as these technologies become more prevalent, the need for robust security and privacy measures becomes increasingly apparent. This chapter aims to provide a comprehensive exploration of biometric security and privacy algorithms, addressing the challenges and solutions in this rapidly evolving domain. It also explores the usage and security considerations of biometric images in algorithmic approaches, highlighting the challenges and solutions in ensuring reliable and secure authentication.

KEYWORDS: *Biometric Security, Privacy Preservation, Liveness Detection, Fusion Techniques, Cryptographic Methods, Data Privacy, Identity Verification, Biometric Data Protection*

I. INTRODUCTION

Information security for biometric images involves the use of various algorithms and techniques to ensure the confidentiality, integrity, and authenticity of the biometric data. Biometric images, which include fingerprint scans, facial images, iris scans, and other unique physical or behavioural characteristics of individuals, have gained significant popularity in various applications due to their potential for enhanced security and convenience. They are commonly used for identity verification, access control, and authentication purposes. However, their usage also raises important considerations about privacy and security. Here are some key algorithms commonly used in biometric image security:

Encryption Algorithms

Data can be encrypted using encryption to ensure only authorized parties may decode it. On a technical level, it is the process of transforming plaintext that can be read by humans into ciphertext, which is incomprehensible text. In simpler terms, encryption changes readable data to make it seem random. A cryptographic key, which is a collection of numbers that both the sender and the recipient of an encrypted message agree upon, is necessary for encryption.

- **Symmetric Encryption:**
  Biometric picture data is encrypted using algorithms like AES (Advanced Encryption Standard) Before being sent or stored. The encryption keys need to be protected, which calls for a secure key management system. The same key is used in symmetric encryption for both encryption and decryption. The secret key is necessary to maintain the system's security. One of the most used symmetric encryption techniques is AES. AES works with blocks of data and has key lengths of 128, 192, or 256 bits. Due to its relatively small key length, the older symmetric algorithm DES (Data Encryption Standard) is today viewed as less safe. A variation known as Triple DES (3DES) applies DES three times for increased security.
- **Asymmetric Encryption:**
  To guarantee the validity of biometric data, secure key exchange and digital signatures are performed using algorithms like RSA. Asymmetric encryption employs a pair of keys—one for encryption and the other for decryption—and is also referred to as public-key cryptography. The keys are not interchangeable despite having a mathematical connection. Popular asymmetric algorithms like RSA (Rivest-

Shamir-Adleman) are used for secure key exchange, digital signatures, and encryption. The difficulty of factoring huge prime numbers contributes to its security. Digital signatures are often created using the DSA (Digital Signature Algorithm), which is frequently used in conjunction with other methods. Diffie-Hellman While not directly used for encryption, the Key Exchange technique is essential for safe key exchange in asymmetric systems.

II. HASHING ALGORITHMS:

Biometric data can be used to create fixed-size hashes using hash algorithms like SHA-256. These hashes are used to check the accuracy of the data and find any evidence of manipulation. Cryptographic operations known as hashing algorithms take an input or'message' and output a fixed-length string of characters, usually in hexadecimal format. Although hash functions are intended to be quick to calculate, their main objective is to generate a distinct output (hash value) for each distinct input. Many applications, including data integrity checking, password storage, digital signatures, and others, use hashing algorithms.

MD5 Message Digest Algorithm 5 was widely used in the past, but its security has been compromised due to vulnerabilities that allow collision attacks. It's not recommended for security-sensitive applications. SHA-1 (Secure Hash Algorithm 1) has also been found to have vulnerabilities that make it unsuitable for security purposes. Collisions can be found with less computational effort than initially thought, making it insecure. SHA-256 (Secure Hash Algorithm 256) is a member of the SHA-2 family, SHA-256 is a widely used and much more secure hashing algorithm. It produces a 256-bit hash value. SHA-256 is commonly used for data integrity verification, digital signatures, and more. SHA-3 (Secure Hash Algorithm 3) is the latest member of the Secure Hash Algorithm family, designed to provide an alternative to SHA-2. It was selected through a public competition and offers a higher level of security. HMAC (Hash-based Message Authentication Code) is a construction that uses a hash function and a secret key to produce a message authentication code. It's commonly used for verifying the authenticity and integrity of messages. It's commonly used for securely storing passwords.

Secure Key Generation and Management:

Cryptographically secure random number generators are used to generate encryption keys and other cryptographic parameters. Key management systems ensure the safe storage, distribution, and rotation of cryptographic keys. Secure key generation and management are crucial aspects of cryptography and information security. Properly generating, storing, distributing, and rotating cryptographic keys ensures the confidentiality, integrity, and authenticity of data. Cryptographic keys should be generated using a cryptographically secure random number generator (CSPRNG) to ensure unpredictability. A good source of entropy (randomness) is essential for key generation. This can include mouse movements, keyboard presses, hardware noise, and other unpredictable physical phenomena. Longer keys are generally more secure. Key length depends on the algorithm and the level of security required. Keys should be stored in physically secure environments, such as hardware security modules (HSMs) or tamper-resistant devices. Use specialized key management systems or vaults to store keys securely. These systems often have access controls, audit logs, and encryption mechanisms. Keys should be isolated from the data they protect and from other keys to prevent unauthorized access. Implement secure backup and recovery processes for keys to prevent data loss in case of hardware failures or disasters.

III. DIGITAL SIGNATURES:

Digital signatures, often used in combination with asymmetric encryption, help verify the authenticity of biometric images. Algorithms like RSA and ECDSA (Elliptic Curve Digital Signature Algorithm) are commonly used. Digital signatures are cryptographic mechanisms used to ensure the authenticity, integrity, and non-repudiation of electronic documents, messages, or transactions. They provide a way for a sender to digitally sign a document and for a recipient to verify the signature's validity. Digital signatures are widely used in various applications, such as email communication, electronic contracts, software distribution, and more.

- **Working and Benefits of Digital Signatures:**

To create a digital signature, the

sender uses their private key (signing key). This key should not be disclosed to anybody and is kept private. Recipients utilize the sender's public key, also known as the "verification key," to confirm the digital signature. It is open to distribution. The material that the sender wishes to sign and send is known as a message or document. During the signing procedure, the sender uses a cryptographic hash algorithm (such SHA-256) to calculate the message's or document's hash value. The sender then uses their private key to encrypt the hash value to produce the digital signature. The recipient then receives the signed message and the digital signature as part of the verification process. Using the same hash function, the recipient determines the received message's hash value. The recipient decrypts the digital signature using the sender's public key to get a decrypted hash value. The recipient compares the hash value for the decrypted message to the one they calculated for the received message. The digital signature is legitimate and the message has not been changed in transit if the two-hash values match.

Biometric Template Protection:

To prevent reverse engineering, biometric templates are frequently kept in hashed or altered forms. Template security is provided by algorithms like cancellable biometrics and fuzzy commitment techniques. A collection of methods and procedures known as "biometric template protection" are intended to increase the security and privacy of biometric data, particularly while it is kept or transferred. It is intended to prevent the original biometric data from being recreated or utilized inappropriately even if biometric templates are hacked.

Irreversible transformations convert the biometric template into a form that cannot be reversed back to the original data. Hashing applies a cryptographic hash function to the template, generating a fixed-size hash value. This ensures that the original template cannot be reconstructed from the hash. Random Projections project the biometric features onto random directions, creating a transformed version that's difficult to reverse engineer. Cancellable biometrics involve creating multiple representations (transforms) of the original biometric template, which can be different for each application or use. This way, even if one version is compromised, the others remain secure. Bio-Hashing generates multiple transformed versions of the biometric template using different random values. The transformations are stored instead of the original template. Fuzzy Vault combines the biometric template with a secret key to create a cryptographic vault. The vault's contents cannot be extracted without the key.

Homomorphic encryption allows scientific operations to be performed on encrypted data without decryption. This technique can be applied to biometric data to perform computations without exposing the raw data. Feature-level Protection extracts fewer or less sensitive features during enrolment, ensuring that even if compromised, the template doesn't contain all the data needed for recognition. Cryptographic binding binds the biometric template to a specific device, time, or context using cryptographic techniques. This prevents unauthorized use of the template outside the designated context. Tokenization replaces the actual biometric data with a token or reference that points to the data. This way, the sensitive data is kept separate from the reference used for identification. Zero-Knowledge Proofs prove the validity of a claim without revealing any information about the claim itself. This can be applied to biometric authentication to demonstrate knowledge of the biometric data without exposing it. Biometric template protection is a crucial aspect of maintaining privacy and security when dealing with sensitive biometric data. The specific technique or combination of techniques chosen depends on the application, threat model, and regulatory requirements.

Multi-Factor Authentication (MFA):

Users must submit two or more separate authentication factors in order to access a system or account using the security procedure known as multi-factor authentication (MFA), commonly referred to as two-factor authentication (2FA). By introducing a second layer of authentication in addition to a username and password, MFA improves security. Even if an unauthorized user is successful in obtaining one of the authentication factors, this strategy makes it substantially more difficult for them to acquire access. To increase security, combine biometric authentication with other factors such as passwords, tokens, or location-based authentication.

The user attempts to log in to a system or application by providing their username and password (the first factor). Upon successful entry of the username and password, the system prompts the user to provide an additional authentication factor. The user provides the second factor, which could be a code from a mobile app, a text message, a fingerprint scan, etc. If both factors are successfully verified, the user is granted access to the system. Even if attackers trick users into revealing their passwords, they won't be able to access the account without the second factor. MFA is often required by industry regulations and standards to ensure data security. Users can feel more confident about their accounts' security, even if they have to go through an extra step during login.

IV. LIVENESS DETECTION:

To prevent spoofing attacks, liveness detection algorithms analyse biometric images to confirm that they are being captured from a live person, not from a static image or a replica. Liveness detection is a critical component of biometric systems, particularly those involving facial recognition and other biometric modalities, to ensure that the presented biometric data comes from a live and genuine source rather than a static image or a replay attack. It helps prevent spoofing or impersonation attempts that involve presenting a photograph, video, or other non-live representations of the legitimate user. Liveness detection techniques aim to determine the "liveness" of the captured biometric sample, confirming that it originates from a real, living person and not from a fake source.

This method detects subtle movements in the biometric sample, such as blinking of eyes or slight head movements, which are difficult to replicate in static images or videos. By using 3D cameras or other depth-sensing technologies, the system can measure the depth of the face and verify that it has a natural, three-dimensional appearance. Liveness detection algorithms analyse the texture of the skin to identify pores, wrinkles, and other minute details that may be absent in fake representations. Thermal cameras can detect the difference in temperature between a live face and a photograph or mask, helping to distinguish between real and fake faces. Some systems use near-infrared light to detect blood flow beneath the skin's surface, confirming the presence of a living person. Eye Movement Detection involves tracking the movement of the user's eyes, including gaze direction and eye blinks, to verify liveness. Liveness detection is an ongoing area of research and development, as attackers continually find new ways to circumvent security measures. The effectiveness of liveness detection methods depends on the accuracy of the algorithms, the quality of the sensors, and the specific use case. Implementing robust liveness detection mechanisms is essential for maintaining the security and reliability of biometric authentication systems.

V. BIOMETRIC CRYPTOSYSTEMS:

Biometric cryptosystems combine biometric authentication with cryptographic techniques to securely manage keys or perform encryption. Biometric cryptosystems combine the fields of biometrics and cryptography. The goal is to create secure and reliable authentication systems that leverage the strengths of both biometrics and cryptography while addressing their individual limitations. Biometric cryptosystems aim to provide enhanced security, privacy, and usability compared to traditional authentication methods. Biometric data includes physiological or behavioural characteristics unique to an individual, such as fingerprints, facial features, iris patterns, voice, or typing behavior. Biometric Templates Mathematical representations of biometric data used for authentication. These templates are derived from the original biometric data but are transformed to protect the user's privacy. Cryptographic algorithms and techniques are used to secure biometric templates, perform authentication, and ensure the confidentiality and integrity of data.

Enrolment is the first step where the user's biometric data is captured and converted into a biometric template. The template is transformed using cryptographic techniques to protect the original biometric data while retaining enough information for authentication. Cryptographic methods, such as irreversible transformations or fuzzy commitment schemes, are applied to the biometric template to prevent reverse engineering and unauthorized reconstruction of the original data. Protected biometric templates are securely stored or transmitted, ensuring that the user's sensitive biometric data remains private. During authentication, the user's biometric data is captured and converted into a biometric template, which is then transformed using the same cryptographic techniques as during enrolment. The transformed template is compared with the stored template to determine if the user is authentic.

Steganography and Watermarking:

Steganography hides biometric data within other digital media, while watermarking adds invisible marks to the images for authentication and tamper detection. Steganography and watermarking are techniques used to hide information within digital media such as images, audio, and video. While they serve different purposes, both involve embedding data within other data in a way that is not immediately noticeable. Steganography is the art and science of hiding secret information within seemingly innocuous carriers (like images, audio files, or text) in such a way that only the intended recipient knows the existence of the hidden data. It aims to achieve secrecy by making the hidden information blend seamlessly with the cover medium.

Watermarking is a technique used to embed information (a watermark) directly into digital content to provide additional information about the content's origin, ownership, or authenticity. Unlike steganography, watermarks are often visible but can be difficult to remove without damaging the original content. Watermarks are often used to indicate the source of the content or the ownership of intellectual property. Watermarks can be either visible (overlaying the content) or invisible (embedded in a way that is not perceptible to the human eye). Visible watermarks can include logos, text, or patterns, while invisible watermarks are usually embedded using techniques that manipulate the content's features subtly.

Biometric Fusion:

Combining multiple biometric modalities (e.g., fingerprint and facial recognition) can enhance security and accuracy. Biometric fusion, also known as multimodal biometrics or biometric combination, involves the integration of multiple biometric modalities to enhance the accuracy, reliability, and security of authentication and identification systems. By combining information from different biometric traits, the strengths of one modality can compensate for the weaknesses of another, leading to improved overall performance. Different biometric modalities refer to unique physical or behavioural traits used for identification. Common modalities include fingerprints, facial features, iris patterns, voice, and more. By combining multiple modalities, errors and false positives/negatives from one modality can be offset by the accuracy of another, leading to more reliable identification. Fusion can enhance security by requiring attackers to bypass multiple biometric systems, making spoofing and impersonation more difficult. Users may have difficulty enrolling in one biometric modality due to physiological conditions or other factors. Fusion allows users to enrol in multiple modalities, increasing enrolment success rates.

VI. CONCLUSION:

Biometric data protection goes beyond the realm of technology. In this exploration, security measures to prevent unauthorized access to biometric databases and the importance of employing advanced encryption techniques to secure biometric templates are discussed. The vital role of liveness detection and anti-spoofing algorithms in thwarting presentation attacks and ensuring that only live individuals can access sensitive systems. the success of biometric data protection lies in collaboration among researchers, practitioners, policymakers, and users. As we unlock the potential of biometrics, we must ensure that our progress is guided by principles that prioritize security, privacy, and the empowerment of individuals in the digital age.

VII. REFERENCES:

1. Marasco, E., & Campisi, P. (2019). Advances in biometric anti-spoofing. IEEE Transactions on Information Forensics and Security, 14(5), 1176-1193.
2. Raghavendra, R., & Raja, K. B. (2020). A survey on liveness detection techniques in biometric systems. IEEE Access, 8, 176045-176062.
3. Liu, K., Kang, X., & Xie, Y. (2018). Iris image liveness detection based on convolutional neural network. IEEE Access, 6, 50688-50696.

4. Brömme, A., Busch, C., & Rathgeb, C. (Eds.). (2019). Handbook of Biometric Anti-Spoofing: Presentation Attack Detection. Springer.
5. Ratha, N. K., & Bolle, R. M. (2001). Dynamic Time Warping for Speaker Verification. IEEE Transactions on Signal Processing, 49(10), 2452-2461.
6. Rattani, A., & Nandakumar, K. (2016). Fuzzy extractor for biometric data in the cancellable domain. IET Biometrics, 6(6), 371-378.
7. Raghavendra, R., Raja, K. B., & Bhandary, S. V. (2019). Deep learning approaches for biometric liveness detection: A review. IEEE Access, 7, 24914-24931.
8. Kamruzzaman, J., & Islam, S. M. R. (2017). A novel multimodal biometric system using PCA and DWT. IEEE Access, 5, 6483-6495.
9. Sutcu, Y., & Li, Q. (2019). Privacy and Security in Biometrics. Foundations and Trends® in Signal Processing, 13(1-2), 1-224.
10. Nigam, I., & Soni, S. (2018). Biometric Cryptosystem for Image Security Using Binary Encoder. In 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT) (pp. 1190-1194).

ID: 08

ENHANCING SECURITY IN MULTIMODAL BIOMETRICS

Ms.S.JEBAPRIYA
Assistant Professor,
Department of Computer Application,
Fatima College,
Madurai, Tamil Nadu, India.

Dr.M.GANAGA DURGA
Assistant Professor & Research Supervisor,
Department of Computer Application,
Sri Meenakshi Govt. Arts College for Women,
Madurai, Tamil Nadu, India.

ABSTRACT
The demand for strong and secure authentication techniques grows as technology develops. Multimodal biometrics, which involves the integration of multiple biometric traits for identification and verification, has emerged as a promising solution to address security concerns. This chapter explores the various aspects of security in multimodal biometrics, including the challenges faced, potential vulnerabilities, and state-of-the-art solutions. By examining the fusion of different biometric modalities and cryptographic techniques, this chapter aims to provide a comprehensive understanding of how multimodal biometrics can significantly enhance security in modern authentication systems.

KEYWORDS: Multimodal biometrics, security, authentication, vulnerabilities, fusion, cryptographic techniques, anti-spoofing measures.

I. INTRODUCTION

The rise of digital transactions, remote access, and sensitive data exchange has fueled the demand for authentication systems that can confidently validate an individual's identity while thwarting malicious efforts to breach security. Multimodal biometrics has emerged as a groundbreaking solution that combines the power of multiple biometric traits to establish a robust and resilient authentication framework.

Biometrics the unique physiological and behavioral characteristics of an individual, have long been recognized as a promising avenue for identity verification. From fingerprints and iris scans to voice recognition and facial features, biometric traits offer a level of distinctiveness that sets them apart from traditional knowledge-based or token-based authentication methods. However, as biometric systems gained traction, they also revealed vulnerabilities and limitations such as susceptibility to spoofing attacks, environmental variations and privacy concerns.

Multimodal biometrics addresses these challenges by integrating two or more biometric modalities, harnessing their complementary strengths and enhancing overall system security. By requiring the simultaneous presentation of different biometric features for verification, multimodal systems significantly raise the bar for potential attackers, making it exponentially more difficult to impersonate a genuine user. This approach not only increases the system's robustness but also opens the door to innovative and complex fusion techniques that improve accuracy, reliability and resistance to attacks.

Multimodal Biometric Systems

Biometric authentication systems have rapidly evolved as a robust means of verifying an individual's identity. However, as these systems gained popularity they also exposed certain limitations and vulnerabilities that threat actors could exploit. Multimodal biometric systems have emerged as a solution to these challenges, combining the strengths of multiple biometric traits to create more secure and accurate identification and verification processes.

Multimodal biometric systems often referred to as multimodal fusion systems that integrate two or more biometric modalities to enhance overall performance, reliability and security. Each biometric modality captures unique physiological or behavioral attributes that are intrinsic to an individual. By combining these attributes, multimodal systems overcome individual modality limitations, reduce false acceptance rates and provide greater resistance to spoofing attacks.

Key Concepts of Mulltimodal Biometrics:

1. **Complementary Modalities**

   Different biometric modalities capture distinct aspects of an individual's identity. Examples include fingerprint, iris scan, facial recognition, voice recognition, hand geometry, gait analysis and behavioral traits such as signature dynamics or keystroke dynamics. Multimodal systems leverage the unique strengths of these modalities to achieve higher accuracy and security.

2. **Fusion Techniques**

   Multimodal systems employ various fusion techniques to combine the information from different modalities. These techniques can occur at different levels: score-level fusion combines matching scores from individual modalities, feature-level fusion combines extracted features and decision-level fusion combines final decisions from individual modalities.

Advantages:

1. Security is increased by requiring multiple biometric traits for authentication, multimodal systems significantly reduce the likelihood of unauthorized access due to their heightened resistance to spoofing and impersonation.
2. The combination of multiple modalities often leads to improved accuracy, especially when individual modalities might be subject to errors or variations in certain scenarios.
3. Multimodal systems are more robust to noisy data and variations caused by factors such as changes in lighting, pose or physiological conditions.
4. Users may find multimodal systems more user-friendly and natural since they can leverage traits that are easy to present such as a fingerprint or a facial scan.

Security Challenges:

1. Multimodal systems are inherently more resistant to various types of spoofing attacks, as an attacker would need to simultaneously replicate multiple biometric traits.
2. Combining multiple modalities can mitigate the effects of noise and variability inherent in individual modalities.
3. Multimodal systems can perform reliably even in challenging conditions where a single modality might fail (e.g., low-quality fingerprints).

Applications:

- Multimodal biometrics find applications in various sectors including law enforcement, border control, financial services, healthcare, access control and consumer electronics.
- They are used in scenarios where both high security and user convenience are essential.

Types of Biometric Modalities

Biometric modalities are unique physiological or behavioral characteristics that can be used for individual identification and verification. These modalities offer distinct advantages and challenges, and when combined in multimodal systems, they enhance security and accuracy.

Common types of biometric modalities:

1. **Fingerprint Recognition**: Fingerprint biometrics involves capturing the patterns of ridges and valleys on the fingertip's surface. Fingerprint recognition is widely used due to its uniqueness and ease of acquisition.
2. **Iris Recognition**: Iris recognition analyzes the unique patterns in the colored part of the eye. The complex and stable structure of the iris makes it suitable for high-security applications
3. **Facial Recognition**: Facial biometrics analyzes unique facial features, such as the distance between eyes, nose shape, and jaw line. It is convenient for user-friendly authentication and surveillance but can be susceptible to variations in lighting and pose.
4. **Voice Recognition**: Voice biometrics uses the vocal characteristics of an individual, including pitch, tone, and resonance. Voice recognition is often used for remote authentication and can

also detect changes due to factors like age or emotion

5. **Hand Geometry**: Hand geometry captures the shape and size of the hand and fingers. It is commonly used for access control and time attendance systems
6. **Retina Scans**: Retina scans focus on the unique pattern of blood vessels in the retina at the back of the eye. This modality offers high accuracy but requires specialized hardware.
7. **Vein Pattern Recognition**: Vein patterns in the hand or finger are captured using near-infrared light. Vein patterns are difficult to replicate and offer a higher level of security.
8. **DNA Biometrics**: DNA-based biometrics use an individual's genetic information for identification. While highly accurate, DNA-based biometrics may raise privacy and ethical concerns.
9. **Heartbeat Analysis**: Heartbeat-based biometrics uses the unique pattern of a person's heartbeat for identification. It is being explored for continuous authentication.
10. **Brainwave Biometrics**: Brainwave patterns, such as EEG signals, can be used for identification. This modality is still in its early stages of development.

Fusion Techniques for Enhanced Security

Fusion techniques play a crucial role in enhancing the security and accuracy of multimodal biometric systems. These techniques combine the information from different biometric modalities to create a more robust and reliable authentication process. Here are some fusion techniques commonly used for enhanced security.

1. **Score-Level Fusion**

   Score-level fusion combines the matching scores obtained from individual biometric modalities. These scores represent the similarity between the presented biometric data and the stored templates. Fusion methods at this level include:
   - Weighted Sum: Combining scores by assigning different weights to each modality based on their reliability
   - Min-Max Fusion: Selecting the minimum or maximum score among the modalities to make a decision.
   - Sum Rule: Summing up the scores of different modalities to determine the final decision.

2. **Feature-Level Fusion**

   Feature-level fusion involves combining the features extracted from different modalities before matching. This can improve accuracy and robustness by leveraging the complementary information provided by each modality. Fusion techniques at this level include:
   - Early Fusion: Combining features at an early stage of processing, such as concatenating feature vectors.
   - Late Fusion: Extracting features separately for each modality and then combining them before making a decision.

3. **Decision-Level Fusion**

   Decision-level fusion combines the final decisions made by individual biometric classifiers. It can help mitigate errors and enhance overall system performance. Fusion techniques at this level include:
   - Majority Voting: Selecting the modality with the majority of "yes" votes to make the final decision.
   - Borda Count: Assigning weights to different modalities based on their reliability and summing up the weights to make a decision.
   - Dempster-Shafer Theory: Combining evidence from different modalities to calculate a combined belief value.

4. **Multilevel Fusion**

   Multilevel fusion combines the results of fusion at different levels (e.g., score-level and feature-level) to achieve enhanced security and accuracy

Cryptographic Approaches in Multimodal Biometrics

Cryptographic approaches in multimodal biometrics play a pivotal role in enhancing security, protecting user privacy and ensuring the integrity of biometric data. These approaches leverage cryptographic techniques to safeguard biometric templates and authentication processes. Here are some cryptographic approaches commonly used in multimodal biometrics:

Biometric Template Protection

Biometric template protection techniques use cryptographic methods to transform raw biometric data into a secure, irreversible template. This template is used for authentication instead of storing the original biometric data. Common methods include:

- Cryptographic Hashing: Creating a fixed-size hash value from biometric data using one-way hash functions
- Fuzzy Commitment: Generating a commitment to the biometric data, allowing verification without revealing the actual biometric
- Cancelable Biometrics: Applying a reversible and non-invertible transformation to biometric data, enabling the generation of different templates for different applications

Homomorphic Encryption

Homomorphic encryption enables computations to be performed on encrypted data without decrypting it. It can be used to securely process biometric data in encrypted form, ensuring privacy during computation.

Secure Key Management

Cryptographic key management is crucial for ensuring the security of biometric data during storage, transmission, and authentication. Proper key management protects keys used in encryption, decryption, and authentication processes.

Privacy-Preserving Protocols

Privacy-preserving protocols ensure that biometric data remains private during interactions between different entities in the authentication process. These protocols use cryptographic methods to prevent data leakage.

CONCLUSION

Multimodal biometrics holds significant importance in enhancing security across a wide range of applications and industries. By combining the strengths of multiple biometric modalities, multimodal systems offer a holistic and robust approach to authentication and identity verification. Overall, multimodal biometrics provides a robust, versatile and future-proof solution for enhancing security. By leveraging the unique attributes of different biometric traits and employing advanced cryptographic techniques, multimodal systems offer a comprehensive defense against unauthorized access, identity fraud and other security threats.

Enhancing security in multimodal biometrics involves a multifaceted approach that combines fusion techniques and cryptographic methods. By addressing vulnerabilities, protecting user privacy and offering a versatile and robust authentication process, multimodal biometrics stands as a powerful solution to modern security challenges across various domains.

REFERENCES:

1. H. Liu, R. Wang, S. Shan, and X. Chen, “Deep supervised hashing for fast image retrieval,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.(CVPR), Jun. 2016, pp. 2064–2072.
2. Z. Chen, X. Yuan, J. Lu, Q. Tian, and J. Zhou, “Deep hashing via discrepancy minimization,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 6838–6847.
3. Nagar, K. Nandakumar, and A. K. Jain, “Multibiometric cryptosystems based on feature-level fusion,” IEEE Trans. Inf. Forensics Security, vol. 7, no. 1, pp. 255–268, Feb. 2012.
4. Ross and A. K. Jain, “Multimodal biometrics: An overview,” in Proc. 12th Eur. Signal Process. Conf. (EUSIPCO), Sep. 2004, pp. 1221–1224.
5. S. Rane, Y. Wang, S. C. Draper, and P. Ishwar, “Secure biometrics: Concepts, authentication architectures, and challenges,” IEEE Signal Process. Mag., vol. 30, no. 5, pp. 51–64, Sep. 2013.

ID: 09

BLOCKCHAIN-BASED SECURED FRAMEWORK FOR INTERNET OF MEDICAL THINGS (IOMT)

Ms.S.MARY HELAN FELISTA
Assistant Professor & Head,
Department of Computer Application,
Fatima College, Madurai, Tamil Nadu, India.
Part Time Research Scholar,
Sri Meenakshi Government Arts College for Women,
Madurai, Tamil Nadu, India.

Dr.M.GANAGA DURGA
Research Supervisor & Assistant Professor,
Department of Computer Application,
Sri Meenakshi Government Arts College for Women (A),
Madurai, Tamil Nadu, India.

ABSTRACT:
The concept of the Internet of Medical Things (IoMT), which is changing patient care and data management, was created as a result of the Internet of Things' (IoT) rapid spread in the healthcare industry. But as connectivity and data sharing have grown, security and privacy worries have taken center stage. The design and implementation of a blockchain-based secure framework specifically suited for IoMT are covered in this chapter, along with other important issues including data integrity, authenticity, and privacy. The suggested system makes use of the inherent properties of blockchain, including decentralization, immutability, and cryptographic hashing, to provide a reliable and impenetrable environment for medical data.

KEYWORDS: *Blockchain, Internet of Medical Things (IoMT), Security, Privacy, Data Integrity, Authentication, Decentralized, Smart Contracts.*

1. INTRODUCTION:

The Internet of Medical Things (IoMT), which connects medical equipment, sensors, and systems to enable real-time data collecting, analysis, and decision-making, has established a new paradigm in healthcare. To avoid unauthorized access, tampering, and breaches, however, medical data must be protected with strict security measures. The idea of a blockchain-based framework is presented in this chapter as a potential remedy to these security issues.

2. SECURITY CHALLENGES IN IOMT:

This section outlines the key security challenges associated with IoMT, including data integrity, authentication, authorization, and secure data sharing. It highlights the vulnerabilities of centralized systems in managing sensitive medical data and the potential consequences of security breaches.Security challenges in the Internet of Medical Things (IoMT) arise due to the interconnected nature of medical devices, sensors, and healthcare systems. These challenges can potentially compromise patient safety, breach sensitive medical information, and disrupt healthcare operations.

Key security challenges in IoMT:

❖ **Data Privacy and Confidentiality**: Healthcare data is highly sensitive and subject to privacy regulations. Unauthorized access to patient records can lead to identity theft, blackmail, and other malicious activities.

❖ **Data Integrity and Authenticity**: Ensuring that medical data remains accurate and untampered throughout its lifecycle is crucial. Malicious actors altering patient data could lead to incorrect diagnoses, incorrect treatments, and patient harm.

❖ **Device Vulnerabilities**: Many IoMT devices lack strong security measures due to factors such as resource limitations, legacy systems, and focus on functionality over security during development.

❖ **Network Vulnerabilities**:

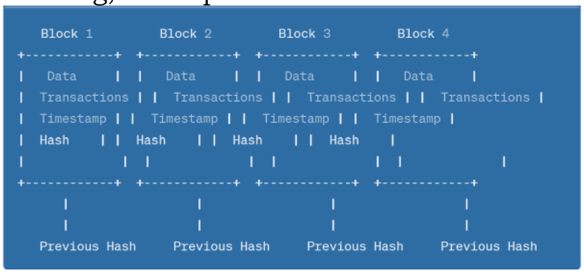
IoMT devices are often connected to hospital networks or the cloud, making them vulnerable to network-based attacks such as Distributed Denial of Service (DDoS) attacks, Man-in-the-Middle (MitM) attacks, and unauthorized access.

- **Lack of Standardization:** The absence of standardized security protocols across various IoMT devices and manufacturers can result in inconsistent security practices and difficulties in managing and securing diverse devices within a network.
- **Supply Chain Vulnerabilities**: The global nature of device manufacturing introduces the risk of compromised devices entering the supply chain, potentially containing malware or vulnerabilities that can be exploited once the devices are deployed.
- **Patch and Update Management**: Keeping IoMT devices up to date with security patches is challenging due to the wide variety of devices, potential impact on patient care, and difficulties in remotely updating devices without disrupting operations.
- **Insufficient Authentication and Authorization:** Weak authentication mechanisms and insufficient access controls can lead to unauthorized personnel gaining access to critical medical devices or patient data.
- **Data Sharing and Interoperability**: While data sharing is essential for coordinated patient care and medical research, it also presents risks if not properly secured. Interoperability challenges can result in data leaks or unauthorized data access.
- **Regulatory Compliance**: Compliance with healthcare regulations, such as HIPAA in the U.S. or GDPR (General Data Protection Regulation) in the EU, adds complexity to IoMT security, as failure to comply can result in severe penalties.
- **Blockchain Integration Challenges**: While blockchain can provide enhanced security for IoMT, integrating this technology into existing healthcare systems presents challenges, including technical compatibility and scalability.
- **Human Factors**: Human error, lack of security awareness, and social engineering attacks can all contribute to security breaches in IoMT environments.

Addressing these security challenges requires a multi-faceted approach that involves adopting best practices in device design, network security, data encryption, access controls, and ongoing security monitoring. Blockchain technology, with its inherent security features, can offer promising solutions to some of these challenges by providing a tamper-proof and transparent platform for securing medical data and device interactions.

3. BLOCKCHAIN TECHNOLOGY OVERVIEW

The core elements of blockchain technology, including blocks, transactions, consensus processes, and cryptographic hashing, are explained in detail.



In this Representation:

- Each block contains a set of data or transactions.
- The blocks are linked by including the hash of the previous block's contents (including the hash of the previous block's hash) in the current block.
- Each block has a timestamp indicating when it was added to the chain.
- The hash is a unique fingerprint generated using a cryptographic hash function, ensuring that any change to the data will result in a completely different hash.
- This structure provides data integrity and immutability since changing any data in a block would require changing all subsequent blocks' data.
- Remember that this is a simplified representation, and actual blockchain implementations may have additional complexities and features.

Overview of blockchain technology:

- **Decentralization**: Decentralization, which refers to the idea that data is not kept on a single central authority or server, is at the heart of blockchain technology.
- **Distributed Ledger:** A distributed ledger, often known as a blockchain, is

ISBN: 978-81-965283-7-9

a sort of digital record of transactions or data that is duplicated and stored among numerous computers in a network. Since each participant has a copy of the ledger, transparency is guaranteed and the possibility of data manipulation is decreased..

- **Blocks and Transactions**: A blockchain is made up of a series of interconnected blocks, each of which contains a collection of transactions. Records of data exchanges, including money transactions, data updates, and other types of information sharing, are called transactions.
- **Cryptographic Hashing**: A cryptographic hashing technique is used to combine all of the transactions in a block and process them to create a unique hash, which is a fixed-length string of characters. This hash serves as a representation of the data in the block and is utilized for security and verification.
- **Immutability:** It is very difficult to change a block's contents after it is added to the blockchain. Due to the amount of computing required and the distributed structure of the network, updating the information in a block would necessitate changing the information in every following block, which is virtually impossible.
- **Consensus Mechanisms**: Consensus mechanisms are procedures that make sure everyone on the network agrees on the blockchain's current state. Various consensus procedures, including Proof of Work (PoW), Proof of Stake (PoS), and more recently, Proof of Authority (PoA) and Practical Byzantine Fault Tolerance (PBFT), are used by different blockchains.

In general, blockchain technology is disrupting many industries and altering how people interact with digital systems by providing a new method of organizing and securing data, transactions, and procedures in a decentralized and transparent way.

4. DESIGN OF THE BLOCKCHAIN-BASED FRAMEWORK:

This section presents the architecture and design principles of the proposed framework. It explains how IoMT devices can interact with the blockchain network securely. The integration of smart contracts for enforcing access control and data sharing policies is elaborated upon. Blockchain technology is a decentralized and distributed digital ledger system that has gained significant attention due to its ability to provide security, transparency, and immutability to various applications. Originally introduced as the underlying technology for cryptocurrencies like Bitcoin, blockchain has evolved into a versatile tool with applications beyond just financial transactions. The design of a blockchain-based framework for securing the Internet of Medical Things (IoMT) involves several key components and considerations. This framework aims to address the security challenges in IoMT, such as data integrity, authentication, and privacy, using the features of blockchain technology.

Here's an outline of the design:

- **Blockchain Selection:** Choose the appropriate type of blockchain based on the use case—public, private, or consortium. Consider factors like data privacy, control, scalability, and regulatory compliance.
- **Network Architecture:** Design the network topology, including nodes, participants, and their roles. Determine the consensus mechanism that aligns with the security and scalability needs of the IoMT application.
- **Data Model:** Define the data structure to be stored on the blockchain. This includes patient records, medical device data, treatment history, and other relevant information. Ensure that sensitive data is appropriately encrypted or hashed for privacy.
- **Identity and Access Management**: Implement robust authentication mechanisms using public-private key pairs for participants and devices. Use digital certificates to ensure the legitimacy of participants and establish secure communication channels.
- **Smart Contracts**: Develop smart contracts to automate processes, enforce access control, and manage patient consent. Smart contracts can handle data access requests, verify identities, and enforce data sharing policies based on patient preferences.
- **Data Integrity and Immutability**: Implement cryptographic hashing of data before it's added to the

ISBN: 978-81-965283-7-9

blockchain. Ensure that each transaction is digitally signed to establish the authenticity of the sender.

- ❖ **Privacy-Preserving Techniques**: Incorporate techniques like off-chain storage, encryption, and zero-knowledge proofs to protect sensitive data. Consider integrating privacy-focused features like confidential transactions or ring signatures.
- ❖ **Consent Management**: Create smart contracts that allow patients to define and manage data access permissions. Implement mechanisms for patients to grant and revoke consent, with clear audit trails for accountability.
- ❖ **Integration with IoMT Devices**: Develop APIs or protocols for IoMT devices to interact securely with the blockchain network. Devices should be able to send data, receive instructions, and verify the legitimacy of other devices.
- ❖ **Security Monitoring and Auditing**: Implement continuous monitoring of the blockchain network for unusual activities. Set up mechanisms to detect and respond to security breaches. Maintain audit logs for compliance purposes.
- ❖ **Regulatory Compliance**: Ensure that the framework aligns with healthcare regulations like HIPAA, GDPR, and others applicable in your region. Design features that facilitate compliance reporting and data auditing.
- ❖ **User Interfaces:** Develop user-friendly interfaces for patients, healthcare providers, and administrators to interact with the blockchain-based system. These interfaces should allow users to manage consent, access data, and monitor their interactions with the network.
- ❖ **Testing and Deployment**: Thoroughly test the framework in a controlled environment before deploying it to a live IoMT ecosystem. Perform penetration testing and vulnerability assessments to identify and address security weaknesses.
- ❖ **Training and Adoption**: Provide training and education to users and administrators about the new framework. Ensure that all participants understand how to securely interact with the blockchain-based system.
- ❖ **Continuous Improvement**: Monitor the performance and security of the framework post-deployment. Continuously assess and address any emerging security threats or scalability challenges. Consider incorporating emerging technologies like AI for enhanced security and analytics.

By addressing data integrity, authentication, and privacy challenges through this blockchain-based framework, the security and reliability of IoMT systems can be significantly enhanced while maintaining patient privacy and data accuracy.

5. DATA INTEGRITY AND IMMUTABILITY:

This section explains how blockchain protects data integrity by creating a distinct fingerprint of medical data through cryptographic hashing.

- ❖ **Data Integrity:** Data accuracy and dependability throughout its full lifecycle are referred to as data integrity. In the context of blockchain, consensus procedures and cryptographic hashing are used to maintain data integrity.
- ❖ **Immutability:** In the context of blockchain, immutability refers to the property that once data is added to a block and the block is put to the blockchain, it cannot be changed, removed, or tampered with without the network's consent.

Benefits of Data Integrity and Immutability in Blockchain:

- o **Trust:** Participants can have trust in the reliability and timeliness of data stored on the blockchain.
- o **Security:** It becomes more challenging for bad actors to covertly modify data.
- o Transparency: The data and its history can be independently verified by each participant.
- o **Auditing:** Without relying on a centralized authority, auditors can quickly validate transactions.
- o **Non-repudiation**: Participants in a transaction or action that is recorded on the blockchain cannot dispute their involvement in it.

In industries like healthcare, where patient data must remain accurate and unaltered, blockchain's data integrity and immutability features can significantly enhance security and trust. This is especially important for medical records, treatment histories, and clinical trial data. As a result, blockchain-based systems are becoming increasingly valuable for ensuring the integrity of sensitive and critical information.

6. AUTHENTICATION AND AUTHORIZATION:

Within the framework, it is investigated how strong authentication and authorisation procedures might be implemented.

Authentication:

Verifying the identity of persons or devices trying to access the system is the process of authentication. Before allowing access to people and devices in the context of IoMT and blockchain, authentication requires establishing their legitimacy.

- Public-Private Key Pair
- Digital Signatures
- Multi-Factor Authentication (MFA)

Increase security by demanding various authentication methods. Combining a password with a special code given to a registered mobile device, for instance.

Authorization:

Based on the identification of the authenticated person or device and established permissions, authorization entails allowing or refusing access to particular resources or actions.

- Role-Based Access Control (RBAC)
- Attribute-Based Access Control (ABAC)
- Access Control Lists (ACLs)
- Patient Consent Management

7. PRIVACY AND CONSENT MANAGEMENT:

This section presents the idea of privacy-preserving transactions utilizing methods like zero-knowledge proofs to handle privacy issues. In the context of smart contracts, it is considered how patients' consent plays a part in allowing access to their medical data. Particularly in delicate industries like healthcare, privacy and consent management are essential components of data processing. It can be difficult to protect patient privacy while upholding the advantages of blockchain technology, which encourages transparency and immutability. Here is how blockchain technologies, particularly in the healthcare industry, solve privacy and consent management:

- **Privacy Challenges in Blockchain:** Protecting sensitive patient data can collide with blockchain's intrinsic transparency. All transactions on a public blockchain are available to all users, possibly disclosing private medical information. Although data visibility in private blockchains is constrained, worries regarding unwanted access and data leakage still exist. Several privacy-preserving strategies are utilized in blockchain systems to overcome these issues. Private information can be kept off-chain, with just hashes or references being kept on the blockchain. In this manner, data privacy is preserved while data integrity is upheld by the blockchain. Before being stored on the blockchain, data might be encrypted. The actual data is encrypted, and only people with the decryption keys have access to it. Ring Signatures: With ring signatures, a user can sign a transaction on behalf of a group while keeping the identity of the group member who signed it a secret.
- **Consent Management:** Management of consent is essential when working with medical data. Patients need to be able to decide who has access to their information and for what objectives. These can store the terms and circumstances of patient permission. Only after the necessary prerequisites have been completed and the patient's agreement has been verified on the blockchain, can access to patient data be allowed. Individual data access rights for various parties can be defined by patients. For instance, patients might grant access to their medical history to their primary care physician but deny it to other healthcare professionals. Patients have the right to withdraw their permission at any moment. If permission is withdrawn, smart contracts can be configured to immediately block access. Blockchain keeps track of every transaction involving consent, creating a transparent audit trail that shows who accessed the data and when.

8. FUTURE DIRECTIONS AND CONCLUSION:

Finding the ideal balance between data security, privacy, and transparency is critical in the healthcare industry. Blockchain technology can provide secure and private healthcare data management while still utilizing its advantages by combining privacy-preserving strategies, consent management, and adherence to legal norms. The chapter's discussion of prospective blockchain technology advancements for IoMT security comes to a close. Future research should focus on emerging themes including blockchain interoperability, AI integration, and hybrid consensus methods.

9. REFERENCES:

1. R. Somasundaram, Mythili Thirugnanam, "Review of security challenges in healthcare internet of things", IEEE, 2021
2. Gulraiz J. Joyia, Rao M. Liaqat, Aftab Farooq, and Saad Rehman , "Internet of Medical Things (IOMT): Applications, Benefits and Future Challenges in Healthcare Domain", Journal of Communications Vol. 12, No. 4, April 2017
3. ZIBIN ZHENG; SHAOAN XIE; HONGNING DAI; XIANGPING CHEN; HUAIMIN WANG "AN OVERVIEW OF BLOCKCHAIN TECHNOLOGY: ARCHITECTURE, CONSENSUS, AND FUTURE TRENDS", 2017 IEEE INTERNATIONAL CONGRESS ON BIG DATA (BIGDATA CONGRESS)
4. Dhiren Patel,Jay Bothra,Vasudev Patel,"Blockchain exhumed",IEEE, 2017
5. Genestier, P., Zouarhi, S., Limeux, P., Excoffier, D., Prola, A., Sandon, S., & Temerson, J.-M. (2017). Blockchain for Consent Management in the eHealth Environment: A Nugget for Privacy and Security Challenges. Journal of the International Society for Telemedicine and EHealth, 5, (GKR);e24:(1-4).
6. Premkumar Duraisamy, K Parvathy, V. Niranjani, Yuvaraj Natarajan, "Improved Recommender System for Kid's Hobby Prediction using different Machine Learning Techniques", 2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT), pp.01-06, 2023.
7. K. Sathyaseelan, Tushar Vyas, Ravikiran Madala, Vijaya Chamundeeswari, Himanshu Rai Goyal, Raja Jayaraman, "Blockchain Enabled Intelligent Surveillance System Model with AI and IoT", 2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), pp.1-7, 2023.

ID: 10

A REAL TIME WASTE MANAGEMENT TECHNIQUE TO DETECT AND SEGREGATE PLASTIC WASTE USING ACNN ALGORITHM

Dr.R.SMEETA MARY
Assistant Professor,
Department of MCA,
Fatima College,
Madurai, Tamil Nadu, India.

Dr.K.PERUMAL
Professor,
Department of Computer Applications,
Madurai Kamaraj University,
Madurai, Tamil Nadu, India.

ABSTRACT

This research paper provides an application of Image Processing based on the concept of machine learning, with the CNN (Convolutional Neural Network) method paired with Multi-Layered Perceptron to detect the custom item. The suggested invention is a technique for detecting non-biodegradable garbage in bins and conveniently separating the non-biodegradable waste from the bins. To characterize the picture as biodegradable garbage or non-biodegradable waste, a Convolutional Neural Network is integrated with a Multi Layered Perceptron. A dataset of domestic wastes necessary are collected. They are trained and tested using Advanced Convolutional Neural Network, 300 - 300 images of each category were collected in order to train the model. Each image was labeled using a labeling tool in the format of ACNN. A machine learning model was created that was trained on the data set of the images of non- biodegradable waste. Following the conclusion of the training, a number of files were received, which were then employed in the application's testing model. On the basis of the output obtained, accuracy was estimated and results were validated.

KEYWORDS: *Machine Learning, Convolutional Neural Network, Advanced Convolutional Neural Network, Multi-Layered perceptron, non-biodegradable waste.*

1. INTRODUCTION

It's as natural as breathing or bathing to produce garbage. It's all part of the routine. All of the issues come when this garbage is not properly disposed of or separated. It has the potential to harm the economy, the environment, and human health in a variety of ways. After segregation, each type of trash is disposed of in a distinct manner. Waste can be recycled, disposed of in landfills, or burned. Waste that is not adequately separated can pollute soil and water sources. When soil is polluted, land fertility declines, resulting in less food and, as a result, more people go to bed hungry. When water bodies become polluted, the water quality deteriorates, marine/river life suffers, and all creatures and humans that rely on water bodies for sustenance suffer as a result. Pathogens such as Cholera and Dysentery, which were formerly considered epidemics, are also caused by water poisoning. When the improper sort of garbage is disposed of or burnt, additional greenhouse gases are created, resulting in pollution of the atmosphere.

Plastics are the most consumable, conveniently available, and light-weight materials, and they're used for a wide range of residential and industrial applications. They are eaten in every walk of life because to their availability and ease of usage. Because these polymers are non-biodegradable, it is critical to recycle and reuse them. Making plastics reusable necessitates the need to sort them efficiently. The two most elite methods of separating plastics for recycling are manual sorting and automated systems. Manual sorting is used in the majority of recycling procedures, with humans identifying and picking plastic debris on a low-speed conveyor belt [1]. This might not be the best solution for high-volume recycling. It's tough to get workers to follow a consistent plastic separation method. It has also proven to be a time-consuming and stressful process. To distinguish recyclable plastics from other types of garbage, automated sorting systems employ a variety of sensing techniques. These detecting systems employ State-of the art technology to sort plastics automatically [2].To automate trash segregation, a variety of strategies are employed. To categorise garbage, some methods employ image processing techniques. Waste is separated based on its size, colour, texture, and other characteristics. Image processing has also been effectively applied in a variety of applications, and are implemented with the help of microcontrollers [3]. Waste segregation based on sensors is also available. To separate garbage, many types of sensors are utilised, including capacitive sensors for detecting plastic, glass, and wood, inductive proximity sensors for detecting metals, gas detection, and optical sensors for identifying food items, among others [4,5].

2. RELATED WORKS

Waste sorting has recently been a popular notion that may be applied in a variety of settings, including the home, business, and so on. Sorting waste is becoming an increasingly serious issue. The current tendency is to properly categorise garbage in order to manage it effectively. Reducing the quantity of garbage transported to landfills by removing recyclable and reusable rubbish can save the environment. Sorting medical and other non-recyclable garbage can help to reduce health problems and land contamination [6].

Rag pickers have traditionally been responsible for rubbish collection and sorting. This job required a large number of workers. They suffer from a variety of health problems as a result of their lack of knowledge and ignorance about the detrimental consequences of these waste products, such as skin cancer and respiratory ailments. Researchers were inspired to create automated garbage segregators as a result of this [7].

RFID is utilised to identify consumer plastics and electronic wastes instantly, allowing us to determine the sort of trash. A wireless interface has been built that allows the user to manage the RFID reader from a distance, allowing for a more sanitary method of waste separation when the user is away from the garbage. RFID is inexpensive and has the capacity to retain critical information about waste items[8].

Hyper spectral imaging systems (HIS) are used to sort waste. Scanning waste products with shortwave infrared (SWIR) hyper spectral imaging systems used to separate cardboard, paper, metal, glass and six plastic waste waste samples. The use of hyper spectral images provides valuable information about the chemical components of objects. This increases the power of separation and system recognition [9].

The program focuses on filtering polycot containers from plastic bottles. They used a visual filtering method using image stabilization data and SVM was used for partitioning. Unlike other methods this paper uses a gray image separated at the back and a binding box around everything. The system achieved phase accuracy of over 96% using images captured on a single gray scale camera [10].

An SVM-based approach to filtering waste packaging materials was developed. This classification system provides a high level of accuracy and can be used in real-time systems. It automatically separates the types of plastic bottles with an accuracy of about 90%. It also produced a 96% separation rate for polyethylene terephthalate (PET) or non-PET plastic types. In addition it demonstrated 92% accuracy of the separation of non-PET plastic types into High-density polyethylene (HPDE) and Polypropylene (PP) [11].

A convolution network attempting to identify the type of waste in images was made [12] when using CNN and SVM. The databases used here mainly consisted of 3 categories of photographs: plastic, paper and metal of each size. Interestingly, in their survey, while CNN achieved only 83% accuracy, SVM showed 94.8% accuracy.

3. PROPOSED SYSTEM

In this method, predict the category of plastic classes belonging to it, using the Deep Learning algorithm. Advanced Convolutional Neural Networks (ACNN) is used to classify incoming plastic image into other classes such as reusable or non-reusable. The plastic types are shown in Fig.1. These types of plastic items are used for segregation and sorting. The element extracting method is used to find the characteristics of an object in an image and to calculate its feature vector that are used for train and test the neural network model.

3.1 COLLECTION OF DATASETS

The data acquisition process is performed as there are no publicly available data sets relating to plastic items. However, images of the data set accurately reflect the condition of recycled plastic bottles. This is not possible with recycled plastic that is considered waste because the images are crushed, damaged, cracked, and so on. The database contains images of recycled and non-recyclable material contains 900 images. Each class contains a variety of 16,200 bottle shots. This is further distinguished by train images and test images that are important in training the ACNN model. The brightness level and shape of each image are different, which introduces variations in the database. Fig.1 shows some sample classes. Data set images include image rotation at all angles, minimum to maximum image brightness control, random image translation, random image measurement, and random photo crop. This type of image enhancement is preferred to target multiple recycled content to maximize database size.

Fig.1 some sample images of plastic

3.2 FEATURE EXTRACTION

ACNN contains of a sequence of convolutional layer along with max-pooling layers, activation layer and each layer has connected with its previous layer. Figure.4 represents the design of CNN model. It is a general, hierarchical feature extractor which will map input image pixels depth into a feature vector. This will be classified by several fully connected layers in the next step. Feature Extraction is used to extract the number of features in a dataset images by creating new features from the existing ones. These extracts set of features should then be able to summarize most of the information contained in the original set of features. In the Fig.3 feature extraction on deep learning techniques. In this way, a summarized form of the original features can be found from a combined original dataset.

4. METHODOLOGY

4.1. DESIGN OF ADVANCED CONVOLUTION NEURAL NETWORK MODEL:

Convolution Neural Network (CNN) is widely used in image classification. Here the convolution neural network is connected to a multilayer preceptor. Unlike other conventional networks with highly interconnected layers CNN shares weights between the receiving fields. This reduces the number of parameters and CNN uses the filtering method. CNN is a network that reads the filters to be used i.e. the features to be extracted, in order to be categorized. The network takes input data, converts it by calculating the estimated value over input and uses non-linear in this modified input to determine the median position. CNN constitute some combination of these three layers.

The convolution layer contains many filters and each filter produces an image that contains a specific feature. So the output of the convolution layer will be a bunch of images called Tensor. The next layer is the maxpool layer which is used to reduce the size of the installation by combining neighboring elements and taking a higher value. Therefore, a two-dimensional window is selected and integrated over each image steps. After each step we find the values, which fall within the window, from these values, the maximum value is selected and after all the steps of convolution a matrix of all these selected values is formed with the minimum size. This will help reduce the number of parameters and thus the calculation value.

The next layer is Fully Connected Layer which has many units. It is the function of the activation in the layer where the input is converted by non-linear activation. Here each unit is connected to each other the unit of the next layer. A fully integrated layout map data has a high dimension. As the size increases, the accuracy of the output also increases. The most commonly used activation function states ReLu (fixed line opening function) in fig 2. The Convolution Neural Network is combined with multilayer

ISBN: 978-81-965283-7-9

perceptron to give Advanced Convolution Neural Network which is suitable for classification prediction problems where inputs are assigned a class or label.

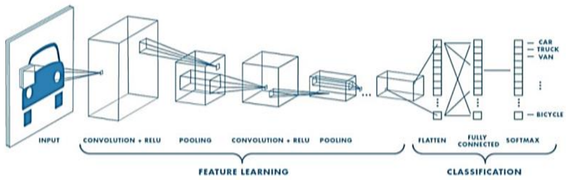


Fig.2 Advanced Convolution Neural Network

4.2 DEEP LEARNING LIBRARIES AND FRAMEWORKS TENSORFLOW:

TensorFlow: It is an open source database arithmetic library using data flow structures. TensorFlow was created and maintained by the Google Brain team within the Google Machine Intelligence research organization Deep Learning. It is currently released under the Apache 2.0 open source license. TensorFlow is designed for large-scale distributed training and testing guidance. The connecting nodes on the graph represent mathematical functions, while the edges of the graph represent a series of multi-sided data (tensors) connected between them. The distributed TensorFlow architecture contains the main distributed resources with kernel usage. These include general functions that include mathematical operations, deception of the same members, flow control and status management functions listed in C ++. TensorFlow is designed for use in both research and development through production systems. It can work on CPU programs and mobile devices as well as large distributed programs for hundreds of nodes.

Keras: Keras is a Python-based Neural Network library based on a framework called TensorFlow. Described as modular, fast and easy to use. Keras does not handle low level calculations. Instead, using another library to do the work, named as Backend. Keras is therefore a threat to the high-level API, which is capable of working in the TensorFlow framework. The Keras High-Level API manages how to create descriptive models, define layers and set multiple input and output models. At this level, Keras also integrates model and loss functions with an optimizer, a training process with appropriate function.

4.3 FLOWCHART OF IMAGE SEGREGATION:

Image classification process based on ACNN is shown in Fig.3. Convolutional Neural Networks with Multilayer perceptron algorithms usually use the published features and learning algorithms to see object context. It's normal used in programs such as image classification with localization, object classification and object acquisition.

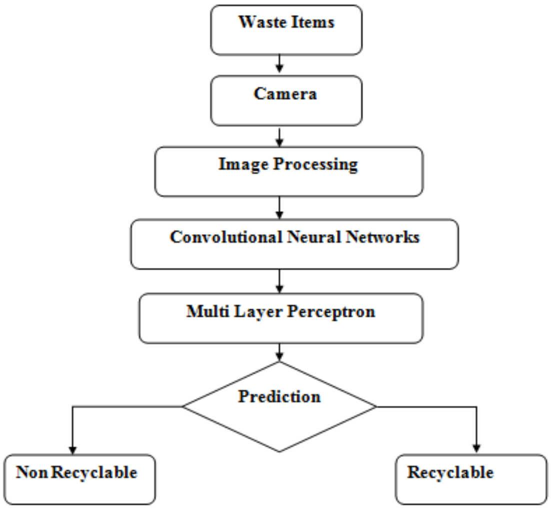


Fig 3 Workflow of Advanced Convolution Neural Network

4.4 TRAINING THE NETWORKS:

Once the network is configured for the acquisition of the object by all parameters, it is ready for training. The descending algorithm of the stochastic gradient is used to train the network. After each repetition, the network connects by reducing the error rate. The loop will terminate if it reaches a minimum number of errors. Here is 0.02. Network weights are adjusted next to each frequency from the initial value based on the result until it converts to a value. The weight of each item is recorded in a backup file. Weight is also used to determine an object.

4.5 TESTING THE NETWORKS:

Pre-trained weight gained from training phase is used in test phase. The inserted image is approved passing through all layers of the neural network as well parameters are available. These numbers are associated with pre-trained weight and identify the provider high compatibility with classes. The plan will consider a label matched next to it.

ISBN: 978-81-965283-7-9

5. RESULT AND DISCUSSION

5.1 FEATURE EXTRACTION

In this series of classifiers in the dense features calculated on different scales on the grid points. The output of these separators is combined using different weighting schemes to obtainan object from an image. The combination of classifiers enhanced results but no single system result and trained in all aspects gave us the best performance, indicating that different features extracted with different scales are needed to achieve something successfully and learning algorithm can do a better job of selecting the right element and its scale accordingly.

Fig.4 A sample for feature extraction

5.2 ACNN CLASSIFICATION

The ACNN segment was deployed on Google's TensorFlow using Keras, which provides a higher output rate over TensorFlow. All images in the database are resized to 64x64 before being fed as network input.

Losses are calculated using cross-entropy binary and the formula used is RMS prop. ACNN trained for approximately 24 hours on an Intel processor. Train classification / verification was 756 photos per class and 100 epochs were used. As Augmented Samples were used for input even when increasing the size of the database, images were highly correlated. This could have resulted in overfitting. The overload problem was solved by modeling the network entropy capacity. Some information stored in the model. A model that can store a lot of information has the potential to be more accurate but also ends up retaining non-essential features.

```
Epoch 1/5
71/71 [==============================] - ETA: 0s -
Epoch 00001: val_acc improved from -inf to 0.79670,
71/71 [==============================] - 367s 5s/st
Epoch 2/5
71/71 [==============================] - ETA: 0s -
Epoch 00002: val_acc did not improve from 0.79670
71/71 [==============================] - 352s 5s/st
Epoch 3/5
71/71 [==============================] - ETA: 0s -
Epoch 00003: val_acc improved from 0.79670 to 0.857
71/71 [==============================] - 351s 5s/st
Epoch 4/5
71/71 [==============================] - ETA: 0s -
Epoch 00004: val_acc did not improve from 0.85714
71/71 [==============================] - 350s 5s/st
Epoch 5/5
71/71 [==============================] - ETA: 0s -
Epoch 00005: val_acc did not improve from 0.85714
71/71 [==============================] - 350s 5s/st
```

Fig.5 Output Predictions of Validation Data

Finally sample image of the plastic were given to the advanced convolutional neural network. Test of the image can be takes place on each layer of the neural network. It contains 50 epochs with operations of 40 steps for each epoch. It classifies the class of the input image and gives the result of image classifier.

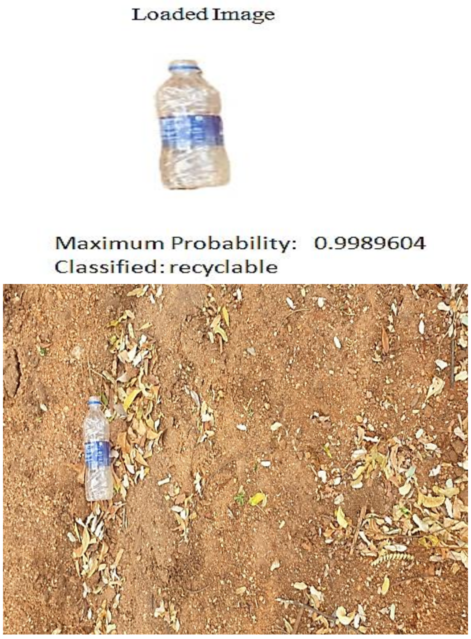


Fig 6 Sample Image1

Fig 6 shows the result of Plastic to check whether it is recyclable or non-recyclable. After loading the image it shows the maximum probability of 0.99896 that is a recyclable plastic and it is classified as recyclable

ISBN: 978-81-965283-7-9

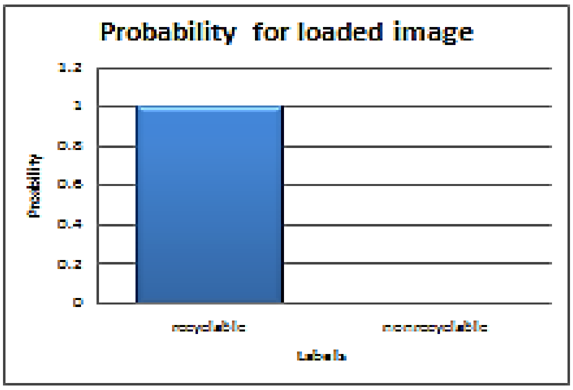


Fig.7 Probability of the sample Image1

Graphical representation of accuracy of result. According to the result the accuracy of detecting plastic 0.998 which means according to our testing the item it detects recyclable Fig 7.

Fig 8 Sample Image2

Fig.8 shows the plastic effect to check whether it is reusable or not reusable. After uploading the image shows a high probability of 0.854363 non recycled plastic and disassembled as non reusable.

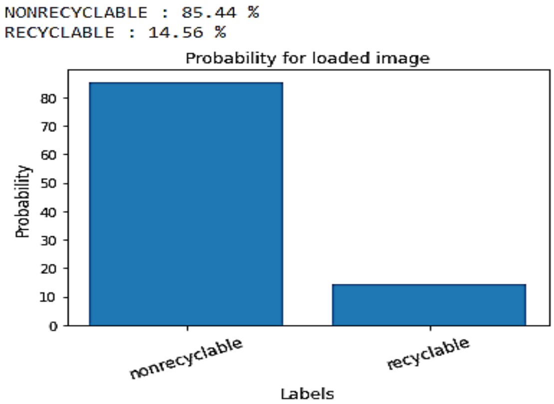


Fig.9 Probability of the sample Image2

According to the result the accuracy of detecting plastic 0.998 which means according to our testing the item it detects recyclable Fig 9.

Epoch	Validation		Training	
	Accuracy	Loss	Accuracy	Loss
10	0.51	1.21	0.53	1.17
20	0.57	1.07	0.62	0.98
40	0.63	1.08	0.71	0.81
60	0.71	0.83	0.74	0.72
80	0.70	0.88	0.78	0.60
100	0.73	0.78	0.82	0.48

Fig.10 Output prediction of validation data

Accuracy, Loss and Precision rate of different types and of bottles are shown. results of train and validation of ACNN model are shown in graph.

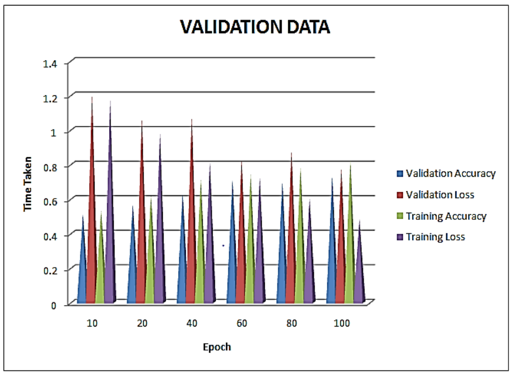


Fig.11 Validation data of the sample Image

6. CONCLUSION AND FUTURE WORK

Recognition and detection of plastic to determine whether it is reusable or not reusable is done using the Advanced Convolutional neural network algorithm. Features are extracted from images and calculated feature vector by feature descriptors using the Convolutional neural network network algorithm. As a result, the accuracy of plastic access is as high as 90.4% in plastic databases. Therefore, the Advanced Convolutional neural network algorithm was very accurate. The use of this provides a good result for the identification of plastic waste. Additional images from a variety of sources will be collected and trained so that the results are more accurate and more accurate. In the future, a robot will be able to take instructions from its default instructions and detect decaying waste and non-perishable waste in bins.

7. REFERENCE

1. Shahrani Shahbudin, Aini Hussain, Dzuraidah Abdul Wahab, Mohd Marzuki Mustafa, Suzaimah Ramli "Support Vector Machines for Automated Classification of plastic bottles" 6th International Colloquium on Signal Processing & Its Applications (CSPA),2010.
2. Wahab, D.A, A. Hussain, E. Scavino, H. Basri and M.M. Mustafa, "Development of a prototype automated sorting system for plastic recycling" Am. J. Applied Sci., 3 (7):1924-1928
3. Sharanya A, Harika U, Sriya N, Kochuvila S Automatic waste segregator2017 Int.Conf on Advances in ComputingCommunications and Informatics (ICACCI)Udupi 1313-1319
4. Chandramohan A, Mendonca J, Shankar N R, Baheti N U, Krishnan N K, Suma M S Automated Waste Segregator 2014 Texas Instruments India Educators' Conference (TIIEC), Bangalore 1-6
5. A A John, Binoy B Nair and Kumar P N Application of clustering techniques for video summarization - An empirical study 2017 Advances in Intelligent Systems and Computing 573 494-506.
6. Islam, M. S., Hannan, M. A., Basri, H., Hussain, A., & Arebey, M. (2014). Solid waste bin detection and classification using Dynamic Time Warping and MLP classifier. Waste management, 34(2), 281-290.
7. Chandramohan, A., Mendonca, J., Shankar, N. R., Baheti, N. U., Krishnan, N. K., & Suma, M. S. (2014, April). Automated Waste Segregator. In 2014 Texas Instruments India Educators' Conference (TIIEC) (pp. 1-6). IEEE.
8. Balramdu, M. P., Kumar, M. M., Murlidhar, M. C. L., Sachin, M. W., Sudamrao, M., Ulhas, P., & Ramkrushna, M. K. N. (2014). Remote Monitoring and Sorting System for Waste Material Management Using RFID. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 3(10).
9. Karaca, A. C., Ertürk, A., Güllü, M. K., Elmas, M., & Ertürk, S. (2013, June). Automatic waste sorting using shortwave infrared hyperspectral imaging system. In 2013 5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS) (pp. 1-4). IEEE.
10. Nawrocky, M., Schuurman, D. C., & Fortuna, J. (2010, May). Visual sorting of recyclable goods using a support vector machine. In CCECE 2010 (pp. 1-4). IEEE.
11. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
12. Sakr, G. E., Mokbel, M., Harwich, A., Khneisser, M. N., & Hadi, A. (2016, November). Comparing deep learning and support vector machines for autonomous waste sorting. In 2016 IEEE International Multidisciplinary Conference on Engineering Technology (IMCET) (pp. 207-212). IEEE.

ID: 11

DEEP LEARNING ALGORITHMS AND ITS APPLICATION

Ms.I.RAZUL BEEVI
Assistant Professor,
Department of Computer Science & Research Center,
Thassim Beevi Abdul Kader College for Women,
Kilakarai, Tamil Nadu, India.

Mr.T.BALAJI
Assistant Professor,
Department of Computer Science,
Government Arts College,
Melur, Tamil Nadu, India.

ABSTRACT

Deep learning is a machine learning technique that teaches computers to do what comes naturally to humans: learn by example. Deep learning is a key technology behind driverless cars, enabling them to recognize a stop sign, or to distinguish a pedestrian from a lamppost. It is the key to voice control in consumer devices like phones, tablets, TVs, and hands-free speakers. Deep learning is getting lots of attention lately and for good reason. It's achieving results that were not possible before. In deep learning, a computer model learns to perform classification tasks directly from images, text, or sound. Deep learning models can achieve state-of-the-art accuracy, sometimes exceeding human-level performance. Models are trained by using a large set of labeled data and neural network architectures that contain many layers.

1. INTRODUCTION

Deep learning is a branch of machine learning which is based on artificial neural networks. It is capable of learning complex patterns and relationships within data. In deep learning, we don't need to explicitly program everything.

It has become increasingly popular in recent years due to the advances in processing power and the availability of large datasets. Because it is based on artificial neural networks (ANNs) also known as deep neural networks (DNNs). These neural networks are inspired by the structure and function of the human brain's biological neurons, and they are designed to learn from large amounts of data. Some of the popular Deep Learning architectures include Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Deep Belief Networks (DBNs).

Training deep neural networks typically requires a large amount of data and computational resources. However, the availability of cloud computing and the development of specialized hardware, such as Graphics Processing Units (GPUs), has made it easier to train deep neural networks.

1.1 Application

Automated Driving: Automotive researchers are using deep learning to automatically detect objects such as stop signs and traffic lights. In addition, deep learning is used to detect pedestrians, which helps decrease accidents.

Aerospace and Defense: Deep learning is used to identify objects from satellites that locate areas of interest, and identify safe or unsafe zones for troops.

Medical Research: Cancer researchers are using deep learning to automatically detect cancer cells. Teams at UCLA built an advanced microscope that yields a high-dimensional data set used to train a deep learning application to accurately identify cancer cells.

Industrial Automation: Deep learning is helping to improve worker safety around heavy machinery by automatically detecting when people or objects are within an unsafe distance of machines.

Electronics: Deep learning is being used in automated hearing and speech translation. For example, home assistance devices that respond to your voice and know your preferences are powered by deep learning applications.

2. ALGORITHMS

2.1 Convolutional Neural Networks (CNNs)

CNNs process the data by passing it through multiple layers and extracting features to exhibit convolutional operations. The **Convolutional Layer** consists of **Rectified Linear Unit** (ReLU) that outlasts to rectify the feature map. **The Pooling layer** is used to rectify these feature maps into the next feed. Pooling is generally a sampling algorithm that is down-sampled and it reduces the dimensions of the feature map. Later, the result generated consists of **2-D arrays** consisting of **single, long, continuous,** and **linear vector** flattened in the map. The next layer i.e., called **Fully Connected Layer** which forms the flattened **matrix** or **2-D** array fetched from the Pooling Layer as input and identifies the image by classifying it.

2.2 Long Short Term Memory Networks (LSTMs)

LSTMs can be defined as **Recurrent Neural Networks** (RNN) that are programmed to learn and adapt for dependencies for the long term. It can memorize and recall past data for a greater period and by default, it is its sole behavior. LSTMs are designed to retain over time and henceforth they are majorly used in time series predictions because they can restrain memory or previous inputs. This analogy comes from their **chain-like** structure consisting of **four** interacting layers that communicate with each other differently. Besides applications of time series prediction, they can be used to construct **speech recognizers, development in pharmaceuticals,** and composition of **music loops** as well.

2.3 Recurrent Neural Networks (RNNs)

RNNs have connections that form directed cycles, which allow the outputs from the LSTM to be fed as inputs to the current phase.

The output from the LSTM becomes an input to the current phase and can memorize previous inputs due to its internal memory. RNNs are commonly used for image captioning, time-series analysis, natural-language processing, handwriting recognition, and machine translation.

2.4 Generative Adversarial Networks (GANs)

GANs are generative deep learning algorithms that create new data instances that resemble the training data. GAN has two components: a generator, which learns to generate fake data, and a discriminator, which learns from that false information.

The usage of GANs has increased over a period of time. They can be used to improve astronomical images and simulate gravitational lensing for dark-matter research. Video game developers use GANs to upscale low-resolution, 2D textures in old video games by recreating them in 4K or higher resolutions via image training.

GANs help generate realistic images and cartoon characters, create photographs of human faces, and render 3D objects

2.5 Radial Basis Function Networks (RBFNs)

RBFNs are special types of feedforward neural networks that use radial basis functions as activation functions. They have an input layer, a hidden layer, and an output layer and are mostly used for classification, regression, and time-series prediction.

The input layer has **neurons** that are sensitive to these data and the nodes in the layer are efficient in classifying the class of data. Neurons are originally present in the hidden layer though they work in close integration with the input layer.

The hidden layer contains **Gaussian transfer** functions that are inversely proportional to the distance of the output from the neuron's center.

The output layer has linear combinations of the **radial-based** data where the Gaussian functions are passed in the neuron as parameter and output is generated.

2.6. Multilayer Perceptrons (MLPs)

MLPs are an excellent place to start learning about deep learning technology. MLPs belong to the class of feedforward neural networks with multiple layers of perceptrons that have activation functions. MLPs consist of an input layer and an output layer that are fully connected. They have the same number of input and output layers but may have multiple hidden layers and can be used to build speech-recognition, image-recognition, and machine-translation software.

- MLPs feed the data to the input layer of the network. The layers of neurons connect in a graph so that the signal passes in one direction.

- MLPs compute the input with the weights that exist between the input layer and the hidden layers.
- MLPs use activation functions to determine which nodes to fire. Activation functions include ReLUs, sigmoid functions, and tanh.
- MLPs train the model to understand the correlation and learn the dependencies between the independent and the target variables from a training data set.

2.7 Self Organizing Maps (SOMs)

Data visualization to reduce the dimensions of data through self-organizing artificial neural networks. Data visualization attempts to solve the problem that humans cannot easily visualize high-dimensional data. SOMs are created to help users understand this high-dimensional information.

- SOMs initialize weights for each node and choose a vector at random from the training data.
- SOMs examine every node to find which weights are the most likely input vector. The winning node is called the Best Matching Unit (BMU).
- SOMs discover the BMU's neighborhood, and the amount of neighbors lessens over time.
- SOMs award a winning weight to the sample vector. The closer a node is to a BMU, the more its weight changes..
- The further the neighbor is from the BMU, the less it learns. SOMs repeat step two for N iterations.

2.8 Deep Belief Networks (DBNs)

DBNs are generative models that consist of multiple layers of stochastic, latent variables. The latent variables have binary values and are often called hidden units.

DBNs are a stack of Boltzmann Machines with connections between the layers, and each RBM layer communicates with both the previous and subsequent layers. Deep Belief Networks (DBNs) are used for image-recognition, video-recognition, and motion-capture data.

- Greedy learning algorithms train DBNs. The greedy learning algorithm uses a layer-by-layer approach for learning the top-down, generative weights.
- DBNs run the steps of Gibbs sampling on the top two hidden layers. This stage draws a sample from the RBM defined by the top two hidden layers.
- DBNs draw a sample from the visible units using a single pass of ancestral sampling through the rest of the model.
- DBNs learn that the values of the latent variables in every layer can be inferred by a single, bottom-up pass.

2.9 Restricted Boltzmann Machines (RBMs)

Developed by Geoffrey Hinton, RBMs are stochastic neural networks that can learn from a probability distribution over a set of inputs. This deep learning algorithm is used for dimensionality reduction, classification, regression, collaborative filtering, feature learning, and topic modeling. RBMs constitute the building blocks of DBNs.

RBMs consist of two layers:

- Visible units
- Hidden units

Each visible unit is connected to all hidden units. RBMs have a bias unit that is connected to all the visible units and the hidden units, and they have no output nodes.

2.10 Autoencoders

Autoencoders are a specific type of feedforward neural network in which the input and output are identical. Geoffrey Hinton designed autoencoders in the 1980s to solve unsupervised learning problems. They are trained neural networks that replicate the data from the input layer to the output layer. Autoencoders are used for purposes such as pharmaceutical discovery, popularity prediction, and image processing.

An autoencoder consists of three main components: the encoder, the code, and the decoder.

- Autoencoders are structured to receive an input and transform it into a different representation. They then attempt to reconstruct the original input as accurately as possible.
- When an image of a digit is not clearly visible, it feeds to an autoencoder neural network.
- Autoencoders first encode the image, then reduce the size of the input into a smaller representation.

- Finally, the autoencoder decodes the image to generate the reconstructed image.

CONCLUSION

Deep learning has evolved over the past five years, and deep learning algorithms have become widely popular in many industries. If you are looking to get into the exciting career of data science and want to learn how to work with deep learning algorithms.To understand these algorithms, a person needs high clarity with mathematical functions discussed in some of the algorithms. These functions are so crucial that the working of these algorithms mostly depends on the calculations done by using these functions and formulae. An aspiring deep learning engineer knows all of these algorithms, and it is highly recommended for beginners to understand these algorithms before moving ahead into artificial intelligence.

REFERENCE

1. https://in.mathworks.com/discovery/deep-learning.html
2. https://www.geeksforgeeks.org/introduction-deep-learning/
3. https://www.simplilearn.com/tutorials/deep-learning-tutorial/deep-learning-algorithm
4. https://www.javatpoint.com/deep-learning-algorithms

ID: 12

AI AND WIRELESS IOT SENSOR NETWORK BASED SMART AGRICULTURE

Dr.R.HEMALATHA
Head & Associate Professor,
PG & Research Department of Computer Science,
Tirupur Kumaran College for Women,
Tirupur, Tamil Nadu, India.

Ms.G.NIVASHINI
Research Scholar,
PG & Research Department of Computer Science,
Tirupur Kumaran College for Women,
Tirupur, Tamil Nadu, India.

ABSRACT

Artificial Intelligence (AI) is experiencing a demanding growth in smart cities, agriculture, energy management, and weather forecasting due to lack of computing power on sensor devices. Applications of Artificial intelligence are embedded in various and common Internet of Things (IoT) sensors for improve the agricultural sector and reduce management costs. Due to the limited resources of wireless technology, most solutions are designed to efficiently deliver agricultural data to cloud systems, however, it still optimizes resource management and data loading for relay nodes, especially nodes nearest edge boundary is a tricky problem. In addition, due to inaccurate environmental data collection, decision-making process leads to a reduction in the productivity of the optimization process.

Overcome these questions, this work proposes a smart and reliable agricultural model using meta heuristic optimization improve resource management to address these issues. The approach of the proposed model uses decision-making functions to overcome information loss and inconsistencies. In addition, it builds trust in collect agricultural data using secure IoT devices and facilitates reliable communication. in terms of performance measures, the proposed model is simulated to assess its importance relative to the state of the art solutions. It not only collects updated farmland data, but also uses the light weight of artificial intelligence Optimization techniques to reduce operating costs on IoT devices.

Experimental results show that importance of the proposed model for monitoring resource and operational costs on an IoT system.

KEYWORDS: *Smart Agriculture (SA), Wireless Sensors Networks (WSN), Internet of Things (IoT), Fuzzy Logic Control, Information Communication Technology (ICT).*

INTRODUCTION

Water is the premise and the fundamental motor of life on the planet. People use water for modern purposes, disinfection, and water system. Somewhat recently, the yearly water withdrawal gone between 11 billion and 15 billion cubic meters for every year, out of which 69% is utilized in agribusiness. Sadly, the greater part of this water is squandered due to deficient water system control frameworks. As in generally bone-dry and sub-Saharan nations, horticulture in Morocco is the biggest buyer of new water, particularly subsequent to sending off the Green Arrangement The partner proofreader organizing the survey of this original copy and endorsing it for distribution was Emrecan Demirors. This program means to advance farming as an efcient area fit for propelling the economy, fighting neediness, and safeguarding many individuals in country regions efciently and economically. The government provided numerous facilities and assistance to farmers and investors in irrigated agriculture as part of this program to supply sufficient basic food for local consumption and export promotion programs. Not withstanding, the degree of Savvy Farming entrance in Morocco remains extremely low. Likewise, financial backers in the watered horticultural activities begun sounding alerts of extreme exhaustion of groundwater not too far off and the absence of a financially savvy realtime information assortment in water system frameworks in cultivating fields that will empower them to benefit from the high level present day advances. There is a need to foster a practical furthermore, manageable information obtaining framework for shrewd horticulture applications in sub-Saharan elds for practical, efcient, furthermore, brilliant farming.

The framework ought to utilize sustainable energy and the most recent Data and Correspondence Advancements (ICT) that can support the aridity of a horticultural climate. The data acquisition system can play a crucial role in increasing agricultural productivity and crop quality when combined with reasonable control and management and data analytics. Above all, not at all like the customary strategies for inundated agribusiness that broadly utilize underground water repositories what's more, resort to petroleum products, particularly gas, as a source of energy. Brilliant horticulture (SA) takes advantage of present day administration frameworks to justify water utilization and take on sustainable power sources as a wellspring of energy, in this way delivering the rural area eco-accommodating. In this paper, we are introducing a vital and practical SA arrangement.

When drawn nearer from a Digital Actual Framework (CPS) point of view, our answer depends on four fundamental ICTs: 1. A Remote Sensors Network (WSNs) screen, continuously, the plant natural conditions, e.g., climate and soil conditions, 2. A Remote Actuators Organization (WAN) follows up on electric machines, for example, water siphons and lights, 3. A Cloudbased IoT stage for ongoing information stockpiling, handling, also, perception; and 4. A Fluffy Rationale Control module settles on checking water system spans in view of the ongoing procured contrasts among wanted and encompassing soil dampness. We sent our open-source programming arrangement for information procurement, incitation, and control. Utilizing readily available, inexpensive nano-Arduinos, we constructed our cost-effective sensors and actuators. At last, we utilized a current free Cloud-based IoT stage, e.g., NodeRED. In this paper, and by taking into account ideal water-ground use, sustainable power reconciliation, and open-source ICTbased brilliant water system, we intend to contribute towards the smooth infiltration of Brilliant Horticulture into oppressed sub-Saharan nations. We are further imagining, through this work, to set a strong foundation for laying out a devoted Cloud-based and HPC (Elite Execution Processing) stage to gather continuous information about water-table use. This information, which falls under the enormous information class, as it bears the enormous information 3Vs (Volume, Speed, and Assortment), and alongside suitable Huge Information Examination apparatuses, will colossally help with advancing eco-accommodating shrewd agribusiness.

The paper commitments are as per the following: - We present a genuine world, savvy, and simple todeploy general engineering for a Savvy Homestead testbed that the examination local area can undoubtedly take on and adjust for further testing and improvement. In detail how to utilize open-source programming and influence Distributed computing for remote sensor information capacity and handling. - We exhibit the combination of our created savvy remote sensors and actuators utilizing offthe- rack equipment in a certifiable contextual investigation.- We show how incorporating sustainable power into savvy ranches clears the way towards embracing our answer in off-lattice locales, e.g., dry and sub-Saharan regions.

LITERATURE SURVEY

The sending of IoT in farming can possibly influence our general public and the remainder of the world. These days, we see climate, soil, and water evaporating as land that is basic to horticulture declines, making it increasingly hard to deliver food. Agriculturalists will profit from utilizing Web of Things (IoT) innovation, which will assist them with eliminating created badlands while too further developing creation. This figure could originate from the quantity of missions the farmhouse cars have performed, or from how much fertilizer utilized in the treating the soil cycle. A savvy horticulture framework might be characterized as a food framework that uses arising sustenance that is uncontaminated and is open to a wide number of individuals.

The Internet of Things (IoT) plays a larger role in agriculture as a result of the Smooth Agricultural IoT platform's expansion of the entire farming system. Despite the fact that the Internet of Things (IoT) is used in farming, the interconnectedness of devices and services has not only saved agriculturalists time but also enormous amounts of liquid and power. It can protect every now and again experienced geologies including dampness, high temperature, soil, and so forth. And provides surveillance in real time via a crystal-clear map. In horticulture, embracing Web of Things (IoT) will yield different advantages. For instance, the farmhouse autos have achieved various missions. Savvy horticulture is in this way essentially a coordinated, uncontaminated technique for arising nourishment that supports swarms. The farming system is extended by the smooth agricultural system, which not only watches the soil but also physically monitors it [5,6]. Despite the fact that singular gadgets and systems administration aren't straightforwardly saving the agriculturalists' time, the Web of Things (IoT) is adversely influencing inefficient spending on resources like Fluid and Power. The general objective of this model is to save regular geological highlights like dampness, temperature, soil, and other data, and give an ongoing perfectly clear reconnaissance. Notwithstanding the benefits depicted above, agribusiness will profit from executing Web of Things (IoT).

Tunable Fields: Accuracy horticulture is a way or practice that makes the cultivating system more right and figured out how to raise live stocks and develop crops. Its utilization and articles like sensors, self-supporting vehicles, PC equipment, control frameworks, advanced mechanics and numerous others. The primary added substances are in this method. Accuracy cultivating utilizing IoT is introduced. Accuracy agribusiness has become one of the most notable horticultural IoT drives lately and this method has started to be utilized by an enormous number of associations.

Data Analytics: The anticipated data set framework currently has insufficient parking spot to store current realities from the IoT sensors. In the cunning agribusiness machine the cloud based basically realities carport furthermore, a stop IoT stage has a significant impact. These designs are supposed to assume an indispensable part in completing higher games. In the IoT universe, sensors are the fundamental wellspring of huge realities. Analytical tools are used to transform the numbers into useful information. The examination of records assists with assessing temperature, livestock conditions and yield circumstances. The measurements gathered utilize the specialized advances and go with better choices for this reason. You can comprehend the plants' constant standing utilizing IoT gadgets, by gathering current realities from sensors. You will acquire an insight utilizing prescient investigation to settle on better choices on reaping. The style examination permits ranchers to comprehend future climatic circumstances and vegetation gathering. IoT empowered ranchers to keep up with the nature of the vegetation and the fruitfulness of the land in the rural industry, hence endlessly further developing how much items remarkable.

Environment Conditions: Environment assumes a crucial part for horticulture. Furthermore, mixed up environment know-how falls apart the amount and the top notch of yield creation significantly. Yet, IoT replies permit you to realize the climate circumstance progressively. Inside and outside the rural fields, sensorsare mounted. They gather natural insights that are utilized to choose the right plants that can develop what's more, support in exact climatic circumstances. The whole IoT air comprises of sensors that can find conditions like stickiness, precipitation, temperature and all the more precisely progressively. There are different sensors that must be hit on every one of these boundaries and designed to accommodate your cunning rural requirements. These sensors uncover what is happening and climatic states of the yields around them. Assuming any alarming climatic circumstances are laid out, the boat is a caution. The absence of actual presence in disturbing climatic circumstances, which in the long run expands efficiency and assists ranchers with acquiring higher farming authorisations, is eliminated.

Savvy Nursery: Nursery farming is a procedure supplementing crop yields, greens, final products and so forth. Nurseries handle natural boundaries in two ways, either physically or by a corresponding control instrument. Notwithstanding, these methodologies are significantly less fruitful in light of the fact that manual mediation has gambles, including creation misfortune, energy misfortune and work costs. A savvy, IoT-based nursery presently doesn't just track yet directs the climate. There is a requirement for human activity. Different sensors that are in accordance with plant necessities are utilized to screen the climate in an astute nursery. A cloud server then, at that point, makes a PC for remote access while partner IoT utilization. The cloud server secretly empowers records to be handled and deals with the stream. Farmers can get the best and most efficient solutions from this design with almost no or little manual work required.

Drones for agriculture: Logical advances have nearly changed farming tasks also, there is pattern aggravation with the appearance of farming robots. The fitness assessment, crop inspection, planting, crop spraying, and field assessment are all carried out with ground and airborne drones. With the right methodology and arrangement in light of genuine reality, the robot age has given the farming industry a solid push and redesign. Drones with warm or multi-ghostly sensors select regions where water system changes are vital. At the point when plants start to develop, sensors demonstrate their wellbeing and measure the file of their plants. Shrewd robots have ultimately decreased the natural effect. The innovation based Savvy cultivating utilizing Robots is introduced. The outcomes were such that there was an extraordinary decrease in the synthetic impacts of groundwater and Domesticated animals Web observing correspondences permit ranchers to get materials about their domesticated animals area, appropriate life and government assistance. They are able to identify the position of their livestock thanks to this calculation. For example, recognizing creatures that are awkward to separate the group, forestalling the illness from creating for the entire creature. The ability of farmers to locate their farm animals using Internet of Things (IoT)-based sensors enables them to transport lower labor costs by a significant amount.

IOT IRRIGATION SYSTEM

Crop docking and air temperature estimation sensors are associated with the organization passage through a Remote Sensor Organization (WSN). ZigBee is a well known decision with various applications since it is easy to carry out and customize. The association with the Web will be at a lower cost since this product just requires a restricted amount of information limit. 4G LTE versatile correspondences are being utilized to associate with the Web remotely. Versatile organization information is shipped off the cloud-based web benefits that are bought into, and the information is made accessible to other cloud-based web services [13, 14].

Programming that can incorporate a wide range of information and consolidate it with CWSI models to make water system list values can break down rural land information and, in light of this data, can apply CWSI models to survey water need. Weather conditions administration and satellite symbolism are other significant wellsprings of data, and the outcomes are water system record values for every area. The information gave to the organization entryway is shipped off a regulator for water system the executives, after which it is sent back to the door. The two information and results, as well as making changes to terminals, might be seen through particular web applications, and ranchers may likewise utilize these web applications to give different ranchers and experts admittance to their information what's more, results.

IOT CAN IMPROVE AGRICULTURE

The Web of Things (IoT) is utilized by ranchers to carry out specific IoT techniques for improving their farming. Concentrated punishments for these practices might convince ranchers to leave them. 1. Data; loads of data gathered by smart agricultural sensors about things like weather, healthy soil, crop growth, and animal health. These data can be used to help your company improve the smartness of its country and the overall performance, productivity, and efficiency of its employees. 2. Diminished creation takes a chance because of further developed exchanging over inner procedures. Assuming that you can expect the result of your assembling, you can prepare and be more ready to convey your products on time. If precisely how much undergrowth you will develop, you could convince your items that their deal is expanding. 3. Expanded business effectiveness by process motorization. You will dominate different innovations through the structure period, for example seepage, treating the soil or bug control by applying smooth techniques. 4. Spending plan association and unused diminishing familiarity with the producer by the expanded regulator. On the off chance that you are savvy to see anomalies or control wellness in collect, you can moderate the risks behind your items. 5. Further developed predominance and limit development. Moved along controller finished development improvement and laid out standards of creation greatness and expanding motorization finished with volume.

CHALLENGES OF IOT FOR AGRICULTURE

The is an All brilliant cultivating reactions should begin with information examination. The information you've gathered will be of no use if you can't understand it. Therefore, you'll areas of strength for require investigation, prescient calculations, and gadgets to break down the information and determine functional experiences [18, 19]. The hardware is made of choosing the sensors for your instrument is basic for web of things farming (or make a custom one). How you search for data and how you go with choices all impact your decision. Anything that the circumstance might be, it is feasible to observe the adequacy of your item founded on the quality and consistency of the information you assemble. Since sensors are frequently used in agriculture Internet of Things products and can be easily destroyed, hardware maintenance is an important project. Accordingly, you should verify that your equipment is both solid and easy to keep up with. Then, at that point, you'll need to refresh your sensors on a more regular basis than you'd like. To get to the data on the site, an organization proprietor or homestead manager should have the option to utilize a versatile telephone or a PC anyplace in the globe. Moreover, each connected apparatus should be independent and have adequate remote variety to associate with different gadgets and communicate information to the focal server. The administrations accessible You'll require a vigorous inner foundation to guarantee that your savvy cultivating application moves along as planned (and that the heap of records can deal with it). Moreover, the interior frameworks should be lovely to utilize. If we don't make our system easier to use, someone who interrupts you, steals your information, or even uses your self-satisfying tractors will be more likely to use it.

CONCLUSION

This article describes the management of physical sensors that are dispersed throughout a WSN using sensor-cloud architecture, a more advanced form of cloud computing. The gathering, storing, and processing of data from the installed sensors is the main goal. For the purpose of running workflows in a virtualized sensor-cloud environment, a brand-new scheduling method called Optimum Energy and Resource Aware Workflow Scheduling (OERES) is suggested. The Fuzzy Membership Mutation Elephant Herding Optimization (FMMEHO) algorithm, which allocates the workflow tasks to virtual machines, is used to create this schema. Additionally, this suggested solution deploys and un-deploys the VMs in accordance with the requirements. The suggested approach effectively schedules the tasks with lowered EC, increased RU, and without going against the time limitation. The two changes made to EHO are as follows: Shiny reflection To increase population diversity, understand how to change the initialization and Fuzzy membership functions. These crucial changes can efficiently control energy use and reduce pan time during workflow execution. When a task is correctly scheduled to the VM that executes the parent's duties, the data transfer EC is reduced. The suggested algorithm chooses the best virtual machine based on resource capacity and energy Sefficiency. The algorithm maintains dependence and deadline constraints. The very dynamic nature of the proposed OERES schema makes it easier for services to be automatically provided as needed by users. The CloudSim toolbox is used to simulate the suggested technique. To evaluate the performance, a total of 3 types of workflows have been developed. The parameters RU, TEC, ECT, ETT, and Makespan are used to measure the effectiveness of the approaches. In future, this system can be extended to be applied to other types of sensor network with different workflows.

REFERENCES

1. Ojha, T., Misra, S. and Raghuwanshi, N.S., 2015. Wireless sensor networks for agriculture: The state-of-the-art in practice and future challenges. Computers and electronics in agriculture, 118, pp.66-84.
2. Abbasi, A.Z., Islam, N. and Shaikh, Z.A., 2014. A review of wireless sensors and networks' applications in agriculture. Computer Standards & Interfaces, 36(2), pp.263-270.
3. Nikolidakis, S.A., Kandris, D., Vergados, D.D. and Douligeris, C., 2015. Energy efficient automated control of irrigation in agriculture by using wireless sensor networks. Computers and Electronics in

Agriculture, 113, pp.154-163.
5. Gutiérrez, J., Villa-Medina, J.F., Nieto-Garibay, A. and Porta-Gándara, M.Á., 2013. Automated irrigation system using a wireless sensor network and GPRS module. IEEE transactions on instrumentation and measurement, 63(1), pp.166-176.
6. Chen, N., Zhang, X. and Wang, C., 2015. Integrated open geospatial web service enabled cyber-physical information infrastructure for precision agriculture monitoring. Computers and Electronics in Agriculture, 111, pp.78-91.
7. Malaver, A., Motta, N., Corke, P. and Gonzalez, F., 2015. Development and integration of a solar powered unmanned aerial vehicle and a wireless sensor network to monitor greenhouse gases. Sensors, 15(2), pp.4072-4096.
8. Dong, X., Vuran, M.C. and Irmak, S., 2013. Autonomous precision agriculture through integration of wireless underground sensor networks with center pivot irrigation systems. Ad Hoc Networks, 11(7), pp.1975-1987.
9. Sales, N., Remédios, O. and Arsenio, A., 2015, Wireless sensor and actuator system for smart irrigation on the cloud. In 2015 IEEE 2nd World Forum on internet of things (WF-IoT) ,pp. 693-698.
10. Supreetha, M.A., Mundada, M.R. and Pooja, J.N., 2018, Design of a smart water-saving irrigation system for agriculture based on a wireless sensor network for better crop yield. In International Conference on Communications and Cyber Physical Engineering 2018 (pp. 93-104). Springer, Singapore.
11. Darshna, S., Sangavi, T., Mohan, S., Soundharya, A. and Desikan, S., 2015. Smart irrigation system. IOSR Journal of Electronics and Communication Engineering (IOSR-JECE), 10(3), pp.32-36.
12. Ghosh, S., Sayyed, S., Wani, K., Mhatre, M. and Hingoliwala, H.A., 2016, Smart irrigation: A smart drip irrigation system using cloud, android and data mining. In 2016 IEEE International Conference on Advances in Electronics, Communication and Computer Technology (ICAECCT) ,pp. 236-239.
13. Savić, T. and Radonjić, M., 2018, WSN architecture for smart irrigation system. In 2018 23rd International Scientific-Professional Conference on Information Technology (IT), pp.1-4.
14. Ojha, T., Misra, S. and Raghuwanshi, N.S., 2017. Sensing-cloud: Leveraging the benefits for agricultural applications. Computers and electronics in agriculture, 135, pp.96-107.
15. Tyagi, S., Obaidat, M.S., Tanwar, S., Kumar, N. and Lal, M., 2017, Sensor cloud based measurement to management system for precise irrigation. In GLOBECOM 2017-2017 IEEE Global Communications Conference, pp. 1-6.
16. Zhang, M., Xiong, S. and Wang, L., 2019, Sensor-cloud based precision sprinkler irrigation management system. In China Conference on Wireless Sensor Networks (pp. 184-197). Springer, Singapore.
17. Garg, N., Goraya, M.S.: Task deadline-aware energy-efficient scheduling model for a virtualized cloud. Arab. J. Sci. Eng. 43(2), 829–841 (2017).
18. Zhang, R., VE, S., & Jackson Samuel, R. D. 2020. Fuzzy efficient energy smart home management system for renewable energy resources. Sustainability, 12(8), pp.3115.
19. Abdul Haseeb-ur-rehman, R.M., Liaqat, M., Aman, A.H.M., Ab Hamid, S.H., Ali, R.L., Shuja, J. and Khan, M.K., 2021. Sensor Cloud Frameworks: State-of-the-Art, Taxonomy, and Research Issues. IEEE Sensors Journal, pp.1-26.
20. Liang, Y., Wang, T., Bhuiyan, M.Z.A. and Liu, A., 2017, Research on coupling reliability problem in sensor-cloud system. In International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage (pp. 468-478). Springer, Cham.
21. Chatterjee, S. and Misra, S., 2015, Optimal composition of a virtual sensor for efficient virtualization within sensor-cloud. In 2015 IEEE International Conference on Communications (ICC) ,pp. 448-453.
22. Chatterjee, S., Misra, S. and Khan, S.U., 2015. Optimal data center scheduling for quality of service management in sensor-cloud. IEEE Transactions on Cloud Computing, 7(1), pp.89-101.
23. Chatterjee, S., Ladia, R. and Misra, S., 2015. Dynamic optimal pricing for heterogeneous service-oriented architecture of sensor-cloud infrastructure. IEEE Transactions on Services Computing, 10(2), pp.203-216.

ID: 13

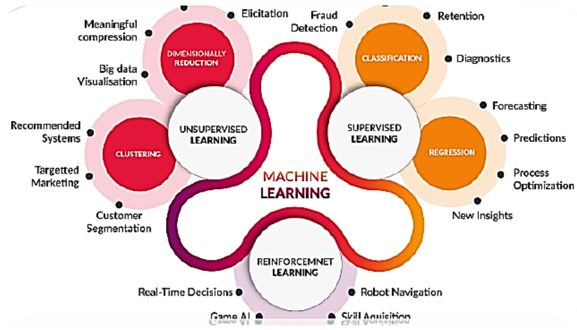
MACHINE LEARNING AND DEEP LEARNING APPLICATIONS

Ms.R.PAVITHRA
Assistant Professor,
Department of Commerce with Information Technology,
Dr.SNS Rajalakshmi College of Arts and Science,
Coimbatore, Tamil Nadu, India.

Dr.S.VENGATESHKUMAR
Associate Professor,
Department of Cyber Security,
Dr. SNS Rajalakshmi College of Arts & Science,
Coimbatore, Tamil Nadu, India.

1. INTRODUCTION

Machine learning is a field of study which can be able to learn the data or experience from the computer to make a predictions based on their own experience. It can develop the algorithms and statistics model to enables the computer to improve their performance in tasks. Machine learning are able to learn the patterns to improve the new data to exposed the data. The purpose of this chapter is to provide the readers to overview the vast range of applications which have at their heart a machine learning problem and to bring some degree of order to the zoo of problem. After that, we will have discussed the problem patterns of statistics and probability which can involve in the training of data. since, they from the programming languages which is machine learning problems that can be phrases to became an amenable to solving. Finally, we will outline a set of fairly basic yet effective algorithms to solve an important problem, namely that of classification.



1.2 Deep learning

Deep learning is the subfield part of machine learning, that can be based on artificial neural network architecture. The artificial neural network or ANN layers are used to interconnected with the nodes are called an neurons, that works can be process of learning the input data networks. The fully connected of deep learning neural network, her is an input a layer and hidden the one or more layers. The output of this learning is one neurons becomes the input to other neurons in the next layer of network and process will continue when the final layer can produce the output of network. The neural network transform of the layers that can be the input data through a series of nonlinear transformation and allowing the networks to learn, complex, representation of neural network. Now a days deep learning has become the most popular to visible areas of machine learning, due to its success in a variety of applications, such as computer vision, natural language processing, and Reinforcement learning. Deep learning can be used for supervised, unsupervised as well as reinforcement machine learning. it uses a variety of ways to process these.

1.3 Approaches to Machine Learning

Supervised Machine Learning

Is the technique of machine learning, which can be learns to which is neural network to make them a predictions of the data classify as based on the datasets? We to put the input both the input and features along with the variables targets. The

ISBN: 978-81-965283-7-9

learning of neural networks can help to make a prediction that based on the cost of errors that comes from the difference between with predicted and the actual target. This is called as a backpropagation. Algorithms of deep learning is like a neural network are uses of many supervised tasks as images, classification, recognition, sentiment analysis, languages, translation, etc.

Unsupervised Machine Learning

Unsupervised machine learning is the techniques of machine learning. This network is uses the machine learning algorithm to discover the patterns or cluster to the unlabelled datasets. This network has no target variable in machine learning. The unsupervised machine learning has to self-determined the hidden patterns and relationship within the datasets. Deep learning algorithms are like to encode the generative models to use for an unsupervised machine tasks like a clustering, dimension able, reduction and anomaly detection.

Reinforcement Machine Learning

Reinforcement Machine Learning is the technique of machine learning, in which learn about the taking suitable decision to maximum a reward in a particular situation. This can interact the agents with a taking actions and observing the resulting rewards. Deep learning can also be used to learning policies to behave in a environment by performing their actions to seeing the results of rewards over a time. The algorithms of reinforcement learning like an Q networks and deep deterministic policy gradient are used in reinforcement learning tasks like an game playing, robotics, etc.

1.4 Types of neural networks

Deep learning is able to learn the automatically features from the data networks, which can be well-suited for tasks such as image recognition, speech recognition and natural language processing. Some of the different types of learning in neural networks are feedforward neural network, convolutional neural network and recurrent neural network.

1.4.1 Feed forward Neural Networks

Feedforward neural network is the one of the type of artificial neural network, this network is characterized by the one or two flow of information network begin with the layers. This network has been widely used for tasks such as image classification, speech recognition and natural language processing.

1.4.2 Recurrent Neural Networks (RNNs)

Recurrent neural network is a type of neural network, that is able to process the sequential data or time series data. Recurrent neural network is maintained the process of where the output information is taken from previous inputs. Which is mostly like the artificial neural network. It is used in speech recognition, video recognition and natural language processing.

1.5 Applications of Deep Learning

The main applications of deep learning can be divided into computer vision, natural language processing (NLP), and reinforcement learning.

In computer vision, Deep learning models can enable machines to identify and understand visual data. Some of the main applications of deep learning in computer vision include:

1.5.1 Object detection and recognition:

Object recognition and detection are used to the techniques of identify and classifying the objects with the images and videos. It is possible for machines to perform their tasks as self-drive cars, surveillance and robotics.

Image classification: Image classification can be used to classify the image into different categories such as animals, plants and buildings, etc. This application is used in medical imaging, quality control and image retrieval.

Image segmentation: Image segmentation is used to involves the image can be converted into an image to collection of regions of pixels that can be represented by the masks or labelled images.

Natural language processing (NLP):

In NLP, the Deep learning model can enable machines to understand and generate human language. Some of the main applications of deep learning in NLP include.

Automatic Text Generation – Deep learning model can learn the corpus of text and new text like summaries, essays can be automatically generated using these trained models.

Language translation: Deep learning models can translate text from one language to another, making it possible to communicate with people from different linguistic backgrounds.

Sentiment analysis: Deep learning models can analyse the sentiment of a piece of text, making it possible to determine whether the text is positive, negative, or neutral. This is used in applications such as customer service, social media monitoring, and political analysis.

Speech recognition: Deep learning models can recognize and transcribe spoken words, making it possible to perform tasks such as speech-to-text conversion, voice search, and voice-controlled devices.

Automatic Text Generation – Deep learning model can learn the corpus of text and new text like summaries, essays can be automatically generated using these trained models.

Language translation: Deep learning models can translate text from one language to another, making it possible to communicate with people from different linguistic backgrounds.

Sentiment analysis: Deep learning models can analyse the sentiment of a piece of text, making it possible to determine whether the text is positive, negative, or neutral. This is used in applications such as customer service, social media monitoring, and political analysis.

Speech recognition: Deep learning models can recognize and transcribe spoken words, making it possible to perform tasks such as speech-to-text conversion, voice search, and voice-controlled devices.

Reinforcement learning

In reinforcement learning, deep learning works as training agents to take action in an environment to maximize a reward. Some of the main applications of deep learning in reinforcement learning include:

Example

Game playing: Deep reinforcement learning models have been able to beat human experts at games such as Go, Chess, and Atari.

Robotics: Deep reinforcement learning models can be used to train robots to perform complex tasks such as grasping objects, navigation, and manipulation.
Control systems: Deep reinforcement learning models can be used to control complex systems such as power grids, traffic management, and supply chain optimization.

1.6 Deep Learning at Work

Deep learning applications are used in industries from automated driving to medical devices.

Automated Driving: Automotive researchers are using deep learning to automatically detect objects such as stop signs and traffic lights. In addition, deep learning is used to detect pedestrians, which helps decrease accidents.

Aerospace and Defence: Deep learning is used to identify objects from satellites that locate areas of interest, and identify safe or unsafe zones for troops.

Medical Research: Cancer researchers are using deep learning to automatically detect cancer cells. Teams at UCLA built an advanced microscope that yields a high-dimensional data set used to train a deep learning application to accurately identify cancer cells.

Industrial Automation: Deep learning is helping to improve worker safety around heavy machinery by automatically detecting when people or objects are within an unsafe distance of machines.

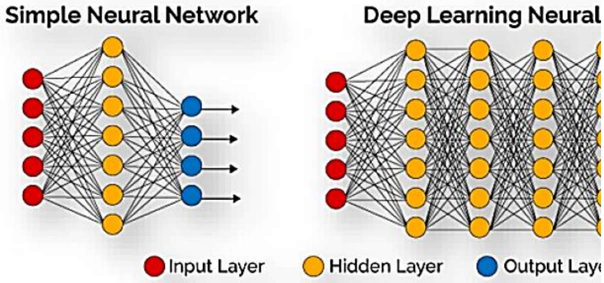
Electronics: Deep learning is being used in automated hearing and speech translation. For example, home assistance devices that respond to your voice and know your preferences are powered by deep learning applications.

How Deep Learning Works

Most deep learning methods use neural network architectures, which is why deep learning models are often referred to as deep neural networks.

The term "deep" usually refers to the number of hidden layers in the neural network. Traditional neural networks (4:37) only contain 2-3 hidden layers, while deep networks can have as many as 150.

Deep learning models are trained by using large sets of labelled data and neural network architectures that learn features directly from the data without the need for manual feature extraction.



Neural networks, which are organized in layers consisting of a set of interconnected nodes. Networks can have tens or hundreds of hidden layers.

One of the most popular types of deep neural networks is known as convolutional neural networks (CNN or ConvNet). A CNN convolves learned features with input data, and uses 2D convolutional layers, making this architecture well suited to processing 2D data, such as images.

CNNs eliminate the need for manual feature extraction, so you do not need to identify features used to classify images. The CNN works by extracting features directly from images. The relevant features are not pertained; they are learned while the network trains on a collection of images. This automated feature extraction makes deep learning models highly accurate for computer vision tasks such as object classification.

1.7 Difference between Machine Learning and Deep Learning

Deep learning is a specialized form of machine learning. A machine learning workflow starts with relevant features being manually extracted from images. The features are then used to create a model that categorizes the objects in the image. With a deep learning workflow, relevant features are automatically extracted from images. In addition, deep learning performs “end-to-end learning” – where a network is given raw data and a task to perform, such as classification, and it learns how to do this automatically.

Another key difference is deep learning algorithms scale with data, whereas shallow learning converges. Shallow learning refers to machine learning methods that plateau at a certain level of performance when you add more examples and training data to the network.

A key advantage of deep learning networks is that they often continue to improve as the size of your data increases.

1.8 Applications of Deep Learning

There are several applications of deep learning across industries. Here, we will discuss some of them in detail.

Self-Driving Cars

Self-driving cars are one of the hottest areas of study and business for the tech demons and deep learning is the power that is rejuvenating self-ruling driving.

Read the blog on How Tesla is making use of Artificial Intelligence in its operations?

Several sets of data are grazed to the system to assemble a model, to prepare the machines to learn, and afterwards test the outcomes in a protected climate. At Pittsburgh, there is an Uber Artificial Intelligence Labs which isn't just operating on preparing driverless vehicles humdrum but also incorporating many creative things. For example, they are trying to make food delivery possible by driverless vehicles. Data from cameras, geo-mapping, sensors are assisting in creating brief and sophisticated models to guide through traffic, identify ways, and real-time components such as traffic volume and road stoppages. Also, the significant worry for self-driving car developers is dealing with uncommon situations. A customary pattern of testing and execution regular to deep learning algorithms is guaranteeing safe driving with increasingly more openness to a great many situations. So, we can assume that in the future deep learning will definitely give many more such intelligent technologies.

Virtual Assistants

Virtual Assistants are one of the very popular applications of deep learning. We all use virtual assistants like Alexa, Siri, search Google Assistant in our day to day life. Every communication with these assistants provides them with a chance to study your voice and emphasize, consequently giving you an optional human interaction experience. Thus, virtual assistants use deep learning to find out more about their subjects going from you dine-out inclinations to your most visited spots or your main tunes. They figure out how to comprehend your orders by assessing common human language to execute them.

Virtual assistants are actually available at your beck-and-call as they can do everything from getting things done to auto-reacting to your particular calls to planning assignments among you and your colleagues. Another power virtual assistants are invested in is interpreting your speech to message, make notes for you, and book arrangements. With deep learning applications like text generation and record synopses, virtual assistants can help you in making or sending proper email duplicates also.

Virtual Recognition

We always face a problem when we have plenty of old pictures but we want some selected images. So, here we faced problems and wasted time and energy in selecting pictures.

So, here comes deep learning that helps in arranging pictures dependent on areas identified in photos, faces, a mix of individuals, or as per occasions, dates, and so forth. Looking for a specific photograph from a library (suppose a dataset as extensive as search Google's image library) requires state-of-the-art visual recognition systems consisting of a few layers from fundamental to advance to recognize elements. Large scale image visual recognition through the deep neural networks are improving development in this section of advanced media management by using convolutional neural networks, Tensor flow, and Python broadly.

Natural Language Processing

NLP is described as the natural manipulation of normal languages, like speech and text, by software and it is the one that aids in perfect communication between human language and computer language.

The area of natural language processing is one of the most crucial and practical applications of deep learning. Through Deep Learning, NLP is attempting to accomplish something very similar by preparing machines to get linguistic nuances and frame suitable reactions. (Related blog- NLP Trends in 2021)

Dealing with questions, classifying text, Twitter analysis, or sentiment analysis at a more extensive level are altogether subsets of NLP where deep learning is acquiring propulsion. Entertainment (Netflix, sports highlights, VEVO, etc)

Netflix and Amazon are improving their deep learning capacities to give a customized insight to their viewers by making their personas figuring in show inclinations, time to access, history, and so forth to prescribe shows that are of getting a kick out of the chance to a specific watcher.

Deep Learning AI is changing the filmmaking cycle as cameras figure out how to examine human non-verbal communication to soak up virtual characters.

Content altering and auto-content creation are now a reality because of deep learning and its commitment to face and pattern recognition.

Fraud Detection

This is the another important application in the deep learning, fraud detection application is useful for the banking and financial sector to finding the fraud complaint in digital transaction. Auto encoders in Kerala and a tensor flow are created to detect the credit card fraud and protection the saving money or dollars of coat in the monetary organization. People can easily trust the advanced technology of banks and online transaction in digital world. Fraud prevention and detection are dependent to design the client transaction of credit scores, distinguished the bizarre and anomalies. The various techniques of regression can be classified under the machine learning methods and neural network for fraud detection.

1.9 CONCLUSION

We have to a simple overview of some techniques and algorithms in machine learning. Furthermore, there are more and more techniques apply machine learning as a solution. In the future machine learning will play an important role in our daily life

REFERENCES

1. Deep Learning by Aaron Courville, Ian Goodfellow, Yoshua Bengio
2. Machine Learning for Algorithmic Trading by Stefan Jansen

ID: 14

DATA STREAM CLASSIFICATION CHALLENGES AND TECHNIQUES: A REVIEW

Ms.M.VIDHYA
Research Scholar,
Department of Computer Science,
University of Kerala,
Kerala, India.

ABSTRACT

Online applications produce immense data streams that flow quickly and with amazing velocity. Due to its enormous influence, stream mining of data has recently gained much attention among the researchers community. This paper contributes a detailed study on the Data Stream Classification Techniques that are happened on top of the Data Stream Mining (DSM). Data stream mining is linked to two fundamental topics of computer science: Data Mining and Data Streams. Data mining is the process of developing algorithms that assist computers in recognizing legitimate and valuable patterns and making rapid and intelligent decisions based on empirical-data. Data mining techniques are effective for vast amounts of static data kept in a system. Data Stream Mining is the process of discovering knowledge/pattern from real/near-real-time continuous and fast changing data streams. Mining huge data streams is a once-in-a-lifetime opportunity as well as a difficult endeavor. Data-stream mining has recently garnered the attention of the research community due to the increased availability of streaming data.

KEYWORDS: *Data Stream Mining, Online Data Classification, Class Imbalance, Concept Drift, Feature Selection.*

1. INTRODUCTION

The Internet of Things (IoT) usage and popularity over the last decade, pointing toward productive and exciting new directions for a whole generation of information devices. IoT technology and its integration with big data have been applied extensively across diverse fields such as smart cities, smart healthcare, intelligent warning systems, and disaster management [1]. The 'Big Data,' coined by John Mashey [2] in 1998. It's expanding beyond our imagination, and that has boosted the incredible growth of computational infrastructures across the world to grow along with the data. The "Big Data" demands new architectures and methods for its management (storage and processing) to enable the extraction of knowledge for enhanced insight and decision making. According to the method of computation, big data analysis can be categorized into two major categories - Offline/Batch computing and Online/Stream computing. In Offline processing, a large amount of data is collected from sources and processed transaction periodically. The main advantage of Offline processing is that repeated jobs can be done fast and manages large repeated jobs easily. The disadvantage is difficult to debug and costly. As the name suggests, Online means "live to action" the processing just on time. The online processing system handles transactions in real-time/near real-time and provides the output instantly. Online processing is interactive; it avoids delay, and users can access data randomly. Autonomous cars [3] are one of the exciting application areas where the hardware, sensors, and algorithms work together efficiently and faster than the human in driving. Real-time and near-real-time applications deal with data that is generated every second and represents an unending stream of events that grows continually and rapidly over time.

2. OVERVIEW OF DATA STREAM CLASSIFICATION

2.1 Definition of Data stream:

A data stream is a potential infinite sequence of data items, which are continuously arrive at a rapid speed. Mathematically, let us define a data stream as D = {(xi, yi)}∞ 1. Here, pair (x_i, y_i) is a data item arrived at a time stamp t, x_i is a n-dimensional feature vector and y = {c1, c2,..., ck} is the ground truth, i.e., a class label (if available) associated with the data item.

In the context of supervised online

classification, we consider a stream of input examples (x) and its corresponding class label (y). These examples arrive one by one from an unknown probability distribution P(x,y). Our goal is to build an online classifier, named M, that can quickly adapt and make predictions as new examples come in. When M makes a prediction (y_i) for an input vector (x_i) at a particular time step (t), it doesn't know the true class label (y_i) at that moment. However, as time progresses, the true labels become available, and we can compare the predictions with the actual values to evaluate the performance of our classifier. The process is repeated with new examples arriving over time. By iteratively updating the classifier with new data, it can continuously learn and adapt to changing patterns in the data stream. It's important to mention that the arrival of new training examples in the data stream is not necessarily uniform or predefined. The time intervals between consecutive time steps may vary, introducing an element of uncertainty and challenge to the online classification process. By continuously refining the classifier with up-to-date feedback, we can maintain its relevance and effectiveness in dynamic environments.

3. DATA STREAM CLASSIFICATION CHALLENGES

3.1 Concept drift

This ability to learn on the fly is especially valuable in scenarios where the underlying data distribution may shift over time, referred to as concept drift.

A drift in the data can occur at any time by switching from one notion to another. The sensor replaced by another sensor with a different calibration in a chemical plant is an example of such a situation. According to the nature of the data, changes could be sudden, recurrent, incremental and gradual. Climate data, for example, may indicate recurring themes as the seasons change, whereas important events may cause a dramatic shift(sudden) in news data. The drift can also occur gradually by incorporating multiple intermediary concepts; the performance degradation of sensors and loss in accuracy is an example of this category.

Alencar et al.,[4] Author suggesting a new approach called Fog-DeepStream that brings together three important fields: Signal Processing (dealing with data), Concept Drift (tracking changes in behavior), and Deep Neural Networks (a type of advanced computer learning). Author start by using something called Wavelet Transform to carefully pick out key information from these streams of data, sort of like summarizing the important parts. Then, we use algorithms that can detect when there's a change in how things are behaving. This helps us decide when to send data and when to hold back, so we're not always flooding the network. And finally, we use Deep Neural Networks to make sense of the data we've selected and saved.

Coelho et al., [5] proposed a method to detect these changes in the data stream. This approach is different from the existing methods that people have used before. Instead of looking directly at the data, we're looking at the space where this data exists. We assume that this space remains the same unless something changes in the way data from different categories show up. imagine the data is spread out in a special structure, kind of like mapping it onto a grid. This structure helps us understand which category is most common in different parts of this grid. When we see that data from one category starts showing up where we'd expect data from another category, we know something's changed – like a sudden shift in how things are happening.

Li et al.,[6] come up with a new way to learn from data as it comes in. The proposed method is called strAEm++DD. It's like a system that combines two good things: one that helps us learn as new data arrives, and another that watches out for changes in the data that might affect our learning. We tested our method on both real-world and made-up data where there's a big difference in how much we have of different categories.

3.2 Class Imbalance

One of the challenges that arise in Stream data analysis is class imbalance. However, some classes of data are significantly underrepresented compared to others. For instance, in applications like Fraud Detection, Anomaly Detection, Outlier Detection, Intrusion Detection, Spam Detection, SaaS subscription churn Prediction, Medical diagnosis. Misclassifying a rare but important event can have serious consequences. That's why addressing class imbalance is crucial for building fair and accurate machine learning models. By understanding and addressing class imbalance, we can improve the performance and fairness of machine learning models, ensuring that they provide reliable predictions across all classes, even in the face of imbalanced data streams.

Liu et al.,[7] To tackle class imbalance issues, we've come up with a new approach. It's called the Transfer Learning Classifier, or TLC for short. This model uses transfer learning, which means it borrows knowledge from one task and applies it to another. The TLC model has three main parts: Active Sampling Module, Real-Time Data Augmentation Module, DenseNet Module. On top of this model, author built a system that can store and analyze real-time data from start to finish.

Mini et al., [8] author proposed approach has three steps: First, we split the data into smaller parts. Then, we use a special technique that combines SMOTE with another method to fix the imbalance problem in each part. For each of these parts, we build a decision tree model to help classify the data. Finally, all these decision trees are put together to make a complete classification model. We tested our approach on lots of different datasets, some of which had up to 4 million instances, and found that our new method works well. It's not only scalable (meaning it works with very big datasets) but also speeds up the processing time.

Sadeghi et al.,[9] author created something called the Dynamic Queues (DynaQ) algorithm. This algorithm helps the computer learn from this tricky, changing data with different categories and varying amounts of data.Our approach uses a smart way of choosing which examples to teach the computer with. We make sure to balance things out by using different techniques, like creating queues for each category and being careful to not include old examples as the computer learns. We also have a way to focus more on the rare categories, like those pandas we mentioned earlier. Our method combines different ways of making decisions and checking if the computer's learning is still on track.

3.3 Feature Selection

Due to the high dimensionality of real-time scenarios, many researchers focus on feature selection based on data stream classification. Feature selection aims to reduce the computation requirements, and it must be helpful to improve computation time. Streaming feature selection can be divided into two types: single feature selection and group feature selection. In 2003, Perkins and Theiler et al.,[10] proposed single feature selection over the streaming data. Xindong Wu et al.,[11] presented single and group feature selection named GFSSF over the incoming streaming data.

Fatma et al.,[12] proposed an Online Feature Selection System(OFS) named Multi-objective Automated Negotiation based Online Feature Selection (MOANOFS) contains two selection levels. In the first level, the confident OFSNeg is selected using the Trust model (so the selection is between agents). In the second level, relevant features are selected using our negotiation method MANOFS (so the selection is between features). Experimental analysis shows that MOANOFS performs better than existing algorithms.

Liu et al.,[13] drawing inspiration from a concept called "neighborhood rough set," which doesn't need us to know everything about the features beforehand, we've come up with a new way to pick out important features from these data streams that have multiple labels. We call it OFS-Mean. This method figures out which neighboring pieces of data are relevant automatically, without needing any preset ideas about the space or the specifics.

Eskandari et al.,[14]Author defined a part of the process as a kind of "yes or no" puzzle, which helps us figure out which features are the most important. To solve this puzzle, we're using something called a binary bat algorithm (BBA), which is pretty good at finding the smallest groups of important features. This new method lets us look at many different combinations of features in an efficient way. Plus, we're also keeping a list of features that we've removed before, so we can check them again if needed. This lets us consider how different features are connected, even if they didn't appear in the data together in a short time span.

4. APPROACHES FOR HANDLING DATA STREAM CLASSIFICATION

4.1 Single/One Pass Approach

A single pass over the data is referred to as a one pass. A one-pass method takes a list of data elements $x_1, x_2, x_3, \ldots x_n$ as input. It can only read these data items once, in the order of the indices from left to right. The size of the memory utilized by the technique and the processing time per data item x_i are critical parameters of a one-pass algorithm. To do this, the method creates and maintains a suitable data structure that is updated for each i upon reading data item x_i.

Wu et al., [15] Many classification methods struggle to properly combine different sets of features for accurate results. To address this issue, we've created a unique setup with two hidden layers: the first layer takes inputs directly, and the second layer connects directly to it. We've also developed a method that involves continuously adjusting the importance of different features and fine-tuning the final output. This method capitalizes on the benefits of our setup.By putting together this specific network structure and our training approach, author come up with an innovative system called ELM-W.

4.2 Ensemble Classification Approach

Ensemble is one of the stream classification methods, and it's a meta-algorithms that associated several machine learning methods with one. In a data stream environment, ensemble classifiers are better than a single classifier but its's time consuming. They improved prediction accuracy by combining more than one classifier using an ensemble approach, despite consuming a lot of memory. Ensemble approaches for streaming data are becoming more popular as a way to improve prediction accuracy.

Wu et al.,[16] proposed a new approach called WENNML. This approach focuses on multi-label data streams. We divide the data into chunks and use them to train two types of classifiers - Active candidate Ensemble Classifiers (AEC) and Passive candidate Ensemble Classifiers (PEC). These classifiers learn from the data in these chunks and get updated as things change. We take into account how different they are and how well they perform when deciding which ones to use.

4.3 Learning-Based Approach

Xie et al.,[17] proposed a clever approach called Energy-based Active Domain Adaptation (EADA). It chooses groups of new data samples that capture the characteristics of the new situation while considering how certain or uncertain each sample is. Also, we've figured out how to make the data from the new situation behave more like the data from the original situation by adjusting something called free energy. This helps close the gap between the two types of data.

5. EMERGING TRENDS IN DATA STREAM CLASSIFICATION

1. Incremental and Online Learning: Traditional machine learning algorithms often assume static datasets, but in the real world, data streams change over time. Researchers have been focusing on developing algorithms that can learn and adapt from incoming data without retraining from scratch. Incremental and online learning methods, such as online ensemble methods and adaptive algorithms, have gained attention to keep models up-to-date with changing data.
2. Concept Drift Detection and Adaptation: Data streams can exhibit concept drift, where the underlying patterns or relationships change over time. Detecting and adapting to concept drift is crucial for maintaining model accuracy. Researchers have been working on methods to monitor drift and adjust models accordingly, ensuring that they remain effective as the data evolves.
3. Imbalanced Data Handling: Many real-world applications involve imbalanced datasets, where one class significantly outnumbers the others. Handling class imbalance in data streams is a challenge, as traditional methods might not be suitable due to the changing nature of the data. Research has been directed towards developing techniques like dynamic resampling and cost-sensitive learning to address this issue.
4. Ensemble Learning and Model Fusion : Ensemble methods, which combine predictions from multiple models, have proven effective in handling data streams. Researchers are exploring ways to combine diverse models, adaptively weight their contributions, and efficiently update ensemble members as new data arrives.
5. Deep Learning for Data Streams: While deep learning has predominantly been applied to static data, efforts are being made to extend these techniques to data streams. Research in this area aims to design deep learning architectures that can adapt and learn from evolving data, while also managing resource constraints in real-time scenarios.
6. Stream Data Visualization: Visualizing data streams can provide insights into trends, patterns, and anomalies. Researchers have been working on

interactive visualization techniques that allow users to explore and understand the evolving data, making it easier to detect meaningful changes.
7. Scalability and Efficiency: As data streams can produce massive amounts of information, there's a focus on developing algorithms that are computationally efficient and scalable. This includes methods to reduce redundant computations, optimize memory usage, and leverage parallel processing techniques.
8. Domain-Specific Applications: The developments in data stream classification are being tailored to various domains, such as finance, healthcare, environmental monitoring, and more. Customized approaches that consider the specific characteristics of each domain are gaining importance.

5.1 Research Gaps and Opportunities in data stream classification

1. Efficient Ensemble Learning: Further advancements in ensemble methods are needed to create ensemble classifiers that can dynamically adapt, update, and optimize the combination of base classifiers to handle changing data streams.
2. Real-time Model Evaluation and Monitoring: Developing techniques to evaluate and monitor the performance of classification models in real-time is critical. This includes methods to measure accuracy, concept drift detection, and model degradation over time.
3. Online Deep Learning: Exploring ways to effectively integrate deep learning approaches into online and incremental learning scenarios is an exciting research opportunity. Developing architectures that can learn from streaming data while managing computation and memory resources is a challenge worth investigating.
4. Semi-Supervised and Unsupervised Approaches: While much of the research has focused on supervised learning, there's room to explore semi-supervised and unsupervised approaches for data stream classification. These methods could leverage unlabeled data to enhance model performance.
5. Multi-Modal and Heterogeneous Streams: Many real-world scenarios involve streams with diverse data types (e.g., text, images, sensor readings). Developing methods to effectively classify heterogeneous data streams and combine information from multiple modalities is a promising research avenue.

6. CONCLUSION

In conclusion, the field of data stream classification has witnessed remarkable advancements and continues to hold immense potential for addressing real-world challenges. Researchers have made significant strides in developing methods that can handle dynamic, evolving data distributions. Techniques for detecting and adapting to concept drift have become increasingly sophisticated, enabling models to maintain accuracy as data changes over time.The journey towards more efficient and effective data stream classification techniques is characterized by the ongoing pursuit of scalability, energy efficiency, and domain-specific solutions. As technologies and industries continue to evolve, the evolution of data stream classification methods will undoubtedly play a pivotal role in enabling real-time insights, informed decision-making, and improved performance across diverse applications.

REFERENCE:

1. Bibri, S. E. (2018). The IoT for smart sustainable cities of the future: An analytical framework for sensor-based big data applications for environmental sustainability. *Sustainable cities and society*, *38*, 230-253.
2. Diebold, F. X. (2012). On the Origin (s) and Development of the Term'Big Data'.
3. Hussain, R., & Zeadally, S. (2018). Autonomous cars: Research results, issues, and future challenges. *IEEE Communications Surveys & Tutorials*, *21*(2), 1275-1313.
4. Alencar, B. M., Canário, J. P., Lobão Neto, R., Prazeres, C., Bifet, A., & Rios, R. A. (2023). Fog-DeepStream: A new approach combining LSTM and Concept Drift for data stream analytics on Fog computing. *Internet of Things*, *22*, 100731. https://doi.org/10.1016/j.iot.2023.100731
5. Amador Coelho, R., Bambirra Torres, L. C., & Leite de Castro, C. (2023). Concept drift detection with quadtree-based spatial mapping of streaming data. *Information Sciences*, *625*, 578-592.

https://doi.org/10.1016/j.ins.2022.12.085

6. Li, J., Malialis, K., & Polycarpou, M. M. (2023). Autoencoder-based Anomaly Detection in Streaming Data with Incremental Learning and Concept Drift Adaptation. *arXiv preprint arXiv:2305.08977.*
7. Liu, Y., Yang, G., Qiao, S., Liu, M., Qu, L., Han, N., Wu, T., Yuan, G., Wu, T., & Peng, Y. (2023). Imbalanced data classification: Using transfer learning and active sampling. Engineering Applications of Artificial Intelligence, 117, 105621. https://doi.org/10.1016/j.engappai.2022.105621
8. Liu, Y., Yang, G., Qiao, S., Liu, M., Qu, L., Han, N., ... & Peng, Y. (2023). Imbalanced data classification: Using transfer learning and active sampling. Engin*eering Applications of Artificial Intelligence*, *117*, 105621.
9. Sadeghi, F., Viktor, H. L., & Vafaie, P. (2023). DynaQ: online learning from imbalanced multi-class streams through dynamic sampling. *Applied Intelligence*, 1-23.
10. Perkins, S., Lacker, K., & Theiler, J. (2003). Grafting: Fast, incremental feature selection by gradient descent in function space. *The Journal of Machine Learning Research*, *3*, 1333-1356.
11. Wang, J., Wang, M., Li, P., Liu, L., Zhao, Z., Hu, X., & Wu, X. (2015). Online feature selection with group structure analysis. *IEEE Transactions on Knowledge and Data Engineering*, *27*(11), 3029-3041.
12. BenSaid, F., & Alimi, A. M. (2021). Online feature selection system for big data classification based on multi-objective automated negotiation. Pattern Recognition, 110, 107629. https://doi.org/10.1016/j.patcog.2020.107629
13. Liu, J., Lin, Y., Du, J. et al. ASFS: A novel streaming feature selection for multi-label data based on neighborhood rough set. Appl Intell 53, 1707–1724 (2023). https://doi.org/10.1007/s10489-022-03366-x
14. Eskandari, S., & Seifaddini, M. (2023). Online and offline streaming feature selection methods with bat algorithm for redundancy analysis. Pattern Recognition, 133, 109007. https://doi.org/10.1016/j.patcog.2022.109007
15. Wu, C., Sang, Y., & Gao, Y. (2023). Extreme Learning Machine Combining Hidden-Layer Feature Weighting and Batch Training for Classification. *Neural Processing Letters*, 1-23.
16. Wu, H., Han, M., Chen, Z., Li, M., & Zhang, X. (2023). A Weighted Ensemble Classification Algorithm Based on Nearest Neighbors for Multi-Label Data Stream. *ACM Transactions on Knowledge Discovery from Data*, *17*(5), 1-21.
17. Xie, B., Yuan, L., Li, S., Liu, C. H., Cheng, X., & Wang, G. (2022, June). Active learning for domain adaptation: An energy-based approach. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 36, No. 8, pp. 8708-8716).

ID: 15

AUGMENTED REALITY APPLICATION FOR INDUSTRIALS/INSTITUTIONS ENGINEERING LAB TRAINING INCORPORATED DEEP LEARNING AND MACHINE LEARNING USING AUGMENTED REALITY TOOLKIT

Dr.U.SRIDEVI
Assistant Professor,
Department of MCA,
Excel Engineering College,
Namakkal, Tamil Nadu, India.

Mr.M.K.NAGESWARAN
Assistant Professor,
Department of MCA,
Excel Engineering College,
Namakkal, Tamil Nadu, India.

ABSTRACT:
Today technological changes make the probability of more complex things made into simple tasks with more accuracy in major areas and mostly in Manufacturing Industry. Internet of things contributes its major part in automation which helps human to make life easy by monitoring and directed to a related person with in a fraction of second. Continuous advances and improvement in computer vision, mobile computing and tablet screens have led to a revived interest in Augmented Reality the Augmented Reality makes the complex automation into an easier task by making more realistic real time animation in monitoring and automation on Internet of Things (eg like temperature, time, object information, installation manual, real time testing).In order to identify and link the augmented content, like object control of home appliances, industrial appliances. The AR-DL will have a much cozier atmosphere and enhance the overall Interactivity of the DL environment. Augmented Reality applications use a myriad of data generated by DL devices and components, AR helps workers become more competitive and productive with the realistic environment in DL. Augmented Reality and Internet of Things together plays a critical role in the development of next generation technologies. This paper describes the concept of how Augmented Reality can be integrated with industry (AR-DL)4.0 and how the sensors are used to monitoring objects/things contiguously round the clock, and make the process of converting real-time physical objects into smart things for the upcoming new era with AR-DL. The proposed framework can be a useful foundation for AR and machine-learning-based frameworks for industrial and educational training.

Keywords: Artificial Intelligence; Augmented Reality; Machine Learning; Object Detection; Computer in Education; Lab Equipment Tutorial.

I. INTRODUCTION:

This work explores the idea of using equipment recognition and an AR-based tutorial to enhance student learning experiences with electrical equipment in their engineering laboratories. Our long-term goal is to develop interactive Smartphone apps for lab equipment such as multimeters, oscilloscopes, wave generators, and power supplies. Object detection using DL methods fits our goal because the app can detect specific electrical equipment in the lab with high precision in real-time using state-of-art DL algorithms. AR technology enables us to create virtual scenarios and integrate 3D models, animations, images, and videos embedded into teaching methods. In the developed app, the interactive visualization was created using the Unity3D game engine. The object detection process of the app employs a deep neural network architecture trained with TensorFlow API.

II. AUGMENTED REALITY

AR allows virtual objects to be overlaid in real-world environments in real time. The below image shows a man using IKEA AR

App to design, improve, and live his dream home. Unlike virtual reality which tries to recreate and replace an entire real-life environment with a virtual one, augmented reality is about enriching an image of the real world with computer-generated images and digital information. It seeks to change perception by adding video, infographics, images, sound, and other details.

Inside a device that creates AR content; virtual 3D images are overlaid on real-world objects based on their geometrical relationship. The device must be able to calculate the position and orientation of objects concerning others. The combined image is projected on mobile screens, AR glasses, etc.

Although the first application was in military and television since the coining of the term in 1990, AR is now applied in gaming, education and training, and other fields. Most of it is applied as AR apps that can be installed on phones and computers. Today, it is enhanced with mobile phone technology such as GPS, 3G and 4G, and remote sensing.

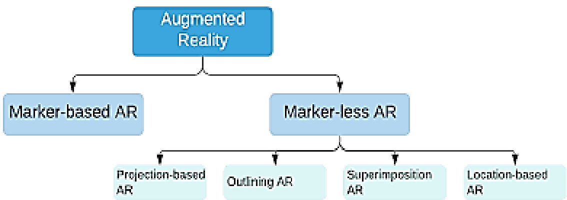


Figure.1. Types of Augmented Reality.

2.1 Marker-based AR

A marker, which is a special visual object like a special sign or anything, and a camera are used to initiate the 3D digital animations. The system will calculate the orientation and position of the market to position the content effectively. For example, a few display virtual information when the device is focused on the marker, while others save that virtual information and allow users to view it again when the device is focused on a different section. The marker-based AR technology leverages images from the actual world or QR codes to extract points, lines, corners, textures, and other properties. These images are used to superimpose and create AR experiences by referencing track points in the physical world.

2.2 Marker-Less AR

It is used in events, business, and navigation apps, **for instance,** the technology uses location-based information to determine what content the user gets or finds in a certain area. It may use GPS, compasses, gyroscopes, and accelerometers as can be used on mobile phones. Marker-less AR collects data from the device hardware such as a camera, a GPS, a digital compass, and an accelerometer for the AR program to function. Marker-less AR applications rely on computer vision algorithms to distinguish objects, and they can function in the real world without specific markers. There are four types of marker-less AR discussed as follows:

(a) *Location-based AR*: In this type of AR, simultaneous localization and mapping (SLAM) technology is used to track the user's location as the map is generated and updated on the user's mobile device. To display AR content in the physical environment, the user must detect a surface with a mobile device. As an example, the world-famous AR-based game app, Pokemon Go, uses SLAM technology that allows its users to battle, navigate, and search for 3D interactive objects based on their geographical locations.

(b) *Superimposition-based AR*: Superimposition-based AR applications can provide an additional view along with the original view of the object. Object recognition is required to determine the type of object to partially or completely replace an object in the user's environment with a digital image. Using HoloLens glasses, surgeons can superimpose images previously gathered through scanners or X-rays on the patient's body during the operation. They can anticipate potential problems using this approach.

(c) *Projection-based AR*: Projection-based AR (also known as projection mapping and augmented spatial reality) is a technique that does not require the use of head-mounted or hand-held devices. This method allows augmented information to be viewed immediately from a natural perspective. Using projection mapping, projection-based AR turns an uneven surface into a projection screen. This method allows for the creation of optical illusions.

(d) *Outlining-based AR*: This type of AR employs image recognition to create contours or forms and highlight components of the real world using special cameras. It is used by human

ISBN: 978-81-965283-7-9

eyes to designate specific items with lines to make situations easier. Vuforia's Model Target is an example of outlining-based AR. Vuforia is a platform that enables developers to quickly incorporate AR technology into their applications. Model Targets allow apps to recognize and track real-world objects based on their shape.

In our project, we built a superimposition-based AR app. We built user interfaces on top of lab equipment, allowing step-by-step instructions to be incorporated into the application for users to understand and learn how to use specific equipment. Using AR technology, immersive experiences are created in a variety of ways. It does, however, have some limitations such as the inability to recognize multiple objects at once. On the other hand, DL models show high performances in recognizing multiple objects at the same time. Integrating AR apps with DL models will help trigger specific AR scenarios based on objects being aimed at with a camera and allow an AR scenario to perform a single tracking without decreasing mobile device performance.

III. DESIGN AND IMPLEMENTATION OF THE AR APP

This section describes the design and development of the AR app that integrates two independent frameworks for object detection and augmented reality as shown in Figure 2. Unity 3D combines the output of these systems by inferring the object detection model with OpenCV and using an AR dataset target with a Vuforia Engine. Furthermore, Unity 3D enables the development of interactive user interfaces. Users can first use their mobile device to infer the object detection model to detect the lab equipment. The inference will classify and localize lab equipment that has been targeted with the mobile camera. When an object is detected, a user interface (UI) button appears, indicating that an AR-guided tutorial is available for the object. Then, an AR scenario will be loaded, allowing students to use their mobile camera to aim at a specific target. Following that, a 3D object will superimpose on top of the physical object, activating UI panels with instructions on how to use the equipment.

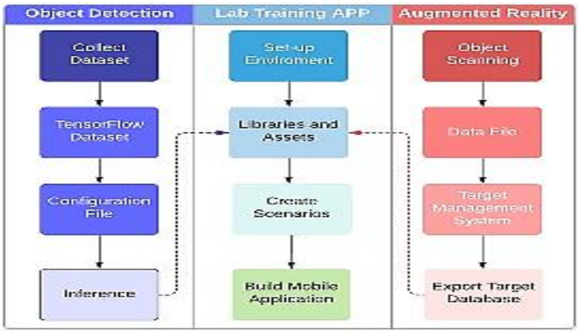


Figure 2. Design framework of AR-based Smartphone app for lab equipment training.

The app development process consists of integrating a number of different independent systems with their frameworks. In the interactive tutorial development framework, Unity3D was used as the primary development software for generating specific UI instructions and creating immersive interactions between the mobile app and the user. The development framework was integrated with a MobileNet-SSD DL model and a marker-less superimposition AR that activates immersive modules containing 2D/3D objects. The detailed framework integration is discussed below.

3.1 Object Detection Framework

MobileNet-SSDv2 architecture was used to build a deep neural network model to detect electrical lab equipment. The architecture comprised MobileNet-v2 as the backbone network, an SSD detector, and feature pyramid network (FPN). MobileNet, as the name implies, is intended for embedded applications on mobile devices to improve accuracy while effectively considering constrained resources. A loss function method was calculated using the predicted and labeled values of the classes and offsets to evaluate how the algorithm models the data. The confidence loss ($Lconfidence$) occurs when attempting to predict a class, which is softmax loss over multiple classes confidences. Localization loss ($Llocalization$) is defined as a mismatch between the ground truth and the intended boundary boxes, where α is the weight coefficient, expressed as:

$$Lloss = Lconfidence + \alpha \times Llocalization \quad (1)$$

MobileNet is a low-latency and low-power model that can be tailored to meet the resource constraints of various use cases. For multi-scale object detection, MobileNetv2 provides a number of feature maps with different dimensions for the backbone detection network to the SSD convolutional layer that uses small convolutional filters to predict scores and class offsets for a fixed set of the standard bounding boxes. MobileNet-SSDv2 extracts features from images, which are then processed through SSD predictor layers that reduce image size to recognize objects at various scales as shown in Figure 3. Mobilenet-SSDv2 detector improves the SSD detector by combining MobileNetv2 and FPN while maintaining memory efficiency.

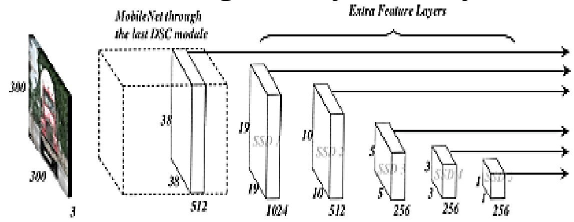


Figure 3. MobileNet SSD Deep Neural Network Architecture.

3.2 TensorFlow Object Detection API

TensorFlow (TF) API, developed by Google Brain, is a framework for creating a DL network. It is a powerful tool that can be used to create a robust object detection framework with a set of standard functions, eliminating the need for users to write code from scratch. It also provides a list of pre-trained models, which are useful not only for inference but also for building models with new data. Model Zoo is a collection of models that have been previously trained using the common objects in context (COCO) dataset. A workflow for training a DL model using the TF API is shown in Figure 5 and can be described through the following steps:

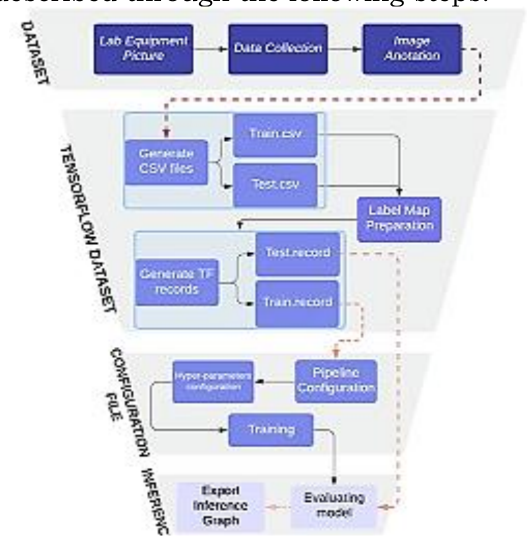


Figure 4. Object Detection Workflow using TensorFlow API.

(a) *Image Dataset*: The model was given input of 643 images collected from various perspective views and in different lighting settings. Each image is of 4032 × 3024 pixels in size. It is necessary to annotate these images before using them to train the model. A software, LabelImg, is used in the annotation process that allows users to draw a rectangle in a specific area of the image. During training, the annotation will help the model precisely locate the object in the image. The outlining will generate and save coordinate points in an XML file.

(b) *TensorFlow Dataset*: To make the computation of the DL framework efficient, TF records use a binary file format. Furthermore, TF records enable the dataset to be stored as a sequence of binary strings that improves the model's performance while using less disk space. We converted the XML files generated by LabelImg into TF binary records using a Python script. The last step in configuring the TF dataset is to create a .pbtxt file containing all of the categorical label classes that will be stored in a TF record file.

(c) *Configuration File*: Multiple pre-trained models based on the common objects in context (COCO) dataset are available in TF. These models can be used to set up DL models prior to training on a new dataset. Table 1 lists several popular architectures with pre-trained models. For instance, *ssd_MobileNet_v1_coco* is the SSD with a MobileNet v1 configuration, *ssd_inception_v2_coco* re presents an SSD with an Inception v2 configuration,
and *faster_rcnn_resnet101_coco* stands for Faster R-CNN with a Resnet-101 (v1) configuration. All these configurations have been derived for the COCO dataset. From Table 1, it can be observed that *ssd_MobileNet_v1_coco* reaches the fastest inference speed of 30 ms but with the lowest mean average precision (mAP).
In contrast, *faster_rcnn_resnet101_coco* has the slowest inference speed but the highest mAP of 32.

3.3 LAB TRAINING APPLICATION FRAMEWORK

With the help of AR and DL, the equipment learning application focuses on teaching and improving the student's learning experience on how to properly use electrical equipment. Unity3D will provide libraries that allow these technologies to be combined on top of assets, animations, and 3D models to create training scenarios that will engage students in learning through experience. The development procedure is shown in **Figure 5** and can be described in the following steps:

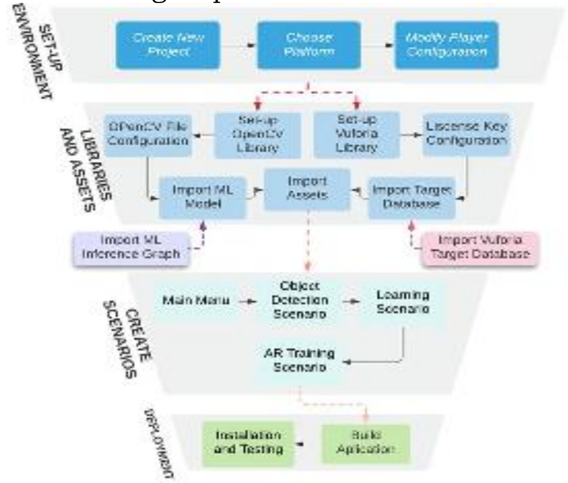


Figure 5. Lab Training Application Framework

(a) Setup Environment:

The setup starts with the creation of a new project using a Unity hub. After creating and opening the project, it is essential to switch to a different build platform because Unity allows us to create once and deploy anywhere. In other words, we can select a platform from the list of available platforms in Unity, such as Windows, WebGL, Android, iOS, or any gaming console. We chose Android as the deployed platform for this project. The platform can be changed in the build settings windows, which can be accessed via the file bar. Additionally, the rendering, scripting, and project configuration must be modified.

(b) Libraries and Assets:

(1) OpenCV Library: OpenCV For Unity is a program that uses AI algorithms to analyze and interpret images on computers or mobile devices. This Unity asset store product allows users to test AI pre-trained models that can be used to run algorithms and executable applications on mobile devices. The model employs a script that requires a binary file of a DL model with trained weights (weights of deep neural networks are not modified in this stage), and a file model network configuration. This script is granted access to the device resource, specifically the camera, so that the script can pass input to the model and start object detection inference, which will generate bounding boxes and labels around the object detected.

(2) Vuforia Engine: This library allows Unity to create AR experiences for mobile devices. It is a collection of scripts and pre-made components for developing Vuforia apps in Unity. It includes API libraries written in the C# language that expose the Vuforia APIs. This library supports all traceable functions as well as high-level access to device hardware such as the device camera.

(3) Assets: They are graphical representations of any items that could be used in the project. It is made up of user interface sprites, 3D models, images, materials, and sounds, all with their own design and functionality. Photoshop is used to create art sprites, such as images for a virtual manual and blender. A 3D modeler software is used to create 3D models.

(c) Scenarios creation

1. Main menu: The application includes a menu scenario, that will allow the user to select various modes based on their preferences. It includes a tutorial that teaches students how to use the application. There is a training mode to help students learn more about lab equipment or electrical components.
2. Object detection: In this case, the DL model is used in conjunction with the OpenCV library in Unity. The application has access to the device's camera from which it will infer the object detection model provided by the object detection framework. Furthermore, depending on the object that is being targeted, the application automatically generates bounding boxes around the desired object with its respective label and confidence. When the user points to the desired equipment, a bottom panel will appear with the option to load the AR experience or continue looking for other lab equipment. The OpenCV library allows us to specify the desired confidence value threshold during the model inference. During model inference, we can specify the desired confidence value threshold using the OpenCV library. The model draws a

green rectangle around the detected equipment. The detection threshold confidence value is set to 90%, which means that the confidence must be greater than or equal to 90% to indicate a detection with a rectangular bounding box. This percentage was chosen because the lab equipment is quite different. The score of 90% would ensure that the lab equipment detected had a high confidence level.

3. AR training scenario: When the application detects an object that has previously been configured in Unity, a 3D model will be superimposed on top of the physical object in the mobile app. It will also include a UI for the user to interact with, allowing them to understand and explore the physical object, while the mobile application provides additional information in the form of holograms

4. EXPERIMENTAL RESULTS

In this section, performance for DL and AR frameworks are discussed.

4.1. Object Detection

The dataset used in this study is a collection of 643 images with annotations. The dataset is divided into four classes: multimeter, oscilloscope, power supply, and wave generator. The collected samples were randomly split into training and test sets with the ratio of 70% and 30%, respectively. We employ commonly used evaluation metrics such as precision, recall, and mAP to evaluate the model performance in the application.

4.2. Evaluation Metrics

Precision: The percentage of positive detections that were correct is referred to as precision. If a model produces no false positives, it has a precision of 1.0. Equation **(2)** describes precision as the True Positive divided by the sum of the True Positive (TP) and False Positive (FP). TP is defined as a correct prediction of the positive class, whereas FP is an incorrect prediction of negative class as the positive class.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad (2)$$

Recall: The percentage of true positives that were correctly identified by the model. A model with a recall of 1.0 produces zero false negatives. Recall can be computed as a ratio of True Positives predictions and the sum of TP and False Negatives (FN), as shown in Equation **(3)**. FN is defined as an incorrect prediction of the positive class as the negative class.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad (3)$$

Intersection over Union (IoU): It is also known as the Jaccard index used for measuring the similarity and diversity of sample sets. In an object detection task, it describes the similarity between the predicted bounding box and the ground truth bounding box. Equation (**4**) expresses IoU in terms of area of the prediction and ground truth bounding boxes.

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \qquad (4)$$

It is important to define a threshold to define what is the correctness of the prediction for IoU.

$$\begin{aligned} &\text{T} \leftarrow \text{Threshold} \\ &\text{IoU} \geq \text{T} \rightarrow \text{Correct} \\ &\text{IoU} < \text{T} \rightarrow \text{Incorrect} \end{aligned} \qquad (5)$$

Mean Average Precision (mAP): It takes under consideration both precision and recall. It is also the area beneath the precision–recall curve. The mAP can be computed by

$$\text{mAP} = \frac{\sum_{k=1}^{n} AP_k}{n} \qquad (6)$$

Where *APk* is the average precision of class *k*, and *n* is the number of classes.

It demonstrates the correct detection of four types of lab equipment in a single shot when the confidence threshold value (i.e., the threshold related to the confidence score to determine whether the detection is an object of interest or not. Confidence scores of the predicted bounding boxes above the threshold value are considered as positive boxes, or vice versa) is greater or equal to 90%.

5. CONCLUSIONS AND FUTURE WORK

In this study, we developed an interactive multimeter tutorial using deep learning and augmented reality. We integrated a deep learning model, namely MobileNet-SSD v2, and an AR target database into a game engine to detect objects automatically. Unity3D was used to create the augmented tutorial, which includes a mobile game infrastructure. The tutorial functions as a virtual manual for the equipment, which provides an immersive experience by projecting holograms on objects recognized by the app via a mobile camera. In the future, we will create tutorials for additional lab equipment. One application will be the addition of a 3D interactive breadboard in the app to help students understand electrical circuits. Another potential enhancement of the proposed AR- and AI-based education tool would be to support remote learning, in which students can learn lab equipment through the AR streaming on their mobile devices or personal computers.

ID: 16

DATA MINING TECHNIQUES WITH DATA SECURITY

Dr.M.RAJESHWARI
Assistant Professor,
Department of Computer Science with Cyber Security,
Erode Arts and Science College,
Erode, Tamil Nadu, India.

Dr.N.SHANMUGAPRIYA
Assistant Professor & Head,
Department of Computer Applications,
Erode Arts and Science College,
Erode, Tamil Nadu, India.

ABSTRACT

Data and Information or Knowledge has a significant role on human activities. Data mining is the knowledge discovery process (KDD) by analyzing the large volumes of data from various perspectives and summarizing it into useful information. Due to the importance of extracting knowledge/information from the large data repositories, data mining has become an essential component in various fields of human life. Advancements in Statistics, Machine Learning, Artificial Intelligence, Pattern Recognition and Computation capabilities have evolved the present days data mining applications and these applications have enriched the various fields of human life including business, education, medical, scientific etc. Hence, this paper discusses the various improvements in the field of data mining from past to the present and explores the future trends.

***KEYWORDS**: Data Mining, Knowledge Discovery Data, Machine Learning, Artificial Intelligence, Pattern Recognition*

I. INTRODUCTION

Data mining is a crucial component of successful analytics initiatives in organizations. The information it generates can be used in business intelligence (BI) and advanced analytics applications that involve analysis of historical data, as well as real time analytics applications that examine streaming data as it's created or collected. Effective data mining aids in various aspects of planning business strategies and managing operations. That includes customer-facing functions such as marketing, advertising, sales and customer support, plus manufacturing, supply chain management, finance and HR. Data mining supports fraud detection, risk management, cyber security planning and many other critical business use cases. It also plays an important role in healthcare, government, scientific research, mathematics, sports and more.

II. HISTORY OF DATA MINING

In the **1990s**, the term "Data Mining" was introduced, but data mining is the evolution of a sector with an extensive history. Early techniques of identifying patterns in data include Bayes theorem (**1700s**), and the evolution of regression (**1800s**). The generation and growing power of computer science have boosted data collection, storage, and manipulation as data sets have broad in size and complexity level. Explicit hands-on data investigation has progressively been improved with indirect, automatic data processing, and other computer science discoveries such as neural networks, clustering, genetic algorithms (**1950s**), decision trees (**1960s**), and supporting vector machines (**1990s**). Data mining origins are traced back to three family lines: Classical statistics, Artificial intelligence, and Machine learning.

CLASSICAL STATISTICS:

Statistics are the basis of most technology on which data mining is built, such as regression analysis, standard deviation, standard distribution, standard variance, discriminatory analysis, cluster analysis, and confidence intervals. All of these are used to analyze data and data connection.

ARTIFICIAL INTELLIGENCE:

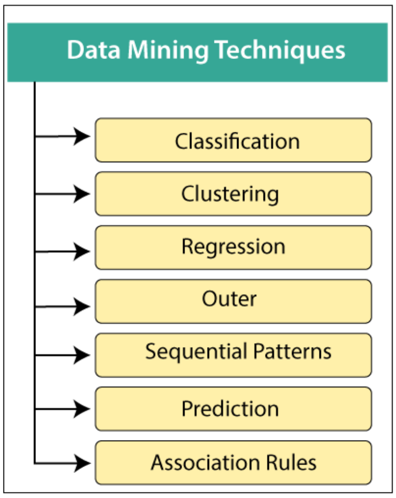
AI or Artificial intelligence is based on heuristics as opposed to statistics. It tries to apply human- thought like processing to statistical problems. A specific AI concept was adopted by some high-end commercial products, such as query optimization modules for **Relational Database Management System (RDBMS)**.

MACHINE LEARNING:

Machine learning is a combination of statistics and AI. It might be considered as an evolution of AI because it mixes AI heuristics with complex statistical analysis. Machine learning tries to enable computer programs to know about the data they are studying so that programs make a distinct decision based on the characteristics of the data examined. It uses statistics for basic concepts and adding more AI heuristics and algorithms to accomplish its target.

III. DATA MINING TECHNIQUES

Data mining includes the utilization of refined data analysis tools to find previously unknown, valid patterns and relationships in huge data sets. These tools can incorporate statistical models, machine learning techniques, and mathematical algorithms, such as neural networks or decision trees. Thus, data mining incorporates analysis and prediction. Depending on various methods and technologies from the intersection of machine learning, database management, and statistics, professionals in data mining have devoted their careers to better understanding how to process and make conclusions from the huge amount of data, but what are the methods they use to make it happen? In recent data mining projects, various major data mining techniques have been developed and used, including association, classification, clustering, prediction, sequential patterns, and regression.



1. CLASSIFICATION:

This technique is used to obtain important and relevant information about data and metadata. This data mining technique helps to classify data in different classes. Data mining techniques can be classified by different criteria, as follows:

1. **Classification of Data mining frameworks as per the type of data sources mined:**
   This classification is as per the type of data handled. For example, multimedia, spatial data, text data, time-series data, World Wide Web, and so on..
2. **Classification of data mining frameworks as per the database involved:**
   This classification based on the data model involved. For example. Object-oriented database, transactional database, relational database, and so on..
3. **Classification of data mining frameworks as per the kind of knowledge discovered:**
   This classification depends on the types of knowledge discovered or data mining functionalities. For example, discrimination, classification, clustering, characterization, etc. some frameworks tend to be extensive frameworks offering a few data mining functionalities together.
4. **Classification of data mining frameworks according to data mining techniques used:**
   This classification is as per the data analysis approach utilized, such as neural networks, machine learning, genetic algorithms, visualization, statistics, data warehouse-oriented or database-oriented, etc.

ISBN: 978-81-965283-7-9

The classification can also take into account, the level of user interaction involved in the data mining procedure, such as query-driven systems, autonomous systems, or interactive exploratory systems.

2. CLUSTERING:

Clustering is a division of information into groups of connected objects. Describing the data by a few clusters mainly loses certain confine details, but accomplishes improvement. It models data by its clusters. Data modeling puts clustering from a historical point of view rooted in statistics, mathematics, and numerical analysis. From a machine learning point of view, clusters relate to hidden patterns, the search for clusters is unsupervised learning, and the subsequent framework represents a data concept. From a practical point of view, clustering plays an extraordinary job in data mining applications. For example, scientific data exploration, text mining, information retrieval, spatial database applications, CRM, Web analysis, computational biology, medical diagnostics, and much more. In other words, we can say that Clustering analysis is a data mining technique to identify similar data. This technique helps to recognize the differences and similarities between the data. Clustering is very similar to the classification, but it involves grouping chunks of data together based on their similarities.

3. REGRESSION:

Regression analysis is the data mining process is used to identify and analyze the relationship between variables because of the presence of the other factor. It is used to define the probability of the specific variable. Regression, primarily a form of planning and modeling. For example, we might use it to project certain costs, depending on other factors such as availability, consumer demand, and competition. Primarily it gives the exact relationship between two or more variables in the given data set.

4. ASSOCIATION RULES:

This data mining technique helps to discover a link between two or more items. It finds a hidden pattern in the data set. Association rules are if-then statements that support to show the probability of interactions between data items within large data sets in different types of databases. Association rule mining has several applications and is commonly used to help sales correlations in data or medical data sets. The way the algorithm works is that you have various data, For example, a list of grocery items that you have been buying for the last six months. It calculates a percentage of items being purchased together.

These are three major measurements technique:

- **Lift:**
  This measurement technique measures the accuracy of the confidence over how often item B is purchased.
  (Confidence) / (item B) / (Entire dataset)
- **Support:**
  This measurement technique measures how often multiple items are purchased and compared it to the overall dataset.
  (Item A + Item B) / (Entire dataset)
- **Confidence:**
  This measurement technique measures how often item B is purchased when item A is purchased as well.
  (Item A + Item B) / (Item A)

5. OUTER DETECTION:

This type of data mining technique relates to the observation of data items in the data set, which do not match an expected pattern or expected behavior. This technique may be used in various domains like intrusion, detection, fraud detection, etc. It is also known as Outlier Analysis or Outilier mining. The outlier is a data point that diverges too much from the rest of the dataset. The majority of the real-world datasets have an outlier. Outlier detection plays a significant role in the data mining field. Outlier detection is valuable in numerous fields like network interruption identification, credit or debit card fraud detection, detecting outlying in wireless sensor network data, etc.

6. SEQUENTIAL PATTERNS:

The sequential pattern is a data mining technique specialized for **evaluating sequential data** to discover sequential patterns. It comprises of finding interesting subsequences in a set of sequences, where the stake of a sequence can be measured in terms of different criteria like length, occurrence frequency, etc. In other words, this technique of data mining helps to discover or recognize similar patterns in transaction data over some time.

7. PREDICTION:

Prediction used a combination of other data mining techniques such as trends, clustering, classification, etc. It analyzes past events or instances in the right sequence to predict a future event

DATA SECURITY

Data security uses tools and technologies that enhance visibility of a company's data and how it is being used. These tools can protect data through processes like data masking, encryption, and redaction of sensitive information. The process also helps organizations streamline their auditing procedures and comply with increasingly stringent data protection regulations. A robust data security management and strategy process enables an organization to protect its information against cyber attacks. It also helps them minimize the risk of human error and insider threats, which continue to be the cause of many data breaches.

BENEFITS OF DATA SECURITY

1. **Keeps your information safe:** By adopting a mindset focused on data security and implementing the right set of tools, you ensure sensitive data does not fall into the wrong hands. Sensitive data can include customer payment information, hospital records, and identification information, to name just a few. With a data security program created to meet the specific needs of your organization, this info stays safe and secure.
2. **Helps keep your reputation clean:** When people do business with your organization, they entrust their sensitive information to you, and a data security strategy enables you to provide the protection they need. Your reward? A stellar reputation among clients, partners, and the business world in general.
3. **Gives you a competitive edge:** In many industries, data breaches are commonplace, so if you can keep data secure, you set yourself apart from the competition, which may be struggling to do the same.
4. **Saves on support and development costs:** If you incorporate data security measures early in the development process, you may not have to spend valuable resources for designing and deploying patches or fixing coding problems down the road.

CONCLUSION

In this Chapter we briefly reviewed the Data Mining Techniques with Data Security. This chapter would be helpful to researchers to focus on the various issues of data mining. Data mining is a very powerful and useful methodology and technology for generating information for decision making. Future developments are expected to make data mining even more powerful and useful. Despite this, data mining is not without limitations.

REFERENCES

1. Heikki, Mannila. 1996. Data mining: Machine Learning, Statistics, and Databases, IEEE
2. Fayadd, U., Piatesky -Shapiro, G., and Smyth, P. 1996. From Data Mining To Knowledge Discovery in Databases, AAAI Press / The MIT Press, Massachusetts Institute of Technology. ISBN 0–262 56097–6 Fayap.
3. Piatetsky-Shapiro, Gregory. 2000. The Data-Mining Industry Coming of Age. IEEE Intelligent Systems.
4. Salmin, Sultana et al. 2009. Ubiquitous Secretary: A Ubiquitous Computing Application Based on Web Services Architecture, International Journal of Multimedia and Ubiquitous Engineering Vol. 4, No. 4, October, 2009
5. Hsu, J. 2002. Data Mining Trends and Developments: The Key Data Mining Technologies and Applications for the 21st Century, The Proceedings of the 19th Annual Conference for Information Systems Educators (ISECON 2002), ISSN: 1542-7382. Available Online: http://colton.byuh.edu/isecon/2002/224b/Hsu.pdf
6. Shonali Krishnaswamy. 2005. Towards Situation-awareness and Ubiquitous Data Mining for Road Safety: Rationale and Architecture for a Compelling Application (2005), Proceedings of Conference on Intelligent Vehicles and Road Infrastructure 2005, pages-16, 17. Available at: http://www.csse.monash.edu.au/~mga

ber/CameraReadyI

7. Kotsiantis, S., Kanellopoulos, D., Pintelas, P. 2004. Multimedia mining. WSEAS Transactions on Systems, No 3, s. 3263-3268.
8. Abdulvahit, Torun., Ebnem, Düzgün. 2006. Using spatial data mining techniques to reveal vulnerability of people and places due to oil transportation and accidents: A case study of Istanbul strait, ISPRS Technical Commission II Symposium, Vienna. Addison Wesley, 1st edition.
9. T. M. Mitchell. 1982. Generalization as Search, Artificial Intelligence, 18(2), 1982, pp.203-226.
10. R. Michalski., I. Mozetic., J. Hong., and N. Lavrac. 1986. The AQ15 Inductive Leaning System: An Overview and Experiments, Reports of Machine Leaning and Inference Laboratory, MLI-86-6, George Maseon University.
11. J. R. Quinlan.1992. Programs for Machine Learning, Morgan Kaufmann.
12. Z. K. Baker and V. K.Prasanna. 2005. Efficient Parallel Data Mining with the Apriori Algorithm on FPGAs. In Submitted to the IEEE International Parallel and Distributed Processing Symposium (IPDPS '05).
13. Jing He. 2009. Advances in Data Mining: History and Future, Third international Symposium on Information Technology Application, 978-0-7695-3859-4/09 IEEE 2009 DOI 10.1109/IITA.2009.204
14. Ali Meligy.2009. A Grid-Based Distributed SVM Data Mining Algorithm, European Journal of Scientific Research ISSN 1450-216X Vol.27 No.3. Pp.313-321 © Euro Journals Publishing, Inc. Available at: http://www.eurojournals.com/ejsr.htm
15. S. Mitra, S. K. Pal, and P. Mitra. 2001. Data mining in soft computing framework: A survey, IEEE Trans. Neural Networks, vol. 13, pp. 3 - 14.
16. Mark, J., Embrechts. 2005. Introduction to Scientific Data Mining: Direct Kernel Methods & Applications, Computationally Intelligent Hybrid Systems: The Fusion of Soft Computing and Hard Computing, Wiley, New York, pp. 317-365
17. Han, J., & Kamber, M. 2001. Data mining: Concepts and techniques. Morgan-Kaufman Series of Data Management Systems. San Diego: Academic Press.

ID: 17

MACHINE LEARNING AND DEEP LEARNING APPLICATIONS

Dr.N.SHANMUGAPRIYA
Assistant professor & Head,
Department of Computer Applications,
Erode Arts and Science College,
Erode, Tamil Nadu, India.

Dr.M.RAJESHWARI
Assistant Professor,
Department of Computer Science with Cyber Security,
Erode Arts and Science College,
Erode, Tamil Nadu, India.

ABSTRACT

Nowadays, large amount of data is available everywhere. Therefore, it is very important to analyze this data in order to extract some useful information and to develop an algorithm based on this chapter. This can be achieved through data mining and machine learning. Machine learning is an integral part of artificial intelligence, which is used to design algorithms based on the data trends and historical relationships between data. Machine learning is used in various fields such as bioinformatics, intrusion detection, Information retrieval, game playing, marketing, malware detection, image deconvolution and so on. This chapter presents in the field of machine learning in various application areas. Deep learning is an emerging area of machine learning (ML) research. It comprises multiple hidden layers of artificial neural networks. The deep learning methodology applies nonlinear transformations and model abstractions of high level in large databases. The recent advancements in deep learning architectures within numerous fields have already provided significant contributions in artificial intelligence. This article presents a state of the art survey on the contributions and the novel applications of deep learning. The following review chronologically presents how and in what major applications deep learning algorithms have been utilized. Furthermore, the superior and beneficial of the deep learning methodology and its hierarchy in layers and nonlinear operations are presented and compared with the more conventional algorithms in the common applications. The state of the art chapter further provides a general overview on the novel concept and the ever-increasing advantages and popularity of deep learning.

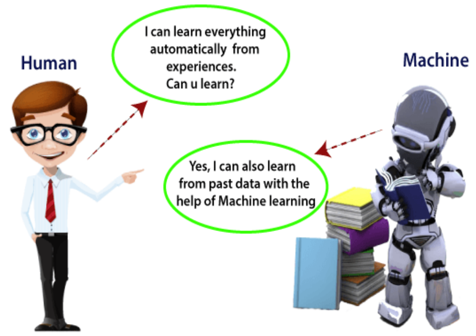
KEYWORDS: *Machine Learning, Deep Learning, Artificial Neural Networks, Intrusion, Detection.*

I. INTRODUCTION

Machine Learning tutorial provides basic and advanced concepts of machine learning. Our machine learning tutorial is designed for students and working professionals. Machine learning is a growing technology which enables computers to learn automatically from past data. Machine learning uses various algorithms for **building mathematical models and making predictions using historical data or information**. Currently, it is being used for various tasks such as **image recognition**, **speech recognition**, **email filtering**, **Face book auto-tagging**, **recommender system**, and many more. This machine learning tutorial gives you an introduction to machine learning along with the wide range of machine learning techniques such as **Supervised**, **Unsupervised**, and **Reinforcement** learning. You will learn about regression and classification models, clustering methods, hidden Markov models, and various sequential models.

WHAT IS MACHINE LEARNING

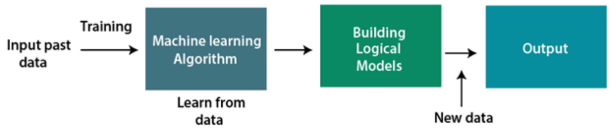
In the real world, we are surrounded by humans who can learn everything from their experiences with their learning capability, and we have computers or machines which work on our instructions. But can a machine also learn from experiences or past data like a human does? So here comes the role of **Machine Learning**.



Machine Learning is said as a subset of **artificial intelligence** that is mainly concerned with the development of algorithms which allow a computer to learn from the data and past experiences on their own. The term machine learning was first introduced by **Arthur Samuel** in **1959**. We can define it in a summarized way as: Machine learning enables a machine to automatically learn from data, improve performance from experiences, and predict things without being explicitly programmed. With the help of sample historical data, which is known as **training data**, machine learning algorithms build a **mathematical model** that helps in making predictions or decisions without being explicitly programmed. Machine learning brings computer science and statistics together for creating predictive models. Machine learning constructs or uses the algorithms that learn from historical data. The more we will provide the information, the higher will be the performance. **A machine has the ability to learn if it can improve its performance by gaining more data.**

HOW DOES MACHINE LEARNING WORK

A Machine Learning system **learns from historical data, builds the prediction models, and whenever it receives new data, predicts the output for it**. The accuracy of predicted output depends upon the amount of data, as the huge amount of data helps to build a better model which predicts the output more accurately. Suppose we have a complex problem, where we need to perform some predictions, so instead of writing a code for it, we just need to feed the data to generic algorithms, and with the help of these algorithms, machine builds the logic as per the data and predict the output. Machine learning has changed our way of thinking about the problem. The below block diagram explains the working of Machine Learning algorithm:



II. FEATURES OF MACHINE LEARNING

- Machine learning uses data to detect various patterns in a given dataset.
- It can learn from past data and improve automatically.
- It is a data-driven technology.
- Machine learning is much similar to data mining as it also deals with the huge amount of the data.

NEED FOR MACHINE LEARNING

The need for machine learning is increasing day by day. The reason behind the need for machine learning is that it is capable of doing tasks that are too complex for a person to implement directly. As a human, we have some limitations as we cannot access the huge amount of data manually, so for this, we need some computer systems and here comes the machine learning to make things easy for us. We can train machine learning algorithms by providing them the huge amount of data and let them explore the data, construct the models, and predict the required output automatically. The performance of the machine learning algorithm depends on the amount of data, and it can be determined by the cost function. With the help of machine learning, we can save both time and money. The importance of machine learning can be easily understood by its uses cases, currently, machine learning is used in **self-driving cars**, **cyber fraud detection**, **face recognition**, and **friend suggestion by Face book**, etc. Various top companies such as Netflix and Amazon have build machine learning models that are using a vast amount of data to analyze the user interest and recommend product accordingly.

Following are some key points which show the importance of Machine Learning:

- Rapid increment in the production of data
- Solving complex problems, which are difficult for a human
- Decision making in various sector including finance
- Finding hidden patterns and extracting useful information from data.

III. CLASSIFICATION OF MACHINE LEARNING

At a broad level, machine learning can be classified into three types:

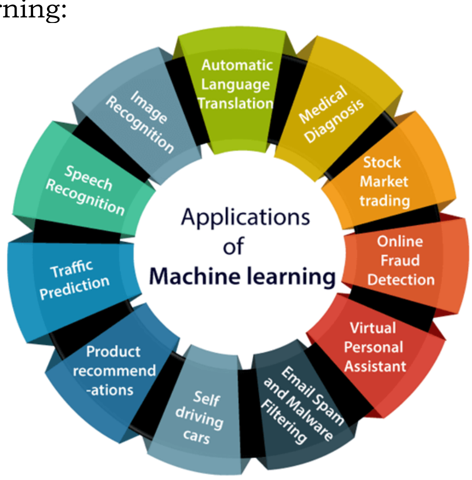
1. **Supervised learning**
2. **Unsupervised learning**
3. **Reinforcement learning**

IV. MACHINE LEARNING AT PRESENT

Now machine learning has got a great advancement in its research, and it is present everywhere around us, such as self-driving cars, Amazon Alexa, Catboats, recommender system, and many more. It includes Supervised, unsupervised, and reinforcement learning with clustering, classification, decision tree, SVM algorithms, etc. Modern machine learning models can be used for making various predictions, including weather prediction, disease prediction, stock market analysis, etc.

APPLICATIONS OF MACHINE LEARNING

Machine learning is a buzzword for today's technology, and it is growing very rapidly day by day. We are using machine learning in our daily life even without knowing it such as Google Maps, Google assistant, Alexa, etc. Below are some most trending real-world applications of Machine Learning:



1. Image Recognition
2. Speech Recognition
3. Traffic prediction
4. Product recommendations
5. Self-driving cars
6. Email Spam and Malware Filtering
7. Virtual Personal Assistant
8. Online Fraud Detection
9. Stock Market trading
10. Medical Diagnosis
11. Automatic Language Translation

V. DEEP LEARNING

Deep learning is based on the branch of machine learning, which is a subset of artificial intelligence. Since neural networks imitate the human brain and so deep learning will do. In deep learning, nothing is programmed explicitly. Basically, it is a machine learning class that makes use of numerous nonlinear processing units so as to perform feature extraction as well as transformation. The output from each preceding layer is taken as input by each one of the successive layers. Deep learning models are capable enough to focus on the accurate features themselves by requiring a little guidance from the programmer and are very helpful in solving out the problem of dimensionality. Deep learning algorithms are used, especially when we have a huge no of inputs and outputs. Since deep learning has been evolved by the machine learning, which itself is a subset of artificial intelligence and as the idea behind the artificial intelligence is to mimic the human behavior, so same is "the idea of deep learning to build such algorithm that can mimic the brain".

DEEP LEARNING APPLICATIONS

1. Virtual Assistants
2. Chatbots
3. Healthcare
4. Entertainment
5. News Aggregation and Fake News Detection
6. Composing Music
7. Image Coloring
8. Robotics
9. Image Captioning
10. Advertising
11. Self Driving Cars
12. Natural Language Processing
13. Visual Recognition
14. Fraud Detection
15. Personalisations
16. Detecting Developmental Delay in Children
17. Colourisation of Black and White images
18. Adding Sounds to Silent Movies
19. Automatic Machine Translation
20. Automatic Handwriting Generation
21. Automatic Game Playing
22. Language Translations
23. Pixel Restoration
24. Demographic and Election Predictions
25. Deep Dreaming

ISBN: 978-81-965283-7-9

CONCLUSION

Deep Learning has found its prominence in almost every sector of business. It is being used in E-Commerce, Healthcare, Advertising, Manufacturing, Entertainment, and many other industries. Deep Learning has revolutionized our lives by making tasks easier.

REFERENCES

1. Aamodt, Agnar, and Enric Plaza. “Case-based reasoning: Foundational issues, methodological variations, and system approaches.” AI communications 7.1 (1994): 39-59.
2. Adebayo, Julius, Justin Gilmer, Michael Muelly, Ian Goodfellow, Moritz Hardt, and Been Kim. “Sanity checks for saliency maps.” arXiv preprint arXiv:1810.03292 (2018).
3. Alain, Guillaume, and Yoshua Bengio. “Understanding intermediate layers using linear classifier probes.” arXiv preprint arXiv:1610.01644 (2016).
4. Alber, Maximilian, Sebastian Lapuschkin, Philipp Seegerer, Miriam Hägele, Kristof T. Schütt, Grégoire Montavon, Wojciech Samek, Klaus-Robert Müller, Sven Dähne, and Pieter-Jan Kindermans. “iNNvestigate neural networks!.” J. Mach. Learn. Res. 20, no. 93 (2019): 1-8.
5. Alberto, Túlio C, Johannes V Lochter, and Tiago A Almeida. “Tubespam: comment spam filtering on YouTube.” In Machine Learning and Applications (Icmla), Ieee 14th International Conference on, 138–43. IEEE. (2015).
6. Alvarez-Melis, David, and Tommi S. Jaakkola. “On the robustness of interpretability methods.” arXiv preprint arXiv:1806.08049 (2018).
7. Ancona, Marco, et al. “Towards better understanding of gradient-based attribution methods for deep neural networks.” arXiv preprint arXiv:1711.06104 (2017).
8. Bishop, C. M. (2006) Pattern Recognition and Machine Learning. Chapter 5: Neural Networks.
   Schmidhuber, J. (2015). Deep Learning in Neural Networks: An Overview. Neural Networks 61: 85-117.
   Bengio, Y., LeCun, Y., Hinton, G. (2015). Deep Learning. Nature 521: 436-44.
   Goodfellow, I., Bengio, Y. and Courville, A. (2016) Deep Learning. MIT Press.
9. A little bit more complete list of references and online resources is here.
10. An extensive list of references can also be found at https://github.com/terryum/awesome-deep-learning-papers and http://deeplearning.net/reading-list/

ID: 18

BLOCKCHAIN TECHNOLOGY: A REVIEW, ARCHITECTURE, CONSENSUS, ISSUES AND CHALLENGES

Ms.D.SEEMA DEV AKSATHA
Assistant Professor,
Department of Computer Science,
Sri Krishna Arts and Science,
Coimbatore, Tamil Nadu, India.

Dr.R.PUGAZENDI
Assistant Professor,
Department of Computer Science,
Government Arts College,
Salem, Tamil Nadu, India

ABSTRACT

The BlockChain is the latest and outlook technology in current economy. A BlockChain is principally a distributed database of records or public journal of all businesses or digital procedures that have been executed and pooled among contributing parties. It provides provenance, immutability and definiteness for the transfer of value within a business network. It enables value exchange in real time, reducing costs and errors. Based on a network consensus approach, whereby trust between the parties involved in a transaction is provided by cryptography. The cause for the concern in BlockChain Technology is it's spread out and scattered features that give security, anonymity, and data integrity with no in-between agency on top of things of transactions. BlockChain is used more and more for record-keeping, endorsing, and validating digital assets (Financial assets, land, etc.) and dealings, governing exchanges, recording data, and managing IDs among multiple parties in a very trusted, decentralized, and secure manner. In this chapter the concept of BlockChain Technology(BT), Architecture and Applications, Advantages and Disadvantages of BlockChain Technology are reviewed.

KEYWORDS: *BlockChain, Decentralized, Ledger, Applications, Consensus.*

I. INTRODUCTION

As aforementioned, BlockChain technology was introduced by Nakamoto in 2008 as an underlying technology for Bitcoin to record all transactions of Bitcoin and to create security against potential attacks [1]. Fig. 1 presents BlockChain's roadmap from 2008 to 2019. Bitcoin's initial infrastructure based on BlockChain technology appeared in 2009 over a peer-to-peer (P2P) network, which is called the Bitcoin network. Since then, cryptocurrencies have gained worldwide attention, and researchers have harnessed and applied BlockChain technology to domains, such as smart contracts and supply chain management. This evolution has occurred because BlockChain is autonomous, distributed, immutable, and contractual [2, 3].

Fig.1 Roadmap of BlockChain

II. TYPES OF BLOCKCHAIN

Public BlockChain: In which all the participated users can perform read and write operation such as Bitcoin, in which anyone can join and become the part of the BlockChain. However, some public BlockChains are restricted with access the data of Block, it can either to read or write down only not both.

Private BlockChain: Only the trusted participants can access the data of BlockChain, mostly used by private organizations where they don't want any interaction of public on blocks because the blocks may have sensitive or private data of corporation.

Consortium BlockChain: The consortium BlockChains are semi-decentralized. When multiple companies want to become the part of BlockChain, control its operations and access control protocols simultaneously, instead of one particular organization have fully controlling access to it.[8]

III. ARCHITECTURE

A BlockChain is a linked-data structure where each block contains two key fragments: a header and body. The header fragment contains a nonce, a previous hash, a Merkle root hash, a timestamp, and a difficulty target. The body fragment contains a list of transactions. Fig. 2 represents the structure of a BlockChain. The first fragment is always named a genesis, all blocks are linked together by cryptography, and blocks are disseminated between nodes over a network [3].

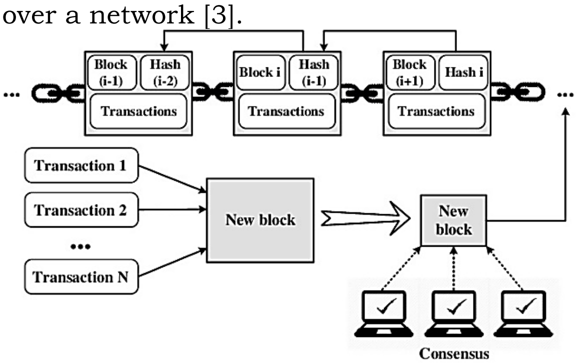


Fig.2 Structure of a BlockChain

Moreover, to observe to the rules of BlockChain technology, all the nodes in the BlockChain network need to have the identical block list, which is represented in Fig. 3.When a new block is added, it broadcasts to all nodes in the network. Each node confirms the new block through a consensus mechanism that endorses a transaction in the block. There are several consensus algorithms to confirm that all nodes have the same BlockChain list, such as resilient of work and resilient of stake [4,5].

Basic Principles of BlockChain Technology

There are several principles of BlockChain technology that are applied to three main layers: the network layer, data layer, and application layers. First, the network layer is well-matched with the P2P network architecture, which supports decentralized networks and scattered network mechanisms. The network layer is accountable for forwarding and validating data between nodes. In addition, BlockChain technology stores the identical chain in all nodes over a network; thus, all nodes are coordinated. Hence, when a new block is generated, it is then verified by a consensus algorithm. If the new block is valid, then it broadcasts to all other nodes. Otherwise, it is discarded. Furthermore, there are quite a few types of consensus algorithms that all function on two principles: (i) the freshness principle achieves fair competition through new resources for every new block that is added, and (ii) the unpredictability principle stops any member from forecasting which node will create a new block. Tab. 2 illustrates some of the consensus algorithms that are used in BlockChain networks [6,7].

IV. APPLICATIONS OF BLOCKCHAIN

BT in Food Industry:

The BlockChain when it gets more and more incorporated into the food industry it'll make the entire process more transparent and safer.

These are the welfares of a translucent food system:

- BT reinforces the safety of food.
- It makes sure about new food because nobody will take any risk for transporting "non-fresh" food in a public decentralized system.
- A very minor quantity of left-over product found because all record of data about food is stored
- Food deceptions are also stopped by using BT since anybody can control anything from this open system.

BlockChain Technology in Cybersecurity:

On 7 September 2017, Equifax, one in all the world's bigger bank line reporting agencies, upset the world once they exposed that they'd just handled an massive cybersecurity breach. They met unauthorized data access from mid-May through July 2017, which they found on July 29. Around 145.5 million consumers were in danger of getting their personal information stolen which included Names of customer ,Social Security identification ,Dates of birth of customer, Residence Addresses, License numbers of Driver's etc.

These are some features of BlockChain that helps in preventing cyberattacks:

Immutability: This is one of the best operative feature in which when data is entered by the user then it cannot be altered by anyone. The BT keeps the information safe by employing its several cryptographic algorithms, digital signatures and various hashing algorithms.

Decentralization and Consensus: BT is a point-to-point, dispersed and scattered block system in which a collection of nodes can be created and to make any changes in any node needs consensus algorithm and verify that these changes then perform all operations.

Voting: Some common problems related to the system of paper ballot are:

- The vote can be stolen or transaction can be changed by inserting fake ballot papers.
- It is very hard to track or access the vote and it takes too much time for counting the votes
- The quantity of surplus paper in paper ballot system may be harm to the atmosphere.

BlockChain technology gives a solution of all these problems:

By using BlockChain technology voting process can be made transparent where anybody can perceive the operation of voting system and casting of vote will be concealed by integrating the BlockChain. The voter can give vote securely by login their personal government-issued ID card and their own webcam. The voter creates their own transaction, and this process is unique for each voter. While casting of vote, the voter can easily confirm that the vote is valid or not and confirm that the votes are not altered.

Applications of BT in Healthcare:
BlockChain Technology plays an important role in healthcare also. The application areas are:

Electronic Medical Records(EMR): Tamper-proof and longitudinal patient record can be achieved and BT provides distributed ledger technology which helps to place all records like vaccines records, results of lab report, treatment strategies, and history of prescription with a decentralized way excepting a central location of all records.

Tokenized Healthcare: The users can share the personal medical information, earn and learn by using their unique medical information. To prevent the diseases or for treatment of patients may also be motivated by using tokenization.

Vehicle Industries: Auto or vehicle enterprises have several application areas for BT. The centralized BT based supply chain with trust-based distribution is that the latest model for a way we construct, procure vehicles for day to day uses. IoT may well be adapted automatically restart BT-based registers to preserve a transparent and immutable vehicle report. This may useful to extend transparency beyond the industry and provide purchasing a "lemon" impossibly. Parts are procured from different vendors and evolving BlockChain applications, IoT for tracking these movement of pieces in an remarkably tamper- proof and authenticated system would enhance the way the vehicles are sold, bought, constructed, and distributed.

Smart Devices: Smart devices presently play a vital role in every arena. As an alternative of storing this data using a central server or cloud-based storage solution, smart appliance data might be stored on the BlockChain. This helps to protect personal information and keep home IoT webs protected. Data can be adapted to advance things like energy costs for a complete grid deprived of linking the data to the human by applying public/ private key cryptographic algorithm to examine out personal identity from presently available data, at the same time keeping the info is authentic.

Supply Chains: It involves globally many parties across time zones and is a multi-layered chain. From food distributors to pharmaceutical enterprises, many supply chains may benefit from employing a combination of IoT and BlockChain to rationalize the process. Handover of ownership and settlement can be tracked in real time amid IoT instruments, the freights themselves or each object individually.

Advantages of BlockChain
Decentralization: BlockChain is a peer-to-peer distributed technology; it eliminates the requisite of third-party as distributer and also avoids all the additional operating expenses and fees of transaction.

Immutability and Data integrity: Each time a new transition is noted in the BlockChain database, it cannot be changed directly or can be removed only after applying consensus. Members of BlockChain can lessen fraud while reinforcing regulatory compliance.
Greater Accessibility and availability: In BT as data is warehoused in a distributed way it cannot be easily accessed by anyone whenever they need the data with accuracy.
Processing Time: By using BT, transaction's processing time is almost reduced from 3 days to minutes or seconds.
Security: Each transaction is allotted with a unique time-stamped cryptographic hash code, it may be a 64- or 128-digits alpha-numeric key signature value that is entered conforming to transaction of each block.
Reliability: In BT possibilities of failure are very less as it is well-ordered by diverse control centres and not on a single point.
Transparency: All the transactions involved in BT are transparent. Everybody can perceive the particulars of other transactions and every node comprises the complete ledger. The shared digital ledger comprises all the information of the actual source, destination, date and time of the block transactions.[9]

V. CHALLENGES IN BLOCKCHAIN TECHNOLOGY

Issue of high Expense: BlockChian Technology devises initial opening charges and hence using BT isn't unbound from cost which is a hindrance of decentralization. The user node must pay money for the transactions and computational power.

Latency issues: For providing the security in BT, multifaceted verification and validation process are accomplished that takes more time for verifying the transactions of BT.

Wasted Resources: During multifaceted verification and mining process it needs enormous computational power like CPU's and GPU's power. The energy used in Bitcoin mining network is near about $15 million per day.[9]

VI. CONCLUSION

The BlockChain technology is currently one of the beneficial and adaptable concerns for our world. BT provides high security of data but takes more time for verification and validation process of transactions. The main features of its privacy and security, traceability and time-stamping has observing its acceptation beside its primary application areas. Currently BlockChain technology potentials the bright future for information technology deprived of the fraud and any deception due to some benefits of the BlockChain technology. The experiments of the BlockChain are large, but the outcomes of the BlockChain using have a greater multitude than drawbacks. It is necessary to keep exploring the BlockChain development and application in the different areas for the nearest future, because this new technology can help to solve many difficult complications, which are worrying and preventing the correct functioning of systems.

VI. REFERENCES

1. S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008. [Online]. Available at: https://www.klausnordby.com/bitcoin/Bitcoin_Whitepaper_Document_HD.pdf.
2. Q. Feng, D. He, S. Zeadally, M. K. Khan and N. Kumar, "A survey on privacy protection in BlockChain system,"Journal of Network and Computer Applications, vol. 126, no. 2, pp. 45–58, 2019.
3. W. Gao, W. G. Hatcher and W. Yu, "A survey of BlockChain: techniques, applications, and challenges," in 2018 27th Int. Conf. on Computer Communication and Networks (ICCCN), pp. 1–11, 2018.
4. X. Liang, S. Shetty, D. Tosh, C. Kamhoua, K. Kwiat et al., "Provchain: A BlockChain-based data provenance architecture in cloud environment with enhanced privacy and availability," in Proc. of the 17th IEEE/ACM Int. sym. on cluster, cloud and grid computing, pp. 468–477, 2017.
5. M. Muzammal, Q. Qu and B. Nasrulin, "Renovating BlockChain with distributed databases: An open-source system," Future Generation Computer Systems, vol. 90, no. Supplement C, pp. 105–117, 2019.
6. Z. Zheng, S. Xie, H. Dai, X. Chen and H. Wang, "An overview of BlockChain technology: Architecture, consensus, and future trends," in 2017 IEEE International Congress on Big Data (BigData Congress), pp. 557–564, 2017.
7. Y. Yuan and F. Y. Wang, "BlockChain

and cryptocurrencies: model, techniques, and applications," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 48, no. 9, pp. 1421–1428, 2018.
8. Rahul Rao Vokerla, Bharanidharan Shanmugam, Et.al "An Overview of BlockChain Applications and Attacks", International Conference on Vision Towards Emerging Trends in Communication and Networking, IEEE 2019.
9. Thomas kitsantas, Athanasios Vazakidis, Et.al. "A Review of BlockChain Technology and Its Applications in the Business Environment", Research Gate, 2019.

ID: 19

DATA MINING TASKS, TECHNIQUES AND APPLICATIONS

Ms.S.SUMATHI
Assistant Professor,
Department of Computer Science,
Karpagam Academy of Higher Education,
Coimbatore, Tamil Nadu, India.

Dr.P.TAMILSELVAN
Associate Professor,
Department of Computer Science,
Karpagam Academy of Higher Education,
Coimbatore, Tamil Nadu, India.

ABSTRACT:
Data mining is a field of intersection of computer science and statistics used to discover patterns in the information bank. Data mining is a new technology that helps businesses to predict future trends and behaviours, allowing them to make proactive, knowledge driven decisions. It is also called as knowledge discovery process. The chapter discusses classification of data mining techniques and the applications in real world. Practically, data mining is really useful for any organization which has huge amount of data. Data mining help regular databases to perform faster. They also help to increase the profit, because of the correct decisions made with the help of data mining.

KEYWORDS: *Data Mining, Data Types, Data Transformation, Data Mining Techniques, Data Mining Applications*

I. INTRODUCTION

What is Data Mining?

Data Mining is a process of finding potentially useful patterns from set of amounts of data. It is a multi-disciplinary skill that uses machine learning, statistics, and Artificial intelligence to extracting information to evaluate the future contents probability. The insights derived from Data Mining are used for marketing, sales, fraud detection, and scientific discovery, and so many things. Data Mining is nothing but all about discovering hidden data and, unsuspected, data previously known. Correct value in a relationship amongst the data. Data mining is also called Knowledge Discovery in Data (KDD), Knowledge extraction, discovery and data pattern analysis, information harvesting, etc.

Types of Data

Data mining can be performed on following types of data

- Relational databases
- Data warehouses
- Advanced DB and information repositories
- Object-oriented and object-relational databases
- Transactional and Spatial databases
- Heterogeneous and legacy databases
- Multimedia and streaming database
- Text databases
- Text mining and Web mining

Business Understanding:

In this phase, business and data-mining goals are established.

- First, you need to understand business and client objectives. You need to define what your client wants (which many times even they do not know themselves)
- Take stock of the current data mining scenario. Factor in resources, assumption, constraints, and other significant factors into your assessment.
- Using business objectives and current scenario, define your data mining goals.
- A good data mining plan is very detailed and should be developed to accomplish both business and data mining goals.

Data Understanding:

- In this phase, sanity check on data is performed to check whether it's appropriate for the data mining goals.

- First, data is collected from multiple data sources available in the organization.
- These data sources may include multiple databases, flat filer or data cubes. There are issues like object matching and schema integration which can arise during Data Integration process. It is a quite complex and tricky process as data from various sources unlikely to match easily. For example, table A contains an entity named cust_no whereas another table B contains an entity named cust-id.
- Therefore, it is quite difficult to ensure that both of these given objects refer to the same value or not. Here, Metadata should be used to reduce errors in the data integration process.
- Next, the step is to search for properties of acquired data. A good way to explore the data is to answer the data mining questions (decided in business phase) using the query, reporting, and visualization tools.
- Based on the results of query, the data quality should be ascertained. Missing data if any should be acquired.

Data Preparation:

- In this phase, data is made production ready.
- The data preparation process consumes about 90% of the time of the project.
- The data from different sources should be selected, cleaned, transformed, formatted, anonymized, and constructed (if required).
- Data cleaning is a process to "clean" the data by smoothing noisy data and filling in missing values.
- For example, for a customer demographics profile, age data is missing. The data is incomplete and should be filled. In some cases, there could be data outliers. For instance, age has a value 300. Data could be inconsistent. For instance, name of the customer is different in different tables.
- Data transformation operations change the data to make it useful in data mining. Following transformation can be applied

Data Transformation:

- Data transformation operations would contribute toward the success of the mining process.
- Smoothing: It helps to remove noise from the data.
- Aggregation: Summary or aggregation operations are applied to the data. I.e., the weekly sales data is aggregated to calculate the monthly and yearly total.
- Generalization: In this step, Low-level data is replaced by higher-level concepts with the help of concept hierarchies. For example, the city is replaced by the county.
- Normalization: Normalization performed when the attribute data are scaled up o scaled down. Example: Data should fall in the range -2.0 to 2.0 post-normalization.
- Attribute construction: these attributes are constructed and included the given set of attributes helpful for data mining.
- The result of this process is a final data set that can be used in modeling.

Modelling

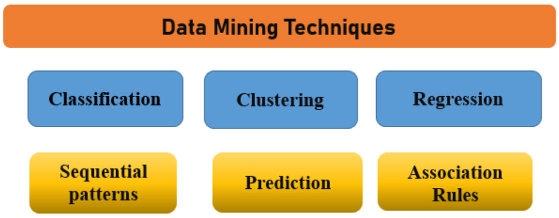
In this phase, mathematical models are used to determine data patterns.

- Based on the business objectives, suitable modeling techniques should be selected for the prepared dataset.
- Create a scenario to test check the quality and validity of the model.
- Run the model on the prepared dataset.
- Results should be assessed by all stakeholders to make sure that model can meet data mining objectives.
- Evaluation:
- In this phase, patterns identified are evaluated against the business objectives.
- Results generated by the data mining model should be evaluated against the business objectives.
- Gaining business understanding is an iterative process. In fact, while understanding, new business requirements may be raised because of data mining.
- A go or no-go decision is taken to move the model in the deployment phase.
- Deployment:
- In the deployment phase, you ship your data mining discoveries to everyday business operations.

ISBN: 978-81-965283-7-9

- The knowledge or information discovered during data mining process should be made easy to understand for non-technical stakeholders.
- A detailed deployment plan, for shipping, maintenance, and monitoring of data mining discoveries is created.
- A final project report is created with lessons learned and key experiences during the project. This helps to improve the organization's business policy.

II. DATA MINING TECHNIQUES



2.1 Data Mining Techniques Diagram

1. Classification:

This analysis is used to retrieve important and relevant information about data, and metadata. This data mining method helps to classify data in different classes.

2. Clustering:

Clustering analysis is a data mining technique to identify data that are like each other. This process helps to understand the differences and similarities between the data.

3. Regression:

Regression analysis is the data mining method of identifying and analyzing the relationship between variables. It is used to identify the likelihood of a specific variable, given the presence of other variables.

4. Association Rules:

This data mining technique helps to find the association between two or more Items. It discovers a hidden pattern in the data set.

5. Outer detection:

This type of data mining technique refers to observation of data items in the dataset which do not match an expected pattern or expected behavior. This technique can be used in a variety of domains, such as intrusion, detection, fraud or fault detection, etc. Outer detection is also called Outlier Analysis or Outlier mining.

6. Sequential Patterns:

This data mining technique helps to discover or identify similar patterns or trends in transaction data for certain period.

7. Prediction:

Prediction has used a combination of the other techniques of data mining like trends, sequential patterns, clustering, classification, etc. It analyzes past events or instances in a right sequence for predicting a future event.

Challenges in Implementation of Data mining

- Skilled Experts are needed to formulate the data mining queries.
- Overfitting: Due to small size training database, a model may not fit future states.
- Data mining needs large databases which sometimes are difficult to manage
- Business practices may need to be modified to determine to use the information uncovered.
- If the data set is not diverse, data mining results may not be accurate.
- Integration information needed from heterogeneous databases and global information systems could be complex

III. DATA MINING EXAMPLES

Now in this Data Mining course, let's learn about Data mining with examples:

Example 1:

Consider a marketing head of telecom service provides who wants to increase revenues of long distance services. For high ROI on his sales and marketing efforts customer profiling is important. He has a vast data pool of customer information like age, gender, income, credit history, etc. But its impossible to determine characteristics of people who prefer long distance calls with manual analysis. Using data mining techniques, he may uncover patterns between high long distance call users and their characteristics.

For example, he might learn that his best customers are married females between the age of 45 and 54 who make more than $80,000 per year. Marketing efforts can be targeted to such demographic.

ISBN: 978-81-965283-7-9

Disadvantages of Data Mining

- There are chances of companies may sell useful information of their customers to other companies for money. For example, American Express has sold credit card purchases of their customers to the other companies.
- Many data mining analytics software is difficult to operate and requires advance training to work on.
- Different data mining tools work in different manners due to different algorithms employed in their design. Therefore, the selection of correct data mining tool is a very difficult task.
- The data mining techniques are not accurate, and so it can cause serious consequences in certain conditions.

Data Mining Applications

Applications	Usage
Communications	Data mining techniques are used in communication sector to predict customer behaviour to offer highly targeted and relevant campaigns.
Insurance	Data mining helps insurance companies to price their products profitable and promote new offers to their new or existing customers.
Education	Data mining benefits educators to access student data, predict achievement levels and find students or groups of students which need extra attention. For example, students who are weak in maths subject.
Manufacturing	With the help of Data Mining Manufacturers can predict wear and tear of production assets. They can anticipate maintenance which helps them reduce them to minimize downtime.
Banking	Data mining helps finance sector to get a view of market risks and manage regulatory compliance. It helps banks to identify probable defaulters to decide whether to issue credit cards, loans, etc.
Retail	Data Mining techniques help retail malls and grocery stores identify and arrange most sellable items in the most attentive positions. It helps store owners to come up with the offer which encourages customers to increase their spending.

IV. SUMMARY:

- Data Mining Definition: Data Mining is all about explaining the past and predicting the future via Data analysis.
- Data mining helps to extract information from huge sets of data. It is the procedure of mining knowledge from data.
- Data mining process includes business understanding, Data Understanding, Data Preparation, Modelling, Evolution, and Deployment.
- Important Data mining techniques are Classification, clustering, Regression, Association rules, Outer detection, Sequential Patterns, and prediction
- R-language and Oracle Data mining are prominent data mining tools and techniques.
- Data mining technique helps companies to get knowledge-based information.
- The main drawback of data mining is that many analytics software is difficult to operate and requires advance training to work on.
- Data mining is used in diverse industries such as Communications, Insurance, Education, Manufacturing, Banking, Retail, Service providers, eCommerce, Supermarkets Bioinformatics.

V. REFERENCES

1. Data Mining: Concepts and Techniques, 3rd Edition. Jiawei Han, Micheline Kamber, Jian Pei. Database Modeling and Design: Logical Design, 5th Edition.
2. Introduction to Data Mining, Psng-Ning Tan, Vipin Kumar, Michael Steinbanch, Pearson Educator.
3. Data Mining Techniques, Arun KPujari, 3rd edition.

ID: 20

AN EVALUATION OF DNA CRYPTOGRAPHY AS A NEW INFORMATION SECURITY FRONTIER

Dr.L.LEELAVATHY
Assistant Professor,
Department of Computer Science,
SRM Trichy Arts & Science College,
Trichy, Tamil Nadu, India.

Mr.M.VENKATESH SARAVANAKUMAR
Assistant Professor,
Department of Computer Science,
SRM Trichy Arts & Science College,
Trichy, Tamil Nadu, India.

ABSTRACT

As technology advances and data becomes increasingly vital in our digital age, the need for secure communication and data protection has never been more critical. Traditional cryptographic methods have served us well, but as the threats to data security evolve, new approaches are required to stay ahead. One such emerging field is DNA cryptography, which leverages the unique properties of DNA molecules to encrypt and decrypt sensitive information. In this article, we delve into the world of DNA cryptography and compare it to conventional cryptographic techniques, exploring its potential as a groundbreaking frontier in information security.

KEYWORDS: *DNA Cryptography, Data Security, DNA Synthesis, Scalability, Biosecurity.*

I. Introduction

Understanding DNA Cryptography:

DNA, short for deoxyribonucleic acid, is the fundamental building block of life, encoding the genetic information of living organisms. It is a molecule that stores information in the form of a sequence of nucleotides: adenine (A), cytosine (C), guanine (G), and thymine (T). DNA possesses remarkable characteristics that make it a candidate for cryptographic applications:

Massive Data Storage: DNA can store an enormous amount of data in a compact form. A single gram of DNA can theoretically store approximately 215 petabytes (215 million gigabytes) of information, far exceeding any current digital storage medium.

Biocompatibility: DNA is biocompatible, non-toxic, and stable over long periods. This property makes it a viable candidate for various applications, including data storage and cryptography.

Parallel Processing: DNA operates in parallel, enabling multiple computations to occur simultaneously. This feature can be harnessed in cryptographic processes for high-speed operations.

II. Comparing DNA Cryptography and Conventional Cryptography

2.1 Encryption Techniques:

Conventional Cryptography: Traditional cryptographic methods utilize algorithms and keys to encrypt and decrypt data. The security of these systems relies on the complexity of the algorithm and the secrecy of the keys.

DNA Cryptography: DNA encryption involves mapping the binary data into DNA sequences and manipulating them using biochemical operations. This approach replaces algorithms and keys with biochemical processes, offering a unique form of encryption.

Security Strength: Conventional Cryptography: The security of traditional cryptographic methods is primarily dependent on the strength of the algorithm and the key length. As computational power increases, brute-force attacks become more feasible, potentially compromising security.

DNA Cryptography: The security strength of DNA cryptography comes from the inherent complexity of biological processes involved in DNA manipulation. The vast parallelism and randomness make it challenging for attackers to decipher the encrypted data.

2.2 Error Tolerance:

Conventional Cryptography: Traditional cryptographic methods are sensitive to data errors, leading to potential decryption failures if even a single bit is altered or lost.

DNA Cryptography: DNA is remarkably error-tolerant due to its redundancy and self-repairing capabilities. It can tolerate minor errors during data manipulation or sequencing, ensuring more robust data retrieval.

2.3 Speed and Efficiency:

Conventional Cryptography: Modern encryption algorithms are highly efficient, and encryption/decryption processes can be completed rapidly, allowing real-time secure communication.

DNA Cryptography: Currently, DNA cryptography is slower compared to conventional methods due to the need for physical biochemical operations. However, advancements in biotechnology could potentially improve the processing speed in the future.

III. Potential Applications of DNA Cryptography:

Secure Data Storage: As mentioned earlier, DNA has an astonishing data storage capacity. DNA-based storage systems have the potential to revolutionize data archiving and long-term preservation. Storing sensitive information, historical records, or valuable intellectual property in DNA form could provide an ultra-durable and secure storage medium.

Authentication and Biometrics: DNA cryptography can be used for biometric authentication purposes. Instead of storing complete DNA sequences, cryptographic representations of specific DNA segments could be utilized for identity verification. This approach could lead to more robust and tamper-resistant biometric systems.

Cryptographically-Secured Bioinformatics: DNA cryptography can play a significant role in bioinformatics applications. Securing genomic data and ensuring privacy in medical research are critical challenges that DNA cryptography could address. By encrypting genomic information, researchers and medical professionals can share and analyze data without compromising patient privacy.

Quantum-Resistant Cryptography: Quantum computers have the potential to break conventional cryptographic methods. DNA-based cryptography, with its unique biochemical processes, may offer an alternative that is resistant to quantum attacks, ensuring long-term data security in the post-quantum computing era.

IV. Challenges and Limitations:

Biotechnological Complexity: DNA manipulation requires sophisticated biotechnological processes, making it more challenging to implement and perform compared to traditional digital cryptography. Current DNA synthesis and sequencing techniques are relatively slow and expensive, limiting the speed and practicality of DNA cryptography.

Error Rates and Redundancy: While DNA possesses error tolerance, it is not entirely immune to errors during data encoding, synthesis, or decoding. High error rates could impact the accuracy of data retrieval and decryption. To counter this, error-correction techniques must be incorporated, adding complexity to the overall system.

Ethical and Social Concerns: DNA cryptography raises ethical considerations, especially regarding the source of DNA used for cryptographic purposes. The use of human DNA, even if non-coding regions, raises concerns about privacy and consent. Ethical guidelines must be established to address potential misuse of DNA data and ensure responsible use.

Interoperability and Standardization: For DNA cryptography to become a viable solution, standardized protocols and formats must be developed to ensure interoperability across different systems and platforms. Without such standards, widespread adoption could be hindered.

Regulatory and Legal Challenges: DNA is a regulated and sensitive biological material in many jurisdictions. Implementing DNA cryptography may require navigating through complex legal frameworks related to genetic data handling and ownership.

V. Current Research and Developments:

DNA Encryption Algorithms: Researchers are actively exploring novel encryption algorithms that leverage the unique properties of DNA sequences. These algorithms aim to strike a balance between security strength, error tolerance, and computational efficiency. Efforts are underway to design encryption schemes that can withstand potential attacks while minimizing the resources required for encryption and decryption.

Error Correction Techniques: To enhance the reliability of DNA-based data storage and encryption, scientists are investigating various error correction methods. These techniques aim to identify and correct errors that may occur during DNA synthesis, sequencing, or other biochemical operations, ensuring accurate data retrieval.

DNA Data Storage Solutions: Several research initiatives focus on developing practical DNA data storage systems. These projects aim to demonstrate the viability of DNA-based storage by encoding and retrieving data from DNA molecules effectively. Such systems could revolutionize large-scale data archiving and preserve valuable information for extended periods.

Biocomputing and DNA-Based Processing: DNA cryptography is not limited to encryption and data storage. Researchers are exploring ways to perform computations and logical operations using DNA molecules. This field, known as biocomputing or DNA-based computing, has the potential to usher in a new era of unconventional, high-speed cryptographic operations.

VI. Future Prospects and Advancements:

Scalability and Efficiency: As biotechnological techniques continue to advance, the scalability and efficiency of DNA cryptography are expected to improve significantly. Reduced costs and faster processes for DNA synthesis and sequencing would make DNA-based encryption more practical for real-world applications.

Hybrid Cryptosystems: Future cryptographic systems may adopt a hybrid approach, combining conventional cryptographic algorithms with DNA-based encryption. This integration could provide an additional layer of security, harnessing the strengths of both methods to create more robust and versatile encryption solutions.

Standardization and Interoperability: The development of standardized protocols and formats for DNA cryptography will be essential to ensure interoperability and widespread adoption. Collaborative efforts by the scientific community and relevant industry stakeholders are crucial to establish a common framework.

DNA-Based Quantum Cryptography: It's possible to investigate DNA cryptography within the framework of quantum cryptography. By combining quantum cryptography methods with DNA-based processes, hybrid schemes with improved security against both classical and quantum attackers may be created.

Ethical and Privacy Considerations: With the potential expansion of DNA-based technologies, the ethical use of genetic data in cryptography must be carefully addressed. Striking a balance between the benefits of DNA cryptography and safeguarding individual privacy and consent will be paramount in shaping responsible practices.

VII. Additional Challenges:

Cost and Accessibility: Currently, DNA sequencing and synthesis technologies can be costly and not readily accessible to everyone. As a result, implementing DNA cryptography on a large scale could be economically prohibitive for some organizations and industries. Reducing the cost and increasing the accessibility of biotechnological tools will be crucial to overcome this challenge.

Data Retrieval Speed: DNA-based data storage and cryptographic processes may not be as fast as traditional digital methods. While DNA offers unparalleled data density,

ISBN: 978-81-965283-7-9

the time required for encoding, decoding, and retrieving data from DNA molecules could hinder real-time applications. Research efforts are needed to optimize the speed of DNA data operations.

Environmental Impact: DNA cryptography, particularly in large-scale data storage applications, may require significant resources and chemical materials, potentially raising environmental concerns. Minimizing the ecological footprint of DNA-based processes will be essential to ensure sustainability and responsible use.

Biological Constraints: DNA is a biological material subject to natural degradation over time. While current DNA storage techniques can preserve data for thousands of years, its longevity is still subject to various environmental factors. Ensuring the long-term stability and integrity of DNA-encoded information will be critical.

VIII. Additional Advantages:

Unbreakable Encryption: Due to the inherent complexity and randomness of DNA-based processes, DNA cryptography could offer a level of encryption that is practically unbreakable with current technologies. This heightened security could be especially valuable for protecting highly sensitive information and critical infrastructure.

Steganography Potential: DNA cryptography offers potential for steganography, a technique where data can be hidden within other data to avoid detection. Concealing sensitive information within DNA sequences could provide an extra layer of covert security.
Bio-inspired Security: DNA cryptography draws inspiration from natural biological processes, tapping into the brilliance of nature's designs. Such bio-inspired security measures have the potential to revolutionize cybersecurity and offer novel solutions to challenging cryptographic problems.

Data Immortality: Unlike traditional digital storage mediums, DNA-encoded data has the potential for long-term immortality. This property could be particularly beneficial for preserving essential historical records, cultural heritage, and other valuable information for future generations.

IX. CONCLUSION:

DNA cryptography represents an exciting and innovative approach to information security, harnessing the unique properties of DNA molecules for encryption, data storage, and computation. While the field is still in its early stages, ongoing research and developments continue to unlock its vast potential.
To realize the promise of DNA cryptography, interdisciplinary collaboration is vital, bringing together experts from cryptography, biotechnology, computer science, and ethics. Addressing the challenges related to cost, speed, scalability, and ethical concerns will be key to advancing this groundbreaking technology and integrating it into practical applications.
As we venture further into the digital frontier, novel cryptographic methods like DNA cryptography offer hope for safeguarding our ever-expanding data-driven world. Embracing innovation, maintaining ethical principles, and remaining committed to security will empower us to harness the power of DNA in securing the data of tomorrow. With continued dedication and exploration, DNA cryptography has the potential to become a cornerstone of modern information security, paving the way for a more secure and resilient digital future.

X. REFERENCES:

1. Boneh, D., & Zhandry, M. (2013). Secure signatures and chosen ciphertext security in a quantum computing world. In Proceedings of the 2013 ACM conference on Computer and Communications Security (CCS '13), 361-372. DOI: 10.1145/2508859.2516737
2. Mo, F., Wang, Y., & Zhang, X. (2015). Cryptanalysis of the DNA-based encryption scheme. IET Information Security, 9(3), 177-180. DOI: 10.1049/iet-ifs.2014.0211
3. Castro, C. C., Yang, L., & Miao, F. (2015). DNA cryptography based on general encoding rules. Journal of Computational and Theoretical Nanoscience, 12(9), 2027-2032. DOI: 10.1166/jctn.2015.4066
4. Rahimi, M., & Arabgol, F. (2018). DNA-Based Cryptography and Security Perspectives. In Handbook of Research on DNA Computing and Molecular Programming (pp. 425-449). IGI Global. DOI: 10.4018/978-1-5225-5109-6.ch019

5. Zhang, X., Li, C., Dong, X., & Zhang, C. (2019). A novel DNA encryption method based on quantum chaotic system. IEEE Access, 7, 175108-175114. DOI: 10.1109/ACCESS.2019.2950362
6. Srinivas, R. R., Gudi, V. R., & Mohan, C. G. (2021). A Novel Approach for Secure DNA Cryptography Using a DNA Molecule Structure and Pseudorandom Number Generator. International Journal of Bio-Inspired Computation, 1(1), 37-53. DOI: 10.4018/IJBIC.2021010103
7. Nielsen, C. B., Cantor, C. R., & Smith, C. L. (2010). Optimizing the encoding of nucleotide sequences for DNA computation. Journal of Physical Chemistry B, 114(49), 16481-16490. DOI: 10.1021/jp1071435
8. Wang, Y., Li, S., & Gao, J. (2013). Research on DNA computing-based secure communication. Mathematical Problems in Engineering, 2013, 475307. DOI: 10.1155/2013/475307
9. Kocar, T., & Guven, B. (2016). Secure image encryption with DNA sequence. 2016 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), 1-5. DOI: 10.1109/CyberSecPODS.2016.7502310
10. Esmaeilzadeh, M., Pourkhalili, F., & Lighvan, M. Z. (2017). DNA sequence encryption based on chaos. IET Information Security, 11(4), 199-205. DOI: 10.1049/iet-ifs.2016.0451
11. Dong, S., Qi, J., Han, J., & Yang, J. (2018). A novel DNA encryption algorithm based on bit-level permutation. Journal of Biomedical Informatics, 87, 1-7. DOI: 10.1016/j.jbi.2018.07.016
12. Xia, Y., Zhang, X., & Zhang, C. (2018). DNA cryptography based on bio-image. Soft Computing, 22(17), 5855-5864. DOI: 10.1007/s00500-017-2925-3
13. Cui, L., Wang, X., & Hu, Y. (2020). DNA cryptography based on encoding of DNA sequences with enhanced error correction. IEEE Access, 8, 145421-145429. DOI: 10.1109/ACCESS.2020.3015649
14. Jin, X., Zhu, L., & Huang, X. (2021). A novel DNA sequence encryption algorithm based on DNA operations and quantum chaotic system. IEEE Access, 9, 21009-21015. DOI: 10.1109/ACCESS.2021.3054538
15. Hu, W., Zhang, X., & Yang, X. S. (2021). DNA cryptography based on chaotic sequences. IEEE Access, 9, 18123-18132. DOI: 10.1109/ACCESS.2021.3050251

ID: 21

MEDICAL IMAGING WITH DEEP LEARNING

Dr.M.CHITRA DEVI
Assistant Professor,
Department of Computer Science,
SRM Trichy Arts and Science College,
Tiruchirappalli, Tamil Nadu, India.

Dr.S.PADMAPRIYA
Associate Professor,
Department of Computer Science,
SRM Trichy Arts and Science College,
Tiruchirappalli, Tamil Nadu, India.

ABSTRACT:
Medical Imaging (MI) plays a pivotal role in diagnosing and treating various diseases. Deep Learning (DL) has emerged as a transformative technology in medical imaging, enabling improved image quality, accurate diagnosis, and enhanced treatment planning. This paper reviews recent advancements in medical imaging enhancement through the application of DL techniques. We explore the integration of Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) to address challenges in image denoising, super-resolution, and image synthesis. Through a complete analysis of the present state-of-the-art methodologies, we highlight the potential of DL in revolutionizing medical imaging, ultimately leading to better patient care and outcomes.

KEYWORDS: Conventional Neural Networks (CNN), Generative Adversarial Networks (GANs), Deep Learning (DL), Medical Imaging (MI), Image Enhancement.

I. INTRODUCTION

Medical imaging has undergone a significant transformation with the integration of deep learning techniques. Deep learning, a subset of machine learning, includes training complex artificial neural networks to perform tasks without explicit programming. This technology has revolutionized the way medical images are analyzed, interpreted, and utilized for various healthcare applications. One of the key applications of deep learning in medical imaging is image segmentation. Deep learning models are trained to accurately identify and delineate specific structures or regions of interest within medical images. This is particularly valuable for tasks such as tumour detection, organ localization, and blood vessel delineation. By automating the segmentation process, deep learning reduces the need for manual intervention, saving time and enhancing accuracy[1]. Another crucial application is image classification. Deep learning algorithms can classify medical images into different categories, aiding in the diagnosis of diseases or conditions. In radiology, for instance, deep learning models can accurately identify specific anomalies, allowing radiologists to make more informed decisions.

Object detection is another area where deep learning excels. It involves pinpointing and localizing specific objects within medical images, such as detecting nodules in lung CT scans or identifying micro calcifications in mammograms. This capability enables earlier and more accurate disease detection, contributing to improved patient outcomes. Deep learning also plays a role in image generation. Generative models, like Generative Adversarial Networks (GANs), can create synthetic medical images that closely resemble real ones. These synthesized images can be used for training purposes, augmenting datasets, and simulating rare conditions, aiding in the development of robust models.

One of the key benefits of deep learning in medical imaging is its ability to learn relevant features directly from raw image data. This eliminates the need for manual feature extraction and engineering, streamlining the analysis process. Moreover, deep learning models excel at recognizing intricate patterns, making them particularly effective at identifying subtle abnormalities that might be missed by human observers.

II. KEY APPLICATIONS OF DEEP LEARNING IN MEDICAL IMAGING:

A. Image Segmentation:

Image segmentation, a critical task in medical imaging, involves delineating specific structures or regions of interest within an image. Deep learning techniques have brought transformative advancements to this area, enabling accurate and efficient segmentation of anatomical structures, tumors, lesions, and other features in medical images. Accurate segmentation is a fundamental step in many medical applications, including diagnosis, treatment planning, and research. Traditional methods often require manual intervention and expert knowledge for feature extraction, making them time-consuming and subject to variability. Deep learning addresses these challenges by automating the segmentation process through data-driven feature learning.

Leveraging Convolutional Neural Networks (CNNs):

At the forefront of deep learning for image segmentation are Convolutional Neural Networks (CNNs). These neural networks are exactly designed to handle image data and have revolutionized the field. CNNs employ convolutional layers to automatically extract pertinent features directly from the raw images. Their ability to capture both local and global patterns makes them particularly well-suited for medical image segmentation tasks [2].

The U-Net Architecture:

A prominent architecture within CNNs for medical image segmentation is the U-Net [3]. This architecture is characterized by a U-shaped structure with both a constricting path (encoder) to internment context and an expanding path (decoder) for accurate localization. The use of skip connections between equivalent layers in the encoder and decoder allows the network to retain fine-grained details during segmentation.

Training and Data Augmentation:

Training deep learning methods for medical image segmentation requires substantial labeled data. However, acquiring such data can be challenging due to the need for expert annotations and privacy concerns. Data augmentation techniques, such as rotations, flips, and scaling, help mitigate the scarcity of labeled data by artificially expanding the training dataset. This augmentation improves the model's ability to generalize to various real-world scenarios.

Semantic and Instance Segmentation:

Deep learning enables two primary types of segmentation: semantic and instance segmentation. Semantic segmentation involves assigning a different class label to each pixel, essentially dividing the image into different regions based on their characteristics. In contrast, instance segmentation goes a step additional by not only identifying regions but also distinguishing between discrete instances of the same class.

B. Image Classification:

Image classification involves assigning labels to images based on their content. In medical imaging, accurate and efficient classification is pivotal for identifying diseases, conditions, and abnormalities. Traditionally, this process required manual interpretation by clinicians, a time-consuming endeavor with inherent subjectivity. The advent of deep learning has alleviated these challenges, enabling automated and highly accurate image classification.

Harnessing Convolutional Neural Networks (CNNs):

At the heart of deep learning for image classification in medical imaging are Convolutional Neural Networks (CNNs). CNNs are tailored to handle image data and have proven to be immensely effective. They employ convolutional layers that automatically extract relevant features from raw images, mimicking the human visual cortex's hierarchical processing [4].

The Role of Transfer Learning:

Transfer learning, a key strategy in deep learning, has been pivotal in medical image classification. Pre-trained CNN models, initially trained on massive datasets for general image recognition tasks, are fine-tuned on medical images. This approach leverages the learned features from the broader dataset to boost the performance of specific medical image classification tasks, even with limited labeled medical data.

C. Object Detection:

Object detection involves identifying and localizing specific objects or regions of interest within images. In medical imaging, this capability is instrumental in swiftly and accurately pinpointing structures, abnormalities, or anomalies that might otherwise be overlooked. Traditional object detection methods often necessitated manual inspection and annotation, a laborious and subjective process. Deep learning-driven object detection mitigates these challenges by automating the task, facilitating quicker and more reliable analysis.

Convolutional Neural Networks (CNNs) and Their Role:

At the heart of deep learning-based object detection lies Convolutional Neural Networks (CNNs), specialized neural networks tailored for image data. CNNs employ convolutional layers to automatically extract intricate features from images, enabling them to comprehend complex patterns and details that are crucial for accurate object localization.

Single-Shot Object Detection and Faster R-CNN:

In medical imaging, single-shot object detection and region-based object detection approaches, like Faster R-CNN (Region Convolutional Neural Network), have garnered significant attention. Single-shot object detection networks, such as YOLO (You Only Look Once), offer real-time detection capabilities by predicting object bounding boxes and class labels directly. Region-based methods, on the other hand, employ a two-step process involving proposal generation and subsequent classification and refinement. Faster R-CNN, a prominent example, demonstrates impressive accuracy in object localization.

D. Image Generation:

Image generation includes creating new images based on existing data patterns. In medical imaging, this process has gained significance for its potential to synthesize realistic images, augment datasets, and enhance image quality [6]. Traditional methods for image generation often relied on mathematical models or simple interpolation, which might not capture complex variations in medical data.

Generative Adversarial Networks (GANs) and Variants:

At the heart of deep learning-based image generation are Generative Adversarial Networks (GANs), a class of neural networks designed to generate new data samples. GANs contain two components: a generator and a discriminator. The generator yields synthetic images, while the discriminator estimates the authenticity of generated images in comparison to real ones. Through adversarial training, the generator strives to produce images that are indistinguishable from real ones.

Conditional GANs in Medical Imaging:

Conditional GANs (cGANs) enhance image generation by allowing control over the features of generated images. In medical imaging, cGANs can synthesize images with specific characteristics, such as variations in disease stages, anatomy, or imaging modalities. This capability is particularly valuable for training robust models and generating diverse datasets.

Data Augmentation and Imbalanced Data:

Deep learning-based image generation contributes to data augmentation, a technique that artificially expands the training dataset by generating new samples. In medical imaging, where labeled data can be scarce, data augmentation aids in training more accurate models. It is especially helpful in addressing class imbalance issues, where certain conditions or anomalies are underrepresented in the dataset.

E. Image Enhancement and Restoration:

Image enhancement and restoration are pivotal processes in medical imaging, aiming to improve image quality, enhance details, and reduce noise or artifacts. Traditional methods often involve complex algorithms that might not adequately capture intricate patterns and variations present in medical images. Deep learning has emerged as a game-changer in this domain by allowing models to learn complex image transformations directly from data.

Convolutional Neural Networks (CNNs) for Image Enhancement:

At the heart of deep learning for image enhancement and restoration are Convolutional Neural Networks (CNNs). These networks are optimized to process image data and have demonstrated remarkable efficacy in improving image quality. By training on pairs of noisy and clean images, CNNs can learn to map noisy images to their corresponding clean versions, effectively removing noise and enhancing details.

Denoising and Noise Reduction:

Noise reduction is a crucial aspect of medical image enhancement, particularly in noisy modalities like X-rays or MRIs. Deep learning models are trained to differentiate between noise and relevant features in images. Denoising autoencoders and CNN-based architectures have shown success in removing noise while preserving clinically relevant information.

Super-Resolution Imaging:

Deep learning-based super-resolution techniques are employed to enhance the resolution of medical images. These techniques predict high-resolution details from low-resolution input images, enabling visualization of finer structures. Super-resolution is particularly useful in applications like microscopy and enhancing the visualization of intricate anatomical features.

Artifact Correction and Removal:

Medical images are prone to various artifacts due to imaging hardware or patient motion. Deep learning models can be taught to detect and correct these artifacts, leading to cleaner and more accurate images. For instance, models can be designed to identify and rectify motion blur or streak artifacts in CT or MRI scans.

F. Registration and Fusion:

Image registration involves aligning multiple images to enable direct comparison or fusion. Fusion, on the other hand, merges complementary information from different images, enhancing the overall understanding of the subject. Traditional registration and fusion methods often rely on intricate mathematical algorithms, and they might struggle to handle complex variations in medical images. Deep learning has developed as a powerful tool to address these challenges.

Convolutional Neural Networks (CNNs) for Image Registration:

Convolutional Neural Networks (CNNs), designed for image data, have been pivotal in image registration tasks. CNNs learn relevant image features and spatial relationships, facilitating accurate and automated image alignment. By training on pairs of images, CNN-based registration models can capture intricate anatomical variations or changes over time.

Deformable Image Registration:

Deep learning has enabled the development of deformable image registration methods. These techniques go beyond rigid or affine transformations, allowing images with varying anatomies or distortions to be accurately aligned. Deformable image registration is particularly valuable in tasks involving organ motion or deformation, such as cardiac or respiratory motion compensation.

Multi-Modal Image Fusion:

Deep learning-driven multi-modal image fusion combines information from different imaging modalities, such as MRI and CT scans[5]. Fusion enhances the diagnostic value by providing a comprehensive view of anatomical and functional information. Deep learning models are trained to identify relevant features from each modality and combine them effectively, aiding in accurate diagnosis and treatment planning.

III. CONCLUSION:

Deep learning has brought revolutionary advancements to medical imaging by enabling more accurate and efficient analysis of medical images. Its potential to improve diagnostic saccuracy, treatment planning, and patient outcomes makes it a crucial area of research and development in the field of healthcare. However, ongoing efforts are needed to address challenges related to data quality, interpretability, and ethical considerations before deep learning becomes fully integrated into clinical workflows.

IV. REFERENCES:

1. K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
2. J. Long, E. Shelhamer, T. Darrell, "Fully Convolutional Networks for Semantic Segmentation", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
3. O. Ronneberger, P. Fischer, T. Brox, “U-

Net: Convolutional Networks for Biomedical Image Segmentation", International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015.

4. A. Krizhevsky, I. Sutskever, G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", Advances in Neural Information Processing Systems (NIPS), 2012.
5. E. Topol, "High-Performance Medicine: The Convergence of Human and Artificial Intelligence", Nature Medicine, 2019.
6. L. Tang, H. Li, L. Gao, "Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique", European Journal of Radiology, 2017.

ID: 22

THE TRANSFORMATIVE ROLE OF MACHINE LEARNING IN BIOINFORMATICS: REVOLUTIONIZING DATA ANALYSIS, INSIGHTS, AND DISCOVERY

Ms.R.AMUTHA
Assistant Professor,
Department of Computer Science,
PSG College of Arts & Science,
Coimbatore, Tamil Nadu, India.

Mr.S.KARTHIK
Assistant Professor,
Department of Computer Science,
PSG College of Arts & Science,
Coimbatore, Tamil Nadu, India.

1. INTRODUCTION

Machine learning has revolutionized various fields, and bioinformatics is no exception. Bioinformatics involves the application of computational techniques to analyse biological data, and machine learning plays a crucial role in extracting meaningful insights from complex biological information. Bioinformatics involves the integration of biology, computer science, and statistics to analyze and make sense of biological data in an interdisciplinary manner. It encompasses various areas such as genomics, proteomics, and structural biology. Machine learning algorithms empower computers to grasp patterns and correlations from data autonomously, without requiring explicit programming.

Sequence the human genome in just one day because of improvements in DNA sequencing methods. In conventional methods, it would have taken about ten years. This is only one of the many significant advances that machine learning has made to bioinformatics.

In the rapidly evolving landscape of modern biology and medicine, the integration of cutting-edge technologies has led to an explosion of data, transforming how we understand and approach complex biological systems. One of the pivotal tools driving this transformation is machine learning. With its capacity to extract patterns, discern relationships, and make predictions from massive and heterogeneous datasets, machine learning has become an indispensable asset in the realm of bioinformatics.

Bioinformatics, the interdisciplinary field at the crossroads of biology, computer science, and data analysis, is tasked with unravelling the intricacies of biological information. From deciphering the human genome's secrets to predicting protein structures and elucidating disease mechanisms, bioinformatics tackles challenges that were once considered insurmountable. However, the sheer volume and complexity of biological data pose significant hurdles that traditional methods struggle to surmount.

This is where machine learning steps in as a transformative force. By harnessing the power of algorithms that adapt and learn from data, machine learning algorithms excel at uncovering hidden patterns, capturing nonlinear relationships, and making predictions in complex and noisy datasets. In the context of bioinformatics, machine learning plays a pivotal role in areas such as genomics, proteomics, structural biology, and drug discovery. The market for AI in bioinformatics is expanding because of the large number of biotech organisations that employ ML consultants to streamline the handling of biological data. It anticipated increasing at a CAGR of 42.7% from 2022 to 2029, reaching $37,027.96. This article introduces machine learning (ML), outlines how it aids biomedical research, and lists potential difficulties you can encounter when implementing this technology.

A part of the larger field of artificial intelligence (AI) is machine learning. Systems can now do activities for which they did not specifically programmed and independently learn from data. Its objective is to enable machines to carry out operations like diagnosing, planning, and anticipating that call for human intelligence. In bioinformatics, machine learning helps researchers make predictions, classifications, and discover hidden patterns in vast biological datasets.

Machine learning comes in two primary flavours:

Machine learning comes in two primary flavours: supervised learning and unsupervised learning. These two approaches represent distinct paradigms within the field of machine learning, each with its own objectives, methods, and applications.

1. Labelled datasets are used in supervised learning to educate algorithms how to use an existing classification system and how to make predictions using it. Decision trees and neural networks are trained using this ML type. The goal of supervised learning is to generalize from the training data in a way that enables the model to accurately predict outcomes for new, unseen data.

2. Dataset with no labels used in unsupervised learning. Instead, algorithms look for patterns in data on their own. In other words, they pick up knowledge that we are unable to give directly. Unsupervised learning is particularly useful when exploring datasets to gain insights and find hidden structures. It is possible to combine labelled data with unlabelled is known as semi supervised learning. Machine learning can be useful while we don't have labelled data for supervised learning approach but still want to use it to direct the learning process.

2. MACHINE LEARNING TECHNIQUES USED IN BIOINFORMATICS

2.1 NATURAL LANGUAGE PROCESSING

Machine learning techniques in bioinformatics contribute to understanding biological processes, accelerating drug discovery, and enabling personalized medicine. These techniques harness the power of data-driven approaches to solve intricate challenges in the life sciences. Some of the algorithms used with either supervised learning or unsupervised learning, while others can be used with both. Processing language naturally, unstructured human discourse can be understood by a variety of tools known as natural language processing (NLP). NLP has the ability to search through vast amounts of biological research, compile data on a particular subject from numerous sources, and translate research results from one language to another. NLP tools are also capable of parsing pertinent biomedical databases in addition to mining research publications. NLP relies heavily on machine learning techniques, including neural networks, recurrent neural networks (RNNs), transformers, and more. Recent advancements, particularly with models like GPT-3, have pushed the boundaries of NLP and enabled more sophisticated language understanding and generation capabilities.

The following are some ways that NLP can help the discipline of bioinformatics - Analysing DNA expression arrays, deciphering genetic variations, annotating protein activities, and searching for novel therapeutic targets. By harnessing NLP's capabilities, bioinformatics researchers can process and leverage the vast amount of textual information available, leading to more comprehensive and informed insights into biological systems, accelerating discoveries, and enhancing the efficiency of research and clinical practice

2.2 NEURAL SYSTEMS

In the context of machine learning, the term "neural systems" usually refers to Artificial Neural Networks (ANNs), refer to computational models motivated by the architecture and functioning of biological neural systems. The building blocks of this multi-layered structure are nodes and neurons. Linkages connect neurons in adjacent layers to one another; however linkages between neurons in the same layer do not exist. Neurons in the input layer take in information, process it, and then transmit it as an input to the layer below. In addition, this procedure keeps going until the output layer receives the processed data.

Perceptron is the name of the most fundamental neural network. It comprises of a single classifier neuron. A linear discrimination function is used by this neuron to categorise input into one of two classes. There is no restriction on the number of layers or nodes in a layer in larger neural networks.

Neural systems have showcased remarkable proficiency across diverse tasks, spanning from image and speech recognition to natural language processing and even competitive gaming. Their ability to automatically learn intricate patterns from data has propelled them to the forefront of modern machine learning research and applications.

In this context, W denotes weights and b represents biases. When all neurons in the input layer are activated, each input neuron multiplies its corresponding weight from the matrix, and the resulting outputs are aggregated with a bias term. This sum is then transmitted to an adjoining hidden layer. While the input-output mechanism follows a similar pattern in successive hidden layers, direct links between neurons within the same layer are typically absent. Activation functions quantify the strength of connections between neighboring neurons across these hidden layers.

In the initial stages of training using batches of data samples, model parameters and attributes can be adjusted through various learning approaches, incorporating suitable activation and rectification functions. Subsequently, the trained network should undergo testing and validation using different sets of data samples to ensure strong robustness and favorable predictive capability. These processes are commonly known as model testing and validation, essential for enhancing the network's reliability and accuracy.

After training completed, the neural network can perform regression or classification task on testing data, while there usually exists the difference between the predicted outputs and actual values. And the difference should be minimized to acquire optimal model performance.

2.3 CLUSTERING

The technique of grouping pieces into different groups based on the given criteria of similarity is known as unsupervised clustering. As a result of this classification, elements in one cluster have a close relationship with one another and are distinct from those in other clusters.

In contrast to supervised classification, we cannot predict how many groups would develop during clustering. One well-known use of this machine learning strategy in bioinformatics is the clustering of genes with comparable expression levels in microarray-based expression profiling of genes. Clustering is a foundational component of unsupervised machine learning, entailing the task of categorizing akin data points into distinct clusters or groups. The objective of clustering is to determine inherent patterns, relationships, or structures in the data without the need for predefined labels. The primary goal of clustering is to reveal hidden structures within data for further analysis or insights.

It's an unsupervised learning approach that helps in exploratory data analysis, data segmentation, and pattern recognition. Various clustering algorithms, such as K-Means, Hierarchical Clustering, and DBSCAN, enable the identification of natural groupings or clusters in data. Clustering finds applications in diverse fields, including customer segmentation, image analysis, gene expression profiling, and anomaly detection, aiding in understanding data and making informed decisions.

Clustering is a versatile technique that plays a vital role in uncovering patterns and insights within data. By grouping similar data points, clustering aids in understanding data structure, segmentation, and exploratory analysis, contributing to informed decision-making in various fields.

2.4 DIMENSIONALITY REDUCTION

Classifications are carried out in Machine Learning classification problems based on criteria or characteristics. Sometimes there are too many variables that influence the outcome, making it challenging to visualise and handle the dataset. By reducing the amount of features, dimensionality reduction algorithms can make the dataset easier to handle. For instance, humidity and rainfall can be two characteristics of a problem involving climate classification. For the sake of convenience, these two can be combined into one component since they are both closely related.

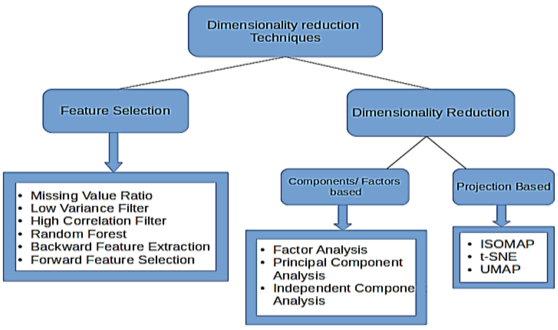


Figure: 2.4.1 Dimensionality Reduction Techniques

Dimensionality reduction stands as a pivotal technique within data analysis and machine learning, encompassing the art of diminishing the count of input features or variables within a dataset while retaining its fundamental information intact. This process is particularly valuable when working with high-dimensional data, as it helps improve computational efficiency, mitigate the curse of dimensionality, and enhance model performance.

Dimensionality reduction consists of two primary parts:

- Feature choice - Through embedding, filtering, or wrapping features, selects a subset of variables to represent the full model.
- Extracting features - reduces the dataset's dimension count. For example, two 2D spaces can be created from a 3D space.

Dimensionality reduction is a valuable technique for simplifying and enhancing the analysis of high-dimensional data. It offers benefits in terms of computational efficiency, model performance, and visualization, although careful consideration should be given to the choice of method and potential information loss.

2.5 DECISION TREE CLASSIFIERS

One of the most well liked traditional supervised learning classifiers is this one. These methods use a recursive method to create a tree model that resembles a flowchart, with each node denoting a test on a feature. The procedure first selects the top node, or the root, and then recursively constructs the tree while considering each parameter. Each sequence's last node is referred to as "the leaf node." It retains the class designation and stands for the final classification.

Decision tree classifiers are powerful machine learning models used for both classification and regression tasks. They work by recursively partitioning the input data into subsets based on feature values, ultimately leading to the prediction of a target variable. Decision trees are particularly advantageous due to their interpretability, ability to handle both categorical and numerical features, and their suitability for complex and nonlinear relationships.

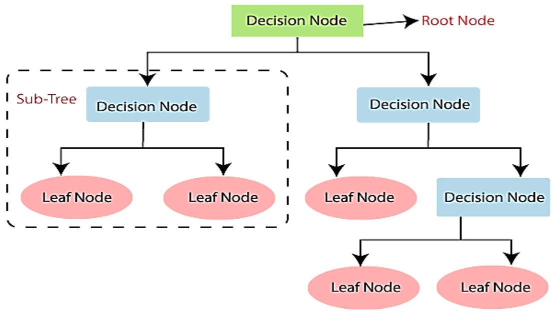


Figure: 2.5.1 Decision Tree Algorithm in Machine Learning

Decision tree models can accomplish classifications without requiring a lot of processing after training, but they do require substantial computational power during training. The key benefit that these classifiers offer the bioinformatics community is that they produce rules that are clear and results that can be explained. Decision tree classifiers are versatile and interpretable models that offer insights into data patterns and relationships. While they can suffer from overfitting and instability, techniques like pruning and ensemble methods mitigate these issues, making decision trees a valuable tool in the machine learning toolbox.

2.6 SUPPORT VECTOR MACHINE

This supervised machine-learning model can address two-group classification issues. These algorithms search for an ideal hyperplane that divides the data into two classes with the smallest distance between the data points in order to categorise the data points.

A Support Vector Machine (SVM) emerges as a potent and extensively applied machine-learning algorithm catering to both classification and regression endeavours. SVMs thrive in unravelling intricate challenges by determining the paramount hyperplane that optimally segregates data into discrete classes or accurately predicts continuous values. Their proficiency extends to high-dimensional realms and scenarios wherein data defies linear separation.

Different classifications of points can be found on either side of the hyperplane. The size of the hyperplane is a function of the quantity of features. The decision boundary is either a line for two features or a 2D plate for three features. Using SVM for classifications with more than three features is challenging because to this trait. The computational identification of functional

ISBN: 978-81-965283-7-9

RNA genes benefits from this method. Based on information about their expression, it can choose the best possible set of genes for cancer diagnosis.

Support Vector Machines are versatile algorithms that can efficiently solve both classification and regression tasks by finding optimal hyperplanes or decision boundaries. They leverage the concept of support vectors and kernel functions to capture intricate relationships in the data, making them a valuable tool in various machine learning applications.

III. APPLICATIONS OF MACHINE LEARNING IN BIOINFORMATICS

Machine learning has various applications in diverse fields, ranging from natural language processing to healthcare. Before machine learning, Bio informatics field faced lot of problems in extracting valuable insights from large datasets. However, today Machine-learning techniques helps to learn the feature of complex datasets and present in an easy manner. This article discuss most commonly used ML techniques in various applications bio informatics applications.

3.1. FACILITATING GENE EDITING EXPERIMENTS

Gene editing is the process of changing a portion of an organism's DNA sequence by deleting, adding, or replacing it. Typically, the CRISPR technique, which is quite potent, is used in this process. However, there is still much room for improvement when it comes to choosing the best DNA sequence to manipulate, and this is where ML might be useful. Researchers can optimise the design of gene editing studies and forecast their results by using machine learning for bioinformatics.

A research team using machine-learning techniques found the most effective combinations of amino acid residues that enable the genome-editing protein Cas9 to interact with the target DNA. Such an experiment would have been too big given the quantity of these variants, but employing an ML-driven like Artificial Neural Networks (ANN), Support Vector Machines (SVM), and random forests (RF)] for the prediction of the CRISPR-Cas9 cleavage sites that will be cleaved by a particular RNA.

3.2. IDENTIFYING PROTEIN STRUCTURE

The study of proteins, their relationships, chemical makeup, and physiological functions is known as proteomics. This field is computationally intensive and uses large biological datasets. For this reason, bioinformatics technologies like machine learning are crucial.

The positioning of amino acids in proteins into the three groups of sheet, helix, and coil using convolutional neural networks is one of the most effective applications in this subject. The theoretical limit of neural networks' accuracy ranges are higher.

Protein model scoring, a crucial task for predicting protein structure, is another application of ML in proteomics. Researchers from Fayetteville State University used machine learning (ML) in their bioinformatics method to enhance protein model score. They utilised an ML to group the protein models in question.

3.3. SPOTTING GENES ASSOCIATED WITH DISEASES

Researchers are increasingly using machine learning in bioinformatics to identify the genes that are possibly to blame for particular illnesses. This is accomplished through the gene expression microarray analysis and RNA sequencing. In research involving cancer, gene identification used to classify tumours by molecularly analysing them and discover genes that are likely to contribute to cancer.

For instance, to evaluate their capacity to predict and categorise cancer kinds, a team of researchers from the University of Washington used a variety of machine learning in bioinformatics methods, including decision trees, support vector machines, and neural networks. Researchers used RNA sequencing data from The Cancer Genome Atlas project and found that the most accurate method for classifying cancers was the linear support vector machine.

Another example is the classification of different forms of breast cancer using machine learning and data on gene expression. This group also used information from the Cancer Genome Atlas project. The samples were divided into triple negative breast cancer, one of the worst types, and non-triple negative breast cancer by the researchers. The Support Vector Machine classifier once more produced the best outcomes.

In relation to non-cancerous illnesses, scientists at the University of Pennsylvania used machine learning to find genes that would be good targets for medications to treat coronary artery disease (CAD). The group identified many single nucleotide polymorphisms (SNPs) connected to CAD using the ML-powered Tree-based Pipeline Optimisation Tool (TPOT). They examined the UK Biobank's genetic data and discovered 28 relevant SNPs.

3.4. TRAVERSING THE KNOWLEDGE BASE IN SEARCH OF MEANINGFUL PATTERNS

Every 2.5 years, cutting-edge sequencing technology doubles the size of genomic databases, and academics are trying to figure out how to make use of this accumulated knowledge. Machine learning in bioinformatics can comb through biological reports and publications to distinguish between various genes and proteins and look up their functions. Additionally, by adding the knowledge it gathers from the literature, it can help annotate protein databases.

One example comes from a team of researchers that used bioinformatics and machine learning to speed up protein model scoring through literature mining. Protein-protein docking structural modelling often yields a number of models that rated according to structural restrictions. In order to find residues that potentially contribute to the generation of these constraints for model scoring, the team employed ML algorithms to search through PubMed papers on protein-protein interactions. Additionally, researchers investigated the capacity of several machine-learning algorithms to evaluate the relevance of all found residues in order to ensure that the constraints are applicable.

3.5. REPURPOSING DRUGS

Drug repurposing, or reprofiling, is a technique scientists use to discover new applications of existing drugs that they were not intended for. Researchers use artificial intelligence (AI) in bioinformatics to analyse drugs on pertinent databases like BindingDB and DrugBank. Drug repurposing can be divided into three main categories:

- Drug-drug interactions research into how different medications interact when taken together.
- Drug-target interactions examine how well a drug can bind directly to the target protein.
- In order to identify hotspots and allosteric locations, protein-protein interaction probes the surface of interacting intracellular proteins.

A deep neural network algorithm created by researchers from China University of Petroleum and Shandong University, and it was applied to the DrugBank database. They sought to investigate the interactions between medication compounds and mitochondrial fusion protein 2 (MFN2), one of the key proteins thought to be responsible for the development of Alzheimer's disease.

4. CHALLENGES PRESENTED BY MACHINE LEARNING IN BIOINFORMATICS

Machine learning in bioinformatics differs from ML in other sectors due to the four factors below, which also constitute the main challenges of applying ML to this field.

4.1 Bioinformatics AI is expensive. For the algorithm to perform properly, we need to acquire a large training dataset. However, it’s rather costly to obtain 10,000 chest scans, or any other type of medical data for that matter.

4.2 Difficulties associated with the training datasets. In other fields, if we do not have enough training data, we can generate synthetic data to expand your dataset. However, this trick might not be appropriate when it comes to human organs. The problem is that our scan generation software might produce a scan of a real human. In addition, if we start using that without the person’s permission, we will be in gross violation of their privacy. Another challenge associated with the training data is that if we want to build an algorithm that works with rare diseases, there will not be much data to work with in the first place.

4.3 The confidence level must be very high. When human life depends on the algorithm’s performance, there is just too much at stake, which does not leave room for error.

4.4 Explain ability issue. Doctors will not be open to using the ML model if they do not understand how it produced its recommendations. We can use explainable AI instead, but these algorithms are not as powerful as some black-box unsupervised learning models.

4.5 Interpretability: Many machine-learning algorithms, especially deep learning models, are considered "black-boxes." In bioinformatics, interpretability is crucial to understand the biological implications of the models' predictions, especially when dealing with gene expression patterns and identifying biomarkers.

4.6 Biological Complexity: Biological systems are intricate and multifaceted, making it challenging to model accurately. Machine learning models may oversimplify these complexities, leading to models that do not fully capture the underlying biological mechanisms.

4.7 Domain Expertise: Effective application of machine learning in bioinformatics requires collaboration between data scientists and domain experts. Domain knowledge is essential for formulating meaningful features, validating results, and ensuring biological relevance.

4.8 Ethical Considerations: Handling sensitive biological data raises ethical concerns about data privacy, consent, and potential misuse. Ensuring compliance with ethical standards and regulations is crucial in bioinformatics.

The domains of biology and medicine can use AI and ML in many different ways. Another area of medicine that utilises ML and AI-based medical solutions is bioinformatics. Managing enormous amounts of diverse data, including gene sequences, protein structures is necessary for bioinformatics. Although many AI bioinformatics models are expensive to run, ML is well renowned for its data processing capabilities.

V. REFERENCES

1. Binhua Tang, Zixiang Pan, Kang Yin,Asif Khateeb(2019). Recent Advances of Deep Learning in Bioinformatics and Computational Biology, frontiers in Genetics, Vol 10, doi.org/10.3389/fgene.2019.00214
2. Noam Auslander, Ayal B. Gussow, Eugene V. Koonin(2021), Incorporating Machine Learning into Established Bioinformatics Frameworks, International Journal of Molecular Sciences, 2021, doi: 10.3390/ijms22062903
3. Siraj-Ud-Doulah(2019), Application of Machine Learning Algorithms in Bioinformatics, Bioinformatics & Proteomics Open Access Journal, Volume 3 Issue 1.
4. K.Jayanthi, C. Mahesh (2019), Need of Machine Learning In Bioinformatics, International Journal of Advances in Electronics and Computer Science, Volume-6, Issue-10, Oct.-2019.

ID: 23

DATA SECURITY

Dr.S.RIZWANA
Assistant Professor & Head,
Department of Computer Science (Unaided),
Erode Arts and Science College,
Erode, Tamil Nadu, India.

Dr.M.MENAKA
Assistant Professor & Head,
Department of Computer Technology,
Erode Arts and Science College,
Erode, Tamil Nadu, India.

ABSTRACT:
Data security is a critical aspect of modern information systems. It encompasses the protection of data from unauthorized access, ensuring confidentiality, availability, and integrity. This abstract provides an overview of the importance of data security, highlighting its significance in safeguarding sensitive information. It explores various types of data security technologies and common practices employed to mitigate risks. Additionally, it touches upon emerging trends in data security, highlighting the need for organizations to stay vigilant in the face of evolving threats. By prioritizing data security measures, organizations can mitigate vulnerabilities, protect their assets, and maintain the trust of their stake holders.

KEYWORDS: *Integrity, Confidentiality, Encryption, Authentication, Masking*

I. INTRODUCTION

In today's digital era, where information holds immense value, the topic of data security has gained significant importance within the field of computer science. Data security revolves around the methods and steps taken to safeguard digital information from unauthorized access, misuse, exposure, disruption, tampering, or loss. With rapid technological progress shaping our world, ensuring the protection, trustworthiness, and accessibility of data has risen to a position of utmost significance.

KEY CONCEPTS OF DATA SECURITY

1. **Confidentiality**: Data confidentiality ensures that sensitive information is only accessible to authorized individuals or entities. Encryption techniques play a pivotal role in securing data by converting it into a coded format that can only be deciphered using the appropriate decryption key.
2. **Integrity**: Data integrity focuses on maintaining the accuracy and trustworthiness of information. It involves measures to prevent unauthorized modifications or alterations to data. Hashing algorithms and digital signatures are used to detect any unauthorized changes to files.
3. **Availability**: Ensuring data availability involves measures to prevent disruptions or denial of access to information. Redundancy, backup systems, and disaster recovery plans are essential components of this aspect of data security.

WHY IS DATA SECURITY IMPORTANT?

1. **Protection Against Threats**: Data security shields sensitive information from cyber threats like hacking and malware, preventing data breaches and unauthorized access.
2. **Privacy Safeguarding**: Ensures personal data remains private,

preventing identity theft and maintaining individual privacy.

3. **Business Continuity**: Secures critical data for seamless operations, building trust among customers, partners, and stakeholders.
4. **Regulatory Compliance**: Adhering to data protection laws avoids legal penalties and maintains ethical business practices.
5. **Reputation Preservation**: Strong data security builds trust, preventing breaches that can damage an organization's reputation and credibility.

TYPES OF DATA SECURITY TECHNOLOGIES

Today, a variety of data security technologies are employed to counter diverse external and internal threats. Employing a multitude of these technologies is crucial for organizations to fortify all potential entry points of threats and ensure data protection. \

Here are a few of these methods:

1. Data Encryption:

Data encryption employs algorithms to transform information into an unreadable form, requiring authorized encryption keys for decryption. This serves as a final safeguard for confidential data, especially in breach scenarios. Crucially, secure storage of encryption keys is essential, with restricted access. Encryption also encompasses security key management functionalities.

2. Authentication:

Authentication verifies user login details against stored database information, encompassing usernames, passwords, biometrics, and more.

It stands as the initial barrier to unauthorized access, offering paramount protection for sensitive data. Simplified by technologies like single sign-on, multi-factor authentication, and breached password detection, authentication ensures robust security while preserving user convenience.

3. Data Masking:

Data masking conceals data partially or wholly, safeguarding it from unauthorized access, both externally and internally. Personally identifiable information (PII), like phone numbers or emails, can be obscured, e.g., by masking the initial digits or letters.

Proxy characters are employed for masking, and the original data is restored only for authorized users. Data masking permits application development using genuine data, while maintaining data security.

4. Tokenization:

Tokenization is similar to data encryption but differs in that it replaces data with random characters, where encryption scrambles data with an algorithm. The "token," which relates to the original data, is stored away separately in a database lookup table, where it is protected from unauthorized access.

5. Data Erasure:

Data erasure occurs when data is no longer needed or active in the system. The erasure process uses software to delete data on a hardware storage device. The data is permanently deleted from the system and is irretrievable.

6. Data Resilience:

Data resilience is determined by the ability of an organization to recover from incidences of a data breach, corruption, power failure, failure of hardware systems, or loss of data. Data centers with backup copies of data can easily get back on their feet after a disruptive event.

7. Physical Access Controls:

Unlike digital access control, which can be managed through authentication, physical access control is managed through control of access to physical areas or premises where data is physically stored, i.e., server rooms and data center locations. Physical access control uses security personnel, key cards, retina scans, thumbprint recognition, and other biometric authentication measures.

COMMON DATA SECURITY RISKS

- **Human Error**: Many data breaches are caused by non-malicious human error that results in the exposure of sensitive data or information. From sharing or granting access to valuable data to losing or mishandling sensitive information, employees can trigger a data breach either by accident or because they are not fully briefed on enterprise security policies.
- **Social Engineering Attacks**: As a primary attack vector for

ISBN: 978-81-965283-7-9

cybercriminals, social engineering attacks manipulate employees into providing PII or access to private accounts. One of the most common forms of social engineering attacks is phishing.

- **Internal Threats**: Malicious or compromised insiders are employees, contractors, vendors, or partners who intentionally or inadvertently put your organization's data at risk. Malicious insiders actively try to steal data or harm your organization for personal gain, while compromised insiders go about their daily routine unaware that their account has been hacked.
- **Ransomware**: Ransomware is malware used by criminals to take over corporate devices and encrypt sensitive data. This data is only accessible with a decryption key that the cybercriminal owns, and they commonly only release this key if a ransom has been paid. Oftentimes, even when organizations pay the ransom, their data ends up being lost. Ransomware: The Good, The Bad, and The Ugly

DATA SECURITY MEASURES

1. **Firewalls and Intrusion Detection Systems (IDS)**: These technologies act as barriers between a network and potential threats. Firewalls filter incoming and outgoing network traffic, while IDS monitors for suspicious activities and alerts administrators to potential breaches.
2. **Multi-Factor Authentication (MFA)**: MFA enhances security by requiring users to provide multiple forms of verification, such as passwords and biometric data, before granting access.
3. **Encryption**: Encrypting data ensures that even if unauthorized users gain access, the information remains unreadable without the appropriate decryption key.
4. **Security Patching**: Regularly updating software and systems with security patches helps mitigate vulnerabilities that could be exploited by cybercriminals.
5. **User Training and Awareness**: Educating users about cybersecurity best practices is crucial. Employees need to understand the risks of phishing, social engineering, and the importance of maintaining strong passwords.

DATA SECURITY TRENDS

- ❖ Adapting to Evolving Threats: Addressing new and sophisticated cyber risks effectively.
- ❖ Proactive Defense: Anticipating vulnerabilities and implementing preventive measures.
- ❖ Compliance and Innovation: Aligning with regulations, integrating new technologies for robust protection.
  - **Zero Trust**: Verify all users and devices continuously, no matter their location.
  - **AI**: Predict threats, automate responses with AI and machine learning.
  - **Cloud Assurance**: Tailored security for data in the cloud
  - **Quantum Encryption**: Developing quantum-resistant encryption methods.
  - **MFA Evolution**: Enhanced multi-factor authentication with biometrics.
  - **Privacy Tech**: Secure data processing while maintaining privacy.
  - **Ransomware Protection**: AI-driven defense and secure backups.
  - **Regulatory Adherence**: Strong data governance for compliance.

CONCLUSION

Data security is a foundational aspect of computer science, underpinning the reliability and trustworthiness of digital interactions. As technological landscapes evolve, so do the challenges and opportunities in securing data. Organizations, individuals, and society at large must remain vigilant in adapting to the dynamic world of data security, ensuring that our digital assets remain safe from the ever-evolving threats of the digital age.

ID: 24

CLOUD DATABASE AND SECURITY

Dr.M.MENAKA
Assistant Professor & Head
Department of Computer Technology,
Erode Arts and Science College,
Erode, Tamil Nadu, India.

Dr.S.RIZWANA
Assistant Professor & Head
Department of Computer Science,
Erode Arts and Science College,
Erode, Tamil Nadu, India.

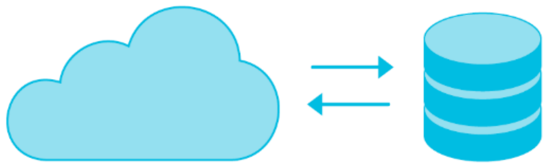
ABSTRACT

Cloud security is a set of security measures designed to protect cloud-based infrastructure, applications, and data. These safeguards protect data privacy by ensuring user and device authentication, data and resource access control, and data access control. Moreover, they also assist with keeping data compliance requirements. In essence, Cloud Database Security protects a company's data from data breaches, distributed denial of service (DDoS) attacks, viruses, hackers, and unauthorized user access or use in cloud environments. Because most businesses are currently adopting cloud computing in some way or another, cloud security has become quite crucial. IT professionals are still wary of shifting more data and apps to the cloud because of security threats, governance, and compliance challenges when data gets kept in the cloud. They are concerned that extremely sensitive data and intellectual property could get compromised due to unintentional leaks or more sophisticated cyber attacks. A significant cloud security component is business content and data protection to solve this issue. Preventing data leaks and theft is vital for retaining your customers' trust and safeguarding the assets that contribute to your competitive advantage. The capacity of cloud database security to protect your big data store and support is critical for any firm moving to the cloud, such as Google cloud storage.

KEYWORDS: *Cloud Database Security, Cloud-based infrastructure, significant cloud security, Cyber-attacks, and Google cloud storage.*

CLOUD DATABASE SECURITY

To understand cloud database security, the concept of a cloud database must first be defined. A cloud database is an organized collection of data managed and hosted in a system that lives on a cloud computing platform, which can be public, private, or a hybrid of the two. A database can be hosted on the cloud via Infrastructure-as-a-Service (IaaS) or Platform-as-a-Service (PaaS). Regardless of which cloud database service is being used, cloud databases are targets for breaches if modern solutions for database security in cloud computing are not in place. Legacy database security solutions focus on endpoint security technologies and network security, monitoring and protection, a defense strategy that does not defend against threats facing data once it has dispersed off-premises. Modern cloud data governance solutions address these threats by providing comprehensive data security controls modern cloud database security solutions provide the data security and governance necessary for data protection in cloud computing. Solutions for data security in cloud computing protect data by providing capabilities like centralized policy-based cloud access management, consistent real time observability, identity federation, unified granular authorization policies, least privilege, and continuous monitoring and data governance in the cloud. While cloud-based data services provide greater accessibility, they tend to leave companies in the dark regarding use of their data. Cloud security governance addresses this challenge by providing visibility into who has access to what data, what the users and applications are doing with the data, and if someone is trying to steal the data.



CLOUD DATABASE SECURITY USES

One of the greatest challenges of securing data in the cloud is the lack of a common standard for exchanging authentication and authorization information with the various data services. This results in a proliferation of shared accounts amongst engineering teams, diminished security for the DevOps and security teams, and a fragmented authorization model. This leaves companies exposed to a large security risk with a wider attack surface. Cloud database security solves this problem with the use of data security governance solutions, which enable identity federation to handle shared accounts; generate logs, metrics, and traces enriched with identity and context information for better visibility into who is accessing what data and what they did with it; and centralize authorization policies in a single place, eliminating the fragmented and broken authorization model. Cloud database security also extends authorization beyond the user level, down to the field level within the database. Users should not be able to access all information within a database e.g. social security numbers should be masked while names and addresses could be available.

COMMON THREATS TO CLOUD DATABASES

The most common cloud data security challenges are data breaches, account hijacking, and data loss. These threats may be carried out by outside attackers and/or malicious insiders

Data breaches: Companies without sophisticated cloud data security solutions are highly susceptible to data breaches, which result in costly non-compliance penalties and damaged customer trust. Nearly half of all data breaches in 2022 occurred in the cloud, a large portion of which was due to stolen or compromised credentials

Account hijacking: Phishing and exploitation of vulnerabilities in third-party software is used to steal login credentials, and results in attackers gaining access to and taking over a user's account, which exposes all files in the user's account and even other user accounts on the cloud

Data loss: Attackers that gain unauthorized access to private or sensitive data may not only view and use the data, but also delete the records completely. If all records are hosted on a central cloud-based datastore, deletions can spread to all user devices connected to the cloud at once

CLOUD DATABASE SECURITY IMPORTANT

Database security in the cloud is more important than ever. The increased adoption of cloud computing has made it all the more critical for businesses to invest in advanced cloud computing data security solutions that can handle the unique threats facing databases on the cloud. Unintentional leaks and sophisticated cyber attacks threaten to compromise sensitive information and intellectual property on the cloud, and security threats, governance challenges, and compliance challenges associated with cloud storage have left many IT professionals wary of shifting more data and apps to the cloud. Advanced data governance in cloud computing solves these issues by extending Identity and Access Management (IAM) controls that help organizations ensure that only authorized persons are behind every action undertaken down to the database level. Legacy database security solutions focus on a "fortress" defense strategy that builds walls and monitors all entries – this strategy is not appropriate for cloud computing. Sensitive data in cloud computing is dispersed across tens and hundreds of databases, data pipelines, and warehouses, which is why it is crucial to adopt modern cloud data governance tools that focus on merging identity management and data observability rather than on "fortress" security.

ISBN: 978-81-965283-7-9

CLOUD DATABASE SECURITY BEST PRACTICES

Organizations can proactively address most cloud database security issues by implementing the following cloud database security best practices

Implement a Culture of Shared Responsibility

In a private data center, the organization is responsible for all IT security issues. However, in a public cloud, IT security is also the responsibility of the cloud provider. In choosing a cloud vendor, businesses should review the company's shared security obligations and determine who is responsible for certain aspects of cloud security. In addition to preventing misunderstandings and misinterpretations, clarity about responsibilities will help prevent security incidents caused by specific security needs falling through the cracks

Cloud Provider about Security

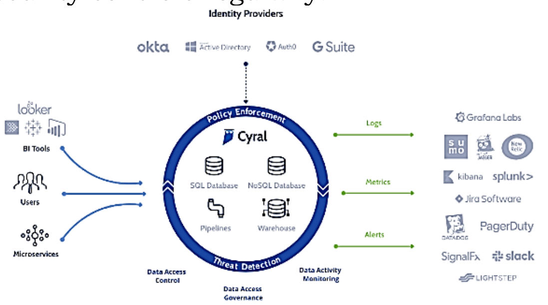
In addition to defining shared responsibilities, enterprises should inquire about their public cloud vendors' security measures and processes. It may seem easy to assume that the largest vendors have their security under control, but security procedures and methods can differ dramatically from one vendor to another. Among the questions you should ask are whether or not the solution provides role-based access as well as identity and access management.

Develop and Enforce Cloud Security Policies

You should have a written policy that explains who can use cloud services, how they can be used, and what data can be stored there. It should also specify how employees should protect data and apps stored in the cloud.

CLOUD DATABASE SECURITY AND CYRAL

Cyral's cloud database security solution enables organizations to seamlessly integrate identity management and data observability, facilitating observation of any data access attempt on any repository, with the full user context, at any given time. Cyral's cloud data access governance ensures that any access attempt triggers an instant matching of the user with their IAM groups, a reference to a single source of policy rules, and the delivery of a password through a password storage solution. All data sources can be monitored in real time without a negative impact on performance, and a single, rich source of logs is available for audits, compliance requirements, and forensics. This information can be used by security and DevOps teams for troubleshooting, forensics, and incident response. Robust data activity monitoring, policy-based cloud access control, least privilege, and identity federation capabilities provide a powerful solution that helps companies establish a secure cloud database while also improving visibility. More data than ever before is being put into cloud-based storage repositories. Leading cloud providers offer an array of storage options, yet databases remain the most common choice in today's enterprises. Because databases are updated so frequently, it's important to review their security controls regularly.



COMMON THREATS TO CLOUD DATABASES

Cloud databases are targets for attackers if they aren't properly secured. For example, in May 2021, security analytics software vendor Cognyte exposed 5 billion data records -- ironically, containing information on previous data breaches at other organizations -- due to a cloud database with weak authentication controls that a security researcher discovered. That same month, 150 million records from the Iranian messaging application Raychat were leaked on the internet following a database exposure in late 2020 and early 2021.

There are numerous threats to cloud databases, with the most common types including the following:

- **Data exposure.** If cloud databases are poorly secured, it's likely that the data in them could be exposed to the internet or other cloud resources. Attackers actively looking for exposed databases can take advantage of this and exfiltrate data for financial gain or other purposes.

- **Exposed APIs.** Many cloud databases offer a wide variety of APIs for administration, integration and synchronization with other data stores. If these APIs are exposed publicly, or poorly secured and left unmonitored, attackers may be able to access and manipulate database content and configurations.
- **Cloud workload hijacking.** Cloud database workloads may run in containers or virtual servers. As a result, databases that aren't properly secured could be exploited by attackers who then compromise the underlying container or OS runtime. This could lead to lateral movement by the attackers and other cloud services also being disrupted, exposed or compromised.
- **Application exploits.** Cloud databases are potentially susceptible to common attacks, such as SQL injection, which can lead to application compromises, escalation of access privileges for user and service accounts, exposure of database details and more. In that way, attackers may be able to expand compromises of cloud environments through traditional application-centric attacks.

CONTROLS AND POLICIES

In addition to implementing layered security controls across your entire network environment, database security requires you to establish the correct controls and policies for access to the database itself. These include:

- **Administrative Controls** to govern installation, change, and configuration management for the database.
- **Preventative Controls** to govern access, encryption, tokenization, and masking.
- **Detective Controls** to monitor database activity monitoring and data loss prevention tools. These solutions make it possible to identify and alert on anomalous or suspicious activities.

Database security policies should be integrated with and support your overall business goals, such as protection of critical intellectual property and your cyber security policies and cloud security policies. Ensure you have designated responsibility for maintaining and auditing security controls within your organization and that your policies complement those of your cloud provider in shared responsibility agreements. Security controls, security awareness training and education programs, and penetration testing and vulnerability assessment strategies should all be established in support of your formal security policies.

DATA PROTECTION TOOLS AND PLATFORMS

- **Data activity monitoring:** The solution should be able to monitor and audit all data activities across all databases, regardless of whether your deployment is on-premise, in the cloud, or in a container. It should alert you to suspicious activities in real-time so that you can respond to threats more quickly.
- **Encryption and tokenization capabilities:** In case of a breach, encryption offers a final line of defense against compromise. Any tool you choose should include flexible encryption capabilities that can safeguard data in on-premise, cloud, hybrid, or multi cloud environments.
- **Data security optimization and risk analysis:** A tool that can generate contextual insights by combining data security information with advanced analytics will enable you to accomplish optimization, risk analysis, and reporting with ease.

CONCLUSION

Cloud database and data security is a promising and emerging technology for the next generation of IT applications. The barrier and hurdles toward the rapid growth of cloud computing are data security and privacy issues. Reducing data storage and processing cost is a mandatory requirement of any organization, while analysis of data and information is always the most important tasks in all the organizations for decision making. So no organizations will transfer their data or information to the cloud until the trust is built between the cloud service providers and consumers.

REFERENCES

1. Kanimozhi, R. (2019). Adaptive and intelligent framework of data protection techniques for cloud storage. International Journal Of Cloud

Computing, 8(1), 50. doi: 10.1504/ijcc.2019.097906

2. Yaseen, Q., Althebyan, Q., Panda, B., & Jararweh, Y. (2016). Mitigating insider threat in cloud relational databases. Security And Communication Networks, 9(10), 1132-1145. doi: 10.1002/sec.1405
3. Connolly, T. M. (2021). Database Systems: A Practical Approach to Design, Implementation and Management 5th (fifth) edition. Addison Wesley.
4. Shcherbinina, Y., Martseniuk, B., & Filonenko, A. (2020). DATABASE SECURITY AND STUDY OF DATA ENCRYPTION METHODS IN CLOUD STORAGE. Системи Управління, Навігації Та Зв'Язку. Збірник Наукових Праць, 3(61), 104-106. doi: 10.26906/sunz.2020.3.104
5. Database Management Challenges in Cloud Environment. (2016). International Journal of Modern Trends in Engineering & Research, 3(9), 208-212. doi: 10.21884/ijmter.2016.3067.9neo0
6. A.A. Izang, A.O. Adebayo, O.J. Okoro and O.O. Taiwo, "SECURITY AND ETHICAL ISSUES TO CLOUD DATABASE," The Journal of Computer Science and its Applications, vol. 24, 2017.
7. Wang, L. (2014). Research on Security of Database in Cloud Computing Environment. Applied Mechanics and Materials, 644-650, 1694-1697. doi: 10.4028/www.scientific.net/amm.644-650.1694
8. A Survey: Cloud Computing Challenges & Security Issues. (2017). International Journal of Modern Trends in Engineering & Research, 4(3), 57-61. doi: 10.21884/ijmter.2017.4079.cfbgf
9. Chetan, & Singh, S. (2016). ENHANCEMENT OF CLOUD DATABASE SECURITY. Far East Journal of Electronics and Communications, 635-645. doi: 10.17654/ecsv3pii16635
10. K. Huang, and R. Tso, "A Commutative Encryption Scheme based on ElGamal Encryption," Database

ID: 25

LEVERAGING BLOCKCHAIN TECHNOLOGY TO OPTIMIZE SUPPLY CHAIN MANAGEMENT

Dr.MAHESH BHALAKRISHNAN
Head,
Department of Management,
Rathinam College of Arts and Science,
Coimbatore, Tamil Nadu, India.

Mr.A.ABDUL REHIMAN
Logistics Manager,
OLA Cell Technology Private Limited,
Bangalore, Karnataka, India.

ABSTRACT:
Supply Chain Management plays a pivotal role in modern business operations by ensuring the efficient flow of goods and services from suppliers to end consumers. However, this complex process often faces challenges such as lack of transparency, trust issues, data inconsistency, and inefficient documentation. Blockchain Technology, known for its decentralized, transparent, and immutable nature, has the potential to revolutionize Supply Chain Management. This paper explores the ways in which Blockchain Technology can be leveraged to optimize supply chain management processes, enhance transparency, improve traceability, and ultimately create a more efficient and secure global Supply Chain Network.

KEYWORDS: *Blockchain, Supply Chain, Smart Contracts, Security, Digital Signature.*

I. INTRODUCTION:

Supply Chain Management involves the coordination of various stages, from raw material extraction to product delivery. Traditional supply chain systems often struggle with issues like data silos, lack of transparency, and the need for intermediaries to ensure trust. Blockchain technology, originally introduced as the underlying infrastructure for cryptocurrencies like Bitcoin, has evolved beyond its initial use case and is now being explored for its potential to revolutionize supply chain management. By providing a decentralized and secure platform for recording and sharing information, blockchain holds the promise of optimizing supply chain operations.

II. BLOCKCHAIN TECHNOLOGY:

Blockchain is a distributed ledger technology that records transactions in a secure, transparent, and immutable manner. It consists of blocks containing transactions, linked in chronological order. Each block is cryptographically linked to the previous one, forming a chain. This decentralized structure eliminates the need for a central authority and ensures that data remains tamper-proof and trustworthy.

III. BENEFITS OF BLOCKCHAIN IN SUPPLY CHAIN MANAGEMENT:

- **Transparency and Traceability:** Blockchain provides end-to-end visibility into supply chain activities. Every transaction and movement of goods is recorded on the blockchain, enabling stakeholders to trace the journey of products from origin to destination.
- **Immutable Records:** Once data is recorded on the blockchain, it cannot be altered without consensus from network participants. This feature ensures data integrity and prevents fraudulent activities.
- **Smart Contracts:** Smart contracts are self-executing contracts with predefined rules. They can automate various supply chain processes, such as payments, verification, and compliance, reducing the need for intermediaries.
- **Reduced Fraud:** The decentralized nature of blockchain reduces the risk of fraud and unauthorized access to sensitive information. Each participant has a copy of the ledger,

making it challenging for malicious actors to manipulate data.

- **Efficient Documentation:** Traditional paper-based documentation can lead to delays and errors. Blockchain digitizes and streamlines documentation, leading to faster and more accurate processing.

IV. USE CASES:

- **Provenance Tracking:** Blockchain can track the origin and history of products, helping to ensure the authenticity and quality of goods. This is particularly useful for industries like luxury goods, pharmaceuticals, and food.
- **Supplier Verification:** Blockchain can securely store and verify supplier information, certifications, and compliance records, reducing the risk of working with unverified or unethical suppliers.
- **Inventory Management:** Blockchain's real-time visibility can improve inventory management by providing accurate and up-to-date information on stock levels, helping to prevent overstocking or stockouts.
- **Logistics and Shipping:** Blockchain can optimize the logistics process by enhancing transparency between different parties involved in transportation, reducing delays, and improving tracking accuracy.

V. CHALLENGES AND CONSIDERATIONS:

- **Scalability:** Blockchain networks must handle a large number of transactions efficiently to be effective in supply chain management.
- **Interoperability:** For widespread adoption, different blockchain platforms and systems need to seamlessly communicate with each other.
- **Data Privacy:** While blockchain ensures data integrity, sensitive information must still be protected, and mechanisms for secure data sharing need to be developed.

VI. CONCLUSION:

Blockchain technology has the potential to transform supply chain management by addressing long-standing issues related to transparency, traceability, and trust. While challenges remain, ongoing developments in blockchain scalability, interoperability, and privacy solutions are paving the way for its widespread adoption. As businesses recognize the benefits of leveraging blockchain for optimizing supply chain processes, collaborations between industry players and technology experts will play a pivotal role in shaping the future of supply chain management.

VII. Future Directions:

Future research can focus on addressing scalability concerns, refining interoperability standards, and developing advanced privacy-preserving techniques. Additionally, exploring the integration of emerging technologies such as the Internet of Things (IoT) and artificial intelligence (AI) with blockchain can further enhance the efficiency and effectiveness of supply chain management processes.

VIII. REFERENCES:

1. Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. Link
2. Buterin, V. (2013). Ethereum: A Next-Generation Smart Contract and Decentralized Application Platform. Link
3. Christidis, K., & Devetsikiotis, M. (2016). Blockchains and Smart Contracts for the Internet of Things. IEEE Access. Link
4. Zheng, Z., Xie, S., Dai, H., Chen, X., & Wang, H. (2017). An overview of blockchain technology: Architecture, consensus, and future trends. IEEE Access. Link
5. Kosba, A., Miller, A., Shi, E., Wen, Z., & Papamanthou, C. (2016). Hawk: The Blockchain Model of Cryptography and Privacy-Preserving Smart Contracts. In 2016 IEEE Symposium on Security and Privacy (SP). Link
6. Al Omar, A., Salah, K., & Pentland, A. (2019). Blockchain technology for supply chain integrity: A systematic literature review and future research agenda. Journal of Cleaner Production. Link
7. Ivanov, D., & Dolgui, A. (2019). A digital supply chain twin for managing the disruption risks. Transportation Research Part E: Logistics and Transportation Review. Link
8. Liao, Z., & Deschamps, F. (2019). Blockchain adoption in agri-food supply chains: The case of chia seeds in Australia. International Journal of Information Management. Link

ID: 26

DATA SECURITY

Mr.T.ELANGOVAN
Assistant Professor,
Department of Computer Technology,
Erode Arts and Science College,
Erode, Tamil Nadu, India.

Ms.G.RUBA
Assistant Professor,
Department of Computer Technology,
Erode Arts and Science College,
Erode, Tamil Nadu, India.

ABSTRACT

Data are today an asset more critical than ever for all organizations may think of. recent advances and trends, such as sensor systems, IoT, cloud computing, and data analytics, are making possible to pervasively, efficiently, and effectively collect data. However for data to be used to their full power, data security and privacy are critical. Data security and privacy have been widely investigated over the past thirty years, today we face new difficult data security and privacy challenges. Some of those challenges arise from increasing privacy concerns with respect to the use of data and from the need of reconciling privacy with the use of data for security in applications such as homeland protection, counterterrorism, and health, food and water security. Other challenges arise because the deployments of new data collection and processing devices, such as those used in IoT systems, increase the data attack surface. In this paper, we discuss relevant concepts and approaches for data security and privacy, and identify research challenges that must be addressed by comprehensive solutions to data security and privacy.

KEYWORDS: *IoT Data Analytics, Data Security and Privacy data masking and redaction of Sensitive Files Encryption.*

DATA SECURITY

Data security is the practice of protecting digital information from unauthorized access, corruption or theft throughout its entire lifecycle. It's a concept that encompasses every aspect of information security from the physical security of hardware and storage devices to administrative and access controls, as well as the logical security of software applications. It also includes organizational policies. When properly implemented, robust data security strategies will not only protect an organization's information assets against cybercriminal activities, but they'll also guard against insider threats and human error, which remain among the leading causes of data breaches today. Data security involves deploying tools and technologies that enhance the organization's visibility into where its critical data resides and how it is used. Ideally, these tools should be able to apply protections such as encryption, data masking and redaction of sensitive files, and should automate reporting to streamline audits and adhering to regulatory requirements.

Encryption

Using an algorithm to transform normal text characters into an unreadable format, encryption keys scramble data so that only authorized users can read it. File and database encryption solutions serve as a final line of defense for sensitive volumes by obscuring their contents through encryption or tokenization. Most solutions also include security key management capabilities.

Data Erasure: More secure than standard data wiping, data erasure uses software to completely overwrite data on any storage device. It verifies that the data is unrecoverable.

Data Masking: By masking data, organizations can allow teams to develop applications or train people using real data. It masks personally identifiable information (PII) where necessary so that development can occur in environments that are compliant.

Data Resiliency: Resiliency is determined by how well an organization endures or recovers from any type of failure—from hardware problems to power shortages and other events that affect data availability (PDF, 256 KB). Speed of recovery is critical to minimize impact.

Data security capabilities and solutions Data security tools and technologies should address the growing challenges inherent in securing today's complex, distributed, hybrid, and/or multicloud computing environments. These include understanding where data resides, keeping track of who has access to it, and blocking high-risk activities and potentially dangerous file movements. Comprehensive data protection solutions that enable enterprises to adopt a centralized approach to Data discovery and classification tools Sensitive information can reside in structured and unstructured data repositories including databases, data warehouses, big data platforms, and cloud environments. Data discovery and classification solutions automate the process of identifying sensitive information, as well as assessing and remediating vulnerabilities.

Data and file activity monitoring File tools analyze data usage patterns, enabling security teams to see who is accessing data, spot anomalies, and identify risks. Dynamic blocking and alerting can also be implemented for abnormal activity patterns.

Automated compliance reporting Comprehensive data protection solutions with automated reporting capabilities can provide a centralized repository for enterprise-wide compliance audit trails.

A comprehensive data security strategy incorporates people, processes, and technologies. Establishing appropriate controls and policies is as much a question of organizational culture as it is of deploying the right tool set. This means making information security a priority across all areas

Regardless of whether your data is stored on-premises, in a corporate data center, or in the public cloud, you need to ensure that facilities are secured against intruders and have adequate fire suppression measures and climate controls in place. A cloud provider will assume responsibility the principle of "least-privilege access" should be followed throughout your entire IT environment

All software should be updated to the latest version as soon as possible after patches or new **Backups** maintaining usable, thoroughly tested backup copies of all critical data is a core component of any robust data security strategy. In addition, all backups should be subject to the same physical and logical security controls that govern access to the primary databases and core systems.

Training employees in the importance of good security practices and password hygiene and teaching them to recognize social engineering attacks transforms them into a "human firewall" Implementing a comprehensive suite of threat management, detection, and response tools and platforms across your on-premises environment and cloud platforms can mitigate risks and reduce the probability of a breach.

AI

AI amplifies the ability of a data security system because it can process large amounts of data. Cognitive computing, a subset of AI, performs the same tasks as other AI systems but it does so by simulating human thought processes. In data security, this allows for rapid decision-making in times the definition of data security has expanded as cloud capabilities grow. Now organizations need more complex solutions as they seek protection for not only data, but applications

Quantum

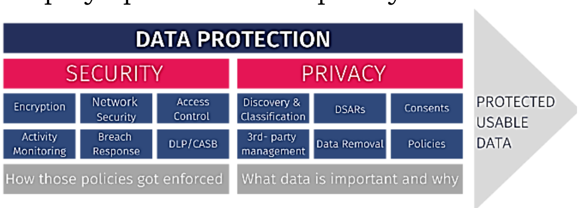
A revolutionary technology, quantum promises to upend many traditional technologies exponentially. Encryption algorithms will become much more faceted, increasingly complex and much more secure.

Data security and other security facets interact the key to applying an effective data security strategy is adopting a risk-based approach to protecting data across the entire enterprise. Early in the strategy development process, taking business goals and regulatory requirements into account, stakeholders should identify one or two data sources containing the most sensitive information, and begin there. After establishing clear and tight policies to protect these limited sources, they can then extend these best practices across the rest of the enterprise's digital assets in a prioritized fashion. Implemented automated data make securing cloud-based infrastructures requires a different approach than the traditional model of situating defenses at the network's perimeter. It demands comprehensive cloud data discovery and classification tools, plus ongoing activity monitoring and risk management. Cloud monitoring tools can sit between a cloud provider's database-as-a-service (DBaaS) solution and monitor data in transit or redirect traffic to your existing security platform. This allows for policies to be applied.

The use of personal computers, tablets, and mobile devices in enterprise computing environments is on the rise despite security leaders' well-founded concerns about the risks that this practice can pose. One way of improving bring your own device (BYOD) security is by requiring employees who use personal devices to install security software to access corporate networks, thus enhancing centralized control over and visibility into data access and movement. Another strategy is to build an enterprise-wide, security-first mindset, encouraging employees to utilize strong passwords, multi-factor authentication, regular software updates, and device backups, along with data encryption by teaching them the value of these actions.

Related solutionsData security solutions

Protect data across multiple environments, meet privacy regulations and simplify operational complexity



SECURITY

Security is protection from, or resilience against, potential harm caused by others, by restraining the freedom of others to act. Beneficiaries of security may be of persons and social groups, objects and institutions, ecosystems or any other entity or phenomenon vulnerable to unwanted change.

Refugees fleeing war and insecurity in Iraq and Syria arrive at Lesbos Island, supported by Spanish volunteers,

Security mostly refers to protection from hostile forces, but it has a wide range of other senses: for example, as the absence of harm as the presence of an essential good as resilience against potential damage or harm e.g. secure foundations; as secrecy as containment e.g. a secure room or cell and as a state of mind e.g. emotional security. The term is also used to refer to acts and systems whose purpose may be to provide security

DATA PRIVACY

With new data privacy laws popping up regularly to ensure websites and online businesses treat their users' data ethically, it can be hard to keep up with definitions, regulations, and legislation.

Data privacy refers to the proper use and processing of personal data by restoring control over their data to individuals. Simply put data privacy enables individuals to decide and limit access to the use and sharing of their personal data.Protecting personal information ensures that the data is kept secure. This concept is where data privacy transitions to data security and protection.

DATA PRIVACY LAWS

If you're a business owner with an online presence, you have probably heard about the numerous data privacy laws recently enacted worldwide.The following are examples of laws that aim to protect users' data privacy online:

DATA SECURITY AFFECTS DATA PRIVACY

Most online businesses and websites collect personal data, from email addresses to phone numbers, credit cards, and log-in details. Ideally, these entities shouldn't keep more information than is necessary, nor should they keep it longer than necessary.However, you cannot operationalize data privacy without ensuring the security of data.

For example, if you fail to protect people's credit card details against hackers and they get access to this data, they can sell it on the dark web. Therefore, data security is a prerequisite to data privacy.

DATA SECURITY

Like data privacy, the phrase "data security" is somewhat vague and not necessarily intuitive. This confusion is particularly true when comparing data privacy, data security, and data protection concepts.

Data security is the concept of protecting digital data from theft, corruption, or unauthorized access throughout its entire lifecycle Data security involves everything from the physical security of the storage devices and hardware to administrative access controls and the security of software applications.It also includes organizational policies and procedures. Correctly implementing data security can protect your data from cybercriminal activities, insider threats, and human error.

Various tools and technologies help protect your data, including

Redaction of sensitive files Data masking Encryption Automated reporting. These tools can help keep your data secure while supporting you in other areas, like streamlining your audits and complying with regulatory require

Data Protection

Once you have ensured appropriate data privacy and security, the next step is providing proper data protection. There are two definitions for "Data protection," narrow and broad

The narrow or more traditional definition of data protection: Maintaining data. Availability by way of backups so you can easily restore data the broader or more modern definition of data protection: It covers data availability immutability, preservation, deletion/destruction, and "data privacy" and "data security. The more data you collect and store, the more important it becomes to create backups for your critical data. For many companies, the timeliness of implementing a backup is also essential ideally, if you have lost critical data, you would want to replace it as soon as possible to avoid losing out on business during your downtime there are several ways to implement a data protection strategy, from using different storage devices to creating cloud backups and archiving.

CONCLUSION

Now that you have a solid understanding of the basic definitions of data privacy, data security, and data protection, let's look at how these three interlinked topics compare, how they are linked, and how they operate in tandem. Another relevant research area which has been the focus of intense research in the past ten years is the area of data security and privacy on the cloud. This area has seen significant research in different directions, such as for example approaches to support privacy-preserving fine-grained attribute-based access control on the cloud and provable possession of data on the cloud. Also the area of data privacy in social networks has received significant focus. One of the key issues emerging from such research is that in social networks collaborative approaches are needed for access control. The reason is that in social networks, a given piece of data, such as a picture, may refer to multiple social network users, and it is thus crucial that all such users be able to express their privacy preferences when sharing the piece of information. In addition, to the research directions mentioned so far in the paper, there are two other additional research directions that we would like to emphasize:
• Data protection from insider threat - protection against insider threat requires combining many different techniques, including context-based access control, anomaly detection in data access and use [54], and user behavior monitoring. User behavior monitoring however may entail privacy issues and therefore it requires a careful trade-off between security risks and individual privacy. • Privacy-aware software engineering – engineering software to provide strong privacy assurance requires, among other things, to identify the code portions that deal with sensitive data, the ability of applications to work on anonymized data and to deal with lack of permissions depending on specific spatial and temporal contexts; also as forensic tools are today able to recover memory contents after applications complete their execution, it is critical that applications scrub memory to permanently delete sensitive data. Finally tools are needed able to create profiles of expected usage of privacy-sensitive data by application programs and use these profiles at run-time to detect anomalies in the data use by the applications. As final remark we would like to mention

REFERENCE

1. Online Etymology Dictionary. "Origin and meaning of secure". www.etymonline.com. Retrieved 2017-12-17.
2. Jump up to:[a] [b] Barry Buzan, Ole Wæver, and Jaap de Wilde, Security: A New

Framework for Analysis (Boulder: Lynne Rienner Publishers, 1998), p. 32
3. House of Commons Defence Committee (2015). "Re-thinking defence to meet new threats". publications.parliament.uk. Retrieved 2017-12-17.
4. General Sir Nicholas Houghton (2015). "Building a British military fit for future challenges rather than past conflicts". www.gov.uk. Retrieved 2017-12-17.
5. FCNL (2015). "Peace Through Shared Security". Retrieved 2017-12-17.
6. Rogers, P (2010). Losing control: global security in the twenty-first century (3rd ed.). London: Pluto Press. ISBN 9780745329376. OCLC 65 8007519.
7. "Reliance spells end of road for ICT amateurs", May 07, 2013, The Australian
8. United Nations General Assembly (2010). "Resolution adopted by the General Assembly on 20 December 2010". www.un.org. Retrieved 2017-12-1
9. “Data, data everywhere”, The Economist, 25 February 2010, available at http://www.economist.com/node/15557443 (Downloaded on April 30, 2012).
10. E. Bertino, “Big Data – Opportunities and Challenges”, Panel Position Paper, Proceedings of the 37th Annual IEEE Computer Software and Applications Conference, COMPSAC 2013, Kyoto, Japan, July 22-26, 2013.
11. J. Manyika, M. Chui, J. Bughin, R. Dobbs, P. Bisson, and A. Marrs. Disruptive technologies: Advances that will transform life, business, and the global economy. http://www.mckinsey.com/insights/business technology/disruptive_technologies, May 2013.
12. E. Bertino, S. Nepal, R. Ranjan, “Building Sensor-Based Big Data Cyberinfrastructures”, IEEE Cloud Computing 2(5): 64-69 (2015).

ID: 27

REVIEW OF MACHINE LEARNING BASED UNDERWATER IMAGE ENHANCEMENT TECHNIQUES

Mr.S.KARTHIK
Assistant Professor,
Department of ECE,
Mahendra Institute of Technology,
Namakkal, Tamil Nadu, India.

Mr.D.P.SAI KRISHNA
Assistant Professor,
Department of ECE,
Mahendra Institute of Technology,
Namakkal, Tamil Nadu, India.

ABSTRACT:
The goal of image enhancement is to improve the quality of a source image so that it may be more easily understood by viewers. The purpose of the augmentation technique is to improve an image's informative value while also changing its aesthetic appeal to the viewer. By "image enhancement," we mean the methodical process of improving the perceivable qualities of images. The software enhances visual properties of Images like edges and contrast to produce a more arresting representation of photographs that is appropriate for close inspection and academic study. To increase the visual quality of photographs, many methods known together as "image enhancement" are employed. Some examples of these methods are increasing the contrast, decreasing the noise, coloring noise, and filtering noise. Image enhancement makes it easier to spot certain characteristics by expanding the range across which they may be perceived. Inadequate color contrast and obstructed eyesight are two of the most often encountered problems in underwater photography. Light scattering and refraction at the boundary between two media of different densities is to blame for the aforementioned difficulties. Light diffusion and the resulting attenuation of color distinction are two optical effects that may be traced back to scattering. It is possible to attribute the observed impacts on underwater imaging to both the water's inherent qualities and the presence of creatures and other chemicals found in the aquatic environment. Numerous approaches and methods have been developed by scientists to enhance the quality of underwater images. This research comprehensively evaluates and analyzes a number of techniques used to improve the standard of underwater photography. The major focus of this study is on analyzing the limitations of conventional methods.

KEYWORDS: *Machine Learning, Image Processing, Underwater Image Enhancement, Contrast Correction, Colour Correction.*

I. INTRODUCTION

Because of the numerous uses that may be achieved with it, underwater photography is becoming an increasingly important field. The applications cover a wide variety of endeavors, including the monitoring of underwater structures, the research of aquatic life forms such as fishes and coral reefs, the search for mineral and biological resources in water bodies, the conduct of aquatic robot inspections, and the participation in underwater archaeology, to name just a few of these activities and their respective focuses of study. However, the quality of the images taken when the subject was submerged in water is noticeably lower. The presence of particles that are suspended inside the water is thought to be the cause of the phenomenon known as light scattering. In addition to this, the frequency of the light being absorbed plays a role in the process. In addition, there are major challenges presented by Images captured underwater, such as color distortion, reduced contrast, and lower visibility. The refraction of light at varying wavelengths as it travels through water is the primary factor responsible for the appearance of color distortion.

This effect causes a change in the

coloring that is seen in Images taken underwater, resulting in a hue that is somewhere between bluish-green and greenish-blue. The random processes of light attenuation and scattering are responsible for the occurrence of the phenomena known as contrast degradation. It is possible to improve and optimize underwater Images in order to simplify the process of information extraction for the purpose of carrying out more research. It is possible to improve the images by making use of image restoration (Zhang et al., 2018) or image enhancement techniques, both of which have the potential to bring forth an improvement in the photographs' overall quality.

Image restoration processes typically involve two basic phases: the removal of noise and the use of an inverse degradation approach to enhance the visual integrity of underwater images. Both of these procedures are designed to improve the quality of the original image. The process of improving a photograph such that it seems to have a higher overall visual quality to human observers is referred to as "image enhancement." This article provides a comprehensive review of numerous techniques that have been used to enhance the quality of underwater photography. In order to improve the Image's quality as a whole, many different image-editing methods are used in order to get the desired effect. These strategies intend to improve the Image's visibility and clarity, making it more clearly observable to human viewers and also boosting its detectability when it is studied by automatic image analysis tools. The enhancement of visibility, the enhancement of Image color and contrast, the preservation of the naturalness of the image, and the enhancement of object prominence are the primary focus areas that demand consideration in image enhancement approaches for underwater photographs. There is a wide variety of software designed specifically for the purpose of improving photographs. These programs use a variety of techniques, such as white balance modification, color correction, histogram equalization, fusion-based approaches, and Retinex-based algorithms.

According to Zheng L et al. (2016), there are many different types of underwater images that are analyzed in the academic literature. These photographs exhibit characteristics such as blurriness, reduced visibility, uneven illumination, lower contrast, and the presence of light spots. In addition, it has been discovered that blue-green light has an effect on particular photographs (Yu, Haifeng, et al., 2020). In addition, there are Images that have varying degrees of brightness, consisting of both dark and light parts (Soni, Om Kumari, et al., 2020). In addition, there are Images that have haze in them (Ancuti, Codruta O., et al., 2017), and there are acoustic Images that have non-uniform lighting (Sharumathi, K., and R. Priyadharsini, 2016), amongst a variety of other instances.

The vast majority of academic study devoted to the improvement of underwater Images makes use of a broad variety of images, but the majority of the time they receive them from internet sources. These studies make use of these Images in order to demonstrate the outcomes of the strategies that they suggest. A limited number of datasets that are open to the public are currently available. According to the documentation found in their scholarly work, Duarte et al. (2016) were able to effectively compile a sizable body of visual data that they collectively refer to as TURBID. The quality of research carried out in the field of underwater photography is one of the principal goals that will be accomplished as a direct result of the creation of this dataset.

The Images in this collection are separated into three distinct categories: milk, deep blue, and chlorophyll. Each group is made up of a different number of photographs; for example, the Milk and DeepBlue groups each include 20 photographs, whereas the Chlorophyll group has 42 photographs. According to the research that was carried out by Li et al. (2019), the UIEB dataset has a collection of 890 photographs that were shot when the subjects were submerged in water. The raw images and their high-resolution equivalents are included in the dataset together with the processed photographs. In addition, the UIEB dataset has its own distinct collection of sixty underwater Images, each of which presents a unique set of difficulties and problems.

II. RELATED WORKS

The academic literature explores the topic of enhancing underwater photography by the utilization of either several images (Narasimhan SG, Nayar SK, 2003) or a single image. The application of several images may be utilized in many manners to exploit their presence and achieve image enhancement through the process of image fusion. Tian et al. (2018) conducted a study that presents a novel methodology for improving target performance. The research conducted by Narasimhan and Nayar (2003) introduces an innovative methodology with the objective of improving the contrast in images that have been degraded due to weather conditions. This approach employs the use of two images in its implementation.

While the use of a dual-image methodology has demonstrated efficacy, it is crucial to acknowledge that this strategy necessitates the acquisition of several Images taken from a same geographic location. Furthermore, the necessity to combine separate Images in a pixel-wise manner imposes limitations on the practicality of these approaches in several real-world scenarios. Image enhancement methods that exclusively utilize a single image for further processing are often favored for their ability to function alone without requiring supplemental images. Consequently, our main focus is in the examination of particular strategies employed to improve visual representations. The table provided offers a succinct summary of the most recent and noteworthy techniques.

Table 1. Summary of Underwater image enhancement techniques

Technique	Method	Pros	Cons
CycleGAN	Generative adversarial network for image translation.	Learns mapping without paired data.	May generate unrealistic artifacts.
U-Net	Convolutional neural network for image-to-image translation.	Captures local and global features.	Limited context for large scenes.
DeepUWEnhance	Multi-scale CNN for enhancing underwater images.	Handles various underwater conditions.	Sensitive to noise and artifacts.
SSDEN	Stacked sparse denoising autoencoders for enhancement.	Removes noise while preserving details.	May over-smooth certain textures.
D3-Net	Depth-guided domain adaptation for image enhancement.	Utilizes depth information for guidance.	Requires depth maps, not always available.
AE-Enhance	Autoencoder-based approach for enhancing low-quality images.	Adaptive enhancement for varying scenes.	May struggle with extreme conditions.
WGAN-GP	Wasserstein GAN with gradient penalty for image enhancement.	Stable training and better convergence.	Computational complexity may be high.

III. HISTOGRAM EQUALIZATION

Histogram Equalization (HE) is a common underwater picture improvement approach. This method yields a gray gradient picture with uniform probability distribution. HDR eliminates image deterioration and component explanations, improving visual representation. Pixel intensities should be uniform for this strategy to operate. It cannot localize unique characteristics. Ketcham writes. AHE is a form of HIM. AHE, introduced in 1976, maps picture blocks. It computes this mapping. The approach works effectively with limited intensity ranges. It improves findings by adding noise. Use modified RGB and HSV channels.

Ghani et al. (2015) used a histogram. Rayleigh Distribution limitations RGB contrast stretching. The authors demonstrate that noise reduction increases data contrast. Ghani et al. (2015) obtained an average mean square error (MSE) of 76.76 and a peak signal-to-noise ratio (PSNR) of 31.13 using this technique. Reza Contrast-Limited Adaptive Histogram Equalization (CLAHE) improves image contrast without artifacts. Noise reduction. Like Histogram Equalization (HE), it suppresses grey images. Noise-reduction and contrast-limiting are applied. Image histogram thresholds control noise signal amplification.

Zheng et al. (2016) enhanced one underwater picture. CLAHE and USM adjustments helped the researchers reach their target. Weighted blending mixes photos from these ways. Histogram equalization and homomorphic filters generate inferior pictures. Garg et al. (2018) improved picture quality with CLAHE and Percentile.

IV. CONTRAST LIMITED ADAPTIVE HISTOGRAM EQUALIZATION

The current investigation employs the

Contrast Limited Adaptive Histogram Equalization (CLAHE) technique, especially utilizing both the Red-Green-Blue (RGB) and Hue-Saturation-Value (HSV) channels. The percentile methodology is employed in conjunction with the Contrast Limited Adaptive Histogram Equalization (CLAHE) technique. The application of color and contrast restoration techniques significantly improved the visual quality of the images, which initially had diminished contrast and shading. Furthermore, the outcomes acquired surpass those achieved by the utilization of Histogram Equalization, Dehazing, Percentile, and CLAHE techniques. According to a study conducted by Soni, Om Kumari, et al. (2020), the findings indicated that the Contrast Limited Adaptive Histogram Equalization (CLAHE) technique demonstrated greater effectiveness in mitigating excessive noise amplification when compared to the exclusive use of the Adaptive Histogram Equalization (AHE) method.

BBHE also retains mean brightness and boosts contrast, according to the study. Zhang et al. (2017) enhanced underwater images with multi-scale fusion. Image enhancement and correction are methodical. White balancing improves underwater picture model recovery quality. Multi-scale fusion combines these two elements. Fusion weights inputs by saturation and contrast. This speeds implementation. Ancuti et al. input underwater noise-affected photographs. Color compression and white balance adjustments make two images from one. Integrating pictures adds image weight maps. Weight maps in composite images capture edge features and contrast from input photographs (Sharumathi, K., and R. Priyadharsini, 2016). Weight map adjustment targets low-frequency components to sharpen the picture. Multi-scale fusion integrates altered images. The underwater camera system doesn't affect this study's method. It sharpens vision.

V. DARK CHANNEL PRIORITY

The Dark Channel Priority (DCP) algorithm has been extensively employed as a significant technique for Image dehazing (He, K., et al., 2011). Nevertheless, it is crucial to recognize that the DCP approach, in isolation, does not possess the capability to directly ascertain the depth of a Image. Therefore, the researchers in the study put forth techniques with the objective of enhancing the quality of underwater images through the capture of transmittance (Peng Y-T, et al., 2017). The aforementioned methodologies encompass the process of capturing the three-dimensional aspects of a Image and employing a singular transmission map to extract the relevant depth data. However, the practical utilization of these devices is confined to particular underwater environments. Soni, Om Kumari, and their colleagues (2020) present an alternative approach in their latest research to address the challenges posed by haze in underwater photographs. This approach entails the employment of a dual transmission map.

VI. ADAPTIVELY CLIPPED CONTRAST LIMITED HISTOGRAM EQUALIZATION

Dixit et al. (2016) suggest using DCP, ACCLAHE, and HF to increase image quality. Improve image quality by reducing noise, uneven lighting, and contrast. It processes atmospheric fog RGB Images. DCP estimates and corrects visual blurriness. Maximum filters define transmission map and dominant ambient light. Increased transmission map smooths images. ACCLAHE improves photographs. The sub-image histogram shows the magnitude of the uppermost bin and the uniform distribution of pixels above it across gray levels. HF benefits more. Lu et al. (2015) studied submerged Images from murky aquatic settings.

VII. RETINEX TEORY

Retinex-based algorithms are a class of computational methodologies utilized in the domain of image enhancement and analysis. The primary goal of this study is to develop a variational framework based on Retinex theory in order to enhance the visual quality of individual underwater images. This work introduces a novel statistical methodology for addressing the issue of color distortion through the utilization of a color correction technique. Subsequently, the color-corrected Image that has been generated is then subjected to the implementation of a variational Retinex model. The process involves the deconstruction of luminosity into its fundamental elements, namely reflectance and illumination. The utilization of the histogram approach serves to improve the quality of reflectance and illumination, so effectively mitigating the problems associated with fuzziness and underexposure. In their study, Fu et al. (2017) propose a two-step methodology to effectively mitigate the difficulties arising from hue shift and contrast deterioration. The current investigation presents a novel method for color correction, which involves the utilization of a piecewise linear transformation as a strategy to achieve color correction. Subsequently, a methodology is implemented to amplify the overall contrast, hence improving the contrast.

VIII. CONCLUSION

Based on the analysis of the underwater situations discussed in this research, it can be inferred that there is a need to improve the quality of underwater imaging. The observed phenomena may be ascribed to a variety of factors, encompassing inequalities in contrast, the absorption and attenuation of light, the purity or turbidity of water, the existence of noise in captured images, and the distance separating the camera from the subject. Additional improvements may be necessary in order to precisely determine and locate probable submerged entities. The findings suggest that the application of wavelet and curvelet modification techniques exhibits improved effectiveness in noise reduction for auditory Images. The utilization of fusion methodologies has exhibited positive outcomes in the field of challenging underwater computer vision assignments, specifically with regards to the improvement of images. The optimization of underwater dehazing and contrast enhancement might possibly be achieved by the utilization of high-frequency (HF) filtering, contrast-limited adaptive histogram equalization (CLAHE), in conjunction with dark channel prior (DCP) and transmission maps. Significant advancements have been achieved in the domain of subaquatic study and inquiry. However, there remains a substantial potential for enhancing image and video processing methodologies to effectively aid and augment underwater research endeavors. Furthermore, it is crucial to gather datasets comprising underwater photographs in order to enhance the efficiency of the evaluation process by employing a standardized collection of Images.

IX. REFERENCES

1. Ancuti, C. O., Ancuti, C., Codruta, O., & De, N. (2017). Colour balance and fusion for underwater image enhancement. IEEE Transactions on Image Processing, 27(1), 379-393.
2. Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., & Tao, D. (2019). An Underwater Image Enhancement Benchmark Dataset and Beyond. IEEE Transactions on Image Processing, 29, 4376-4389.
3. Chiang, J. Y., & Chen, Y. C. (2011). Underwater image enhancement by wavelength compensation and dehazing. IEEE Transactions on Image Processing, 21(4), 1756-1769.
4. Dixit, S., Tiwari, S. K., & Sharma, P. (2016). Underwater image enhancement using DCP with ACCLAHE and homomorphism filtering. In 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES) (pp. 1-6). IEEE.
5. Duarte, A., Codevilla, F., Gaya, J. D. O., & Botelho, S. S. (2016). A dataset to evaluate underwater image restoration methods. In OCEANS 2016-Shanghai (pp. 1-6). IEEE.
6. Garg, D., Garg, N. K., & Kumar, M. (2018). Underwater image enhancement using blending of CLAHE and percentile methodologies. Multimedia Tools and Applications, 77(20), 26545-26561.
7. Fu, X., Fan, Z., Ling, M., Huang, Y., & Ding, X. (2017). Two-step approach for single underwater image enhancement. In Proc. International Symposium on Intelligent Signal Processing and Communication Systems (pp. 789-794).
8. Ghani, A. S. A. A., & Isa, N. A. M. (2015). Underwater image quality enhancement through integrated colour model with Rayleigh distribution. Applied Soft Computing, 27, 219-230.
9. He, K., Sun, J., & Tang, X. (2011). Single image haze removal using dark channel prior. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(12), 2341-2353.
10. Hung-Yu, P. Y. C., Chien-Chuan, H., & Zhuang, Y. Z. (2005). Second International Conference on Innovations in Bio computing and Applications.
11. Iqbal, K., Odetayo, M., James, A., Salam, R. A., & Talib, A. (2010). Enhancing the low quality images using unsupervised colour correction method. In IEEE Int. Conf. Syst. Man Cybern. (SMC) (pp. 1703-1709).
12. Ketcham, D. J. (1976). Real-time image enhancement techniques. In Image processing. International Society for Optics and Photonics.

13. Lu, H., Jiang, J., Ma, K. K., & Yao, H. (2015). Contrast enhancement for images in turbid water. Journal of the Optical Society of America A, 32(5), 886-893.
14. Narasimhan, S. G., & Nayar, S. K. (2003). Contrast restoration of weather degraded images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(6), 713-724.
15. Peng, Y. T., & Cosman, P. C. (2017). Underwater image restoration based on image blurriness and light absorption. IEEE Transactions on Image Processing, 26(4), 1579-1594.
16. Reza, A. M. (2004). Realization of the contrast limited adaptive histogram equalization (CLAHE) for realtime image enhancement. Journal of VLSI signal processing systems for signal, image and video technology, 38(1), 35-44.
17. Sharumathi, K., & Priyadharsini, R. (2016). A survey on various image enhancement techniques for underwater acoustic images. In 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) (pp. 1-6). IEEE.
18. Soni, O. K., & Kumare, J. S. (2020). A Survey on Underwater Images Enhancement Techniques. In IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT) (pp. 1-6). IEEE.
19. Tian, H., Zhu, J., Tan, S., Zhang, Y., Li, Y., & Hou, X. (2018). Rapid underwater target enhancement method based on polarimetric imaging. Optics and Laser Technology, 108, 515-520.
20. Yu, H., Wang, Y., Tao, D., & Wen, Y. (2020). Underwater image enhancement based on DCP and depth transmission map. Multimedia Tools and Applications, 1-18.
21. Zhang, M., & Peng, J. (2018). Underwater image restoration based on a new underwater image formation model. IEEE Access, 6, 58634-58644.
22. Zhang, C., Zhang, X., & Tu, D. (2017). Underwater image enhancement by fusion. In International Workshop of Advanced Manufacturing and Automation (pp. 1-6). Springer, Singapore.
23. Zheng, L., Shi, H., & Sun, S. (2016). Underwater image enhancement algorithm based on CLAHE and USM. In Proc IEEE Int Conf Inform Automation (ICIA) (pp. 585-590).

ID: 28

EMERGING TRENDS IN MACHINE LEARNING-ASSISTED VLSI TESTING AND VERIFICATION

Mr.M.SATHISHKUMAR
Assistant Professor,
Department of ECE,
Mahendra Institute of Technology,
Namakkal, Tamil Nadu, India.

Mr.A.KARTHIK
Assistant Professor,
Department of ECE,
Mahendra Institute of Technology,
Namakkal, Tamil Nadu, India.

ABSTRACT:
Due to the escalating levels of process variability, the integrated circuit (IC) industry is compelled to devise strategies aimed at reducing both the manufacturing time of chips and the intricacy of their designs. The completion of these tasks necessitates a substantial allocation of both temporal and financial resources. The utilization of artificial intelligence (AI) can facilitate the automation of challenging and data-intensive processes in VLSI design and testing. This is achieved via the implementation of diverse learning methodologies. In the realm of designing and manufacturing very large-scale integration (VLSI) systems, artificial intelligence (AI) and machine learning (ML) algorithms play a crucial role in reducing the time and effort involved in processing and managing data across several levels of abstraction. The enhancement of both integrated circuit yield and production time is observed as an outcome. This study undertakes a thorough examination of the methodologies employed in the automated design and manufacture of Very Large Scale Integration (VLSI) circuits using Artificial Intelligence (AI) and Machine Learning (ML) techniques. This study further explores the possible uses of artificial intelligence and machine learning within the domain of extremely large scale integrated circuit design. Implementations should strive to possess both cleverness and speed, while also maximizing efficiency.

KEYWORDS: *Artificial Intelligence (AI) Machine learning (ML) Manufacturing CMOS VLSI Design*

I. INTRODUCTION

Since the 1960s, there has been an exponential growth in the number of transistors that can be manufactured on a single device [1, 2]. The steady decrease in size of transistors over the course of a number of technical iterations has resulted in a significantly enhanced density and usefulness of these components [3]. This has led to a significant development of the microelectronics sector. Technology known as very-large-scale integration (VLSI) has advanced to the point where it is now possible to integrate complicated digital systems on a single processor. Due to the proliferation of portable technology in recent years, there has been a discernible rise in the level of activity surrounding the ever-increasing need for designs that are both power-efficient and feature-rich. As a result of the growing demand in the electronics industry, highly advanced and scalable VLSI circuits are now being developed. Because of the ongoing reduction in device dimensions, integrated circuit (IC) technology has made considerable strides toward further advancement in recent years, which has led to increased device performance. At the moment, there is a movement toward the trend of lowering device dimensions to a gate regime of sub-3 nanometers or even smaller sizes. Device developers have been faced with a variety of hurdles as a result of the dramatic size reduction of CMOS technology, but they have also been given new opportunities for progress. The manufacturing process of semiconductors is becoming increasingly complicated due to the ever-decreasing size of transistors. When we get close enough to the size of atoms, a phenomena called simple scaling stops happening. Despite the fact that the devices are rather little, there are some characteristics of their effectiveness that diminish. During the manufacturing process, there is a rise in permeability [4, 5, 6], a loss in gain [7], and an increased sensitivity to process fluctuations [7]. In addition, there is an increased susceptibility to process fluctuations [7].

There are variances in performance among transistors of the same size as a result of the large rise in the number of process changes, which has a significant influence on the functioning of electronic circuits. Due to the fact that it possesses the properties of a stochastic random variable, the latency of the circuit's propagation is also altered. This adds complexity to the procedures taken for timing closure and has a significant impact on the yield of semiconductors [8]. The increase in process variability at the nanoscale level is a significant factor that contributes to the downward trend in parametric yield. It has been shown that complementary metal-oxide-semiconductor (CMOS) transistors are less tolerant of changes in the manufacturing process than multi-gate field-effect transistors (FETs) [9]. Nevertheless, aggressive scaling has the potential to influence the performance characteristics of the system (10, 11).

It is vital to develop sophisticated and cost-effective design procedures that prioritize exact optimization within the VLSI design cycle in order to sustain future performance trends in circuits and systems. These techniques must be implemented in order to meet the requirements of the VLSI design cycle. The amount of time necessary to finish a semiconductor is directly proportional to the efficiency with which electronic design automation (EDA) technologies are able to circumvent design limitations. Traditional rule-based methods, such those employed in Electronic Design Automation (EDA), take significantly longer time to locate the best possible solution for a specified group of design constraints. In addition to this, it is essential to keep in mind that the methods that are often utilized for completing these jobs rely primarily on manual processes.

As a consequence of this, they are susceptible to time limits and need significant resources, both of which cause delays in the time it takes to bring the product to market. In addition, after collecting the data, designers have a difficult time understanding the underlying functioning, also known as the fundamental causes of problems, and putting the essential remedies into action. Alterations to both the gameplay and the surrounding environment contribute to the difficulty spike [12, 7].

The application of artificial intelligence (AI) has become an important aspect in the resolution of a broad variety of issues across a variety of industries. Emulating human intellect is the foundation of artificial intelligence (AI), which refers to the process of developing robots that are able to reproduce and carry out a variety of activities with varied degrees of difficulty. The topic of artificial intelligence (AI) includes a subset known as machine learning (ML). Acquiring new knowledge, exercising logical thinking, making accurate predictions, and gathering information are some of the goals of artificial intelligence and machine learning. Artificial intelligence and machine learning have the capability of finding and analyzing patterns and trends within enormous datasets, which enables users to make decisions that are both educated and pertinent to their circumstances.

The processing of multidimensional and multivariate data is made possible via the use of machine learning and artificial intelligence algorithms. These techniques provide extraordinary computer efficiency. The range of potential applications for artificial intelligence and algorithms for machine learning is practically infinite, which is not surprising considering the numerous advantages connected with these technologies.

II. EXISTING REVIEWS

AI influenced VLSI design in 1985 [15]. The user briefly reviewed AI approaches in CAD tools at different levels of very large-scale integration (VLSI) design. These researchers pioneered VLSI design and production learning. Knowledge-based systems employ NNs [18]. Data analytics and machine learning optimize physical design [19]. Machine learning's physical design uses and potential are examined [20]. Beerel et al. (21) explored ML algorithm challenges and opportunities in asynchronous CAD and VLSI. The ML-based design advisor would watch and document designers' usage of standard register-transfer level (RTL), logic synthesis, and place route tools. Design advisors identify the optimum solution using strong training engines. The design advisor then advises circuit designers. Design consultants prefer asynchronous CAD and machine learning. Stratigopoulos et al. (22) explored IC testing with machine learning (ML). The researchers illustrated many ML approaches used in IC testing and gave future practitioners important guidance.

III. BRIEF ON VLSI DESIGN FLOW

Fig. 1 depicts a hierarchical digital IC design pipeline for full-custom/semi-custom IC designs. Digital circuit design guidelines abstractly outline functionality, interface, and architecture. Block diagrams demonstrate functionality, time, propagation delays, package type, and design constraints. They sign a vendor-design engineer agreement. Architects design system architecture. RISC/CISC and ALU/FPU counts are covered. This level defines micro-architectural subsystem unit functions. Such descriptions assist architects estimate design performance and power. Verilog HDL/VHDL is used for behavioral design. Behavior separates unctionality from implementation. RTL descriptions verify time.

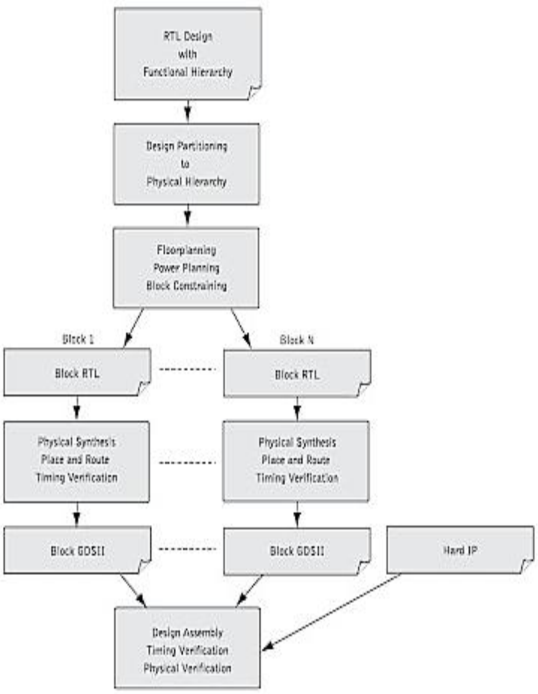


Fig 1. Hierarchical digital IC Design

4. BRIEF ON AI/ML ALGORITHMS

Statistical learning underpins most growing research and technology sectors. Data from each area may be mined for learning patterns and parameter relationships for future analysis and predictions. Statistics answers many real-world issues. AI mimics human behavior. ML self-corrects with new data. ML processes structured and semi-structured data, whereas AI manages all three. ML includes supervised, unsupervised, and reinforcement learning. Every data element with an output label is supervised. Unsupervised learning uses just input variables. Semi-supervised learning employs labeled and unlabeled data [28].

4.1. Supervised Learning

Classification and regression require supervision. Classification generates data class models. Classifiers predict discrete categories [30]. Regression predicts missing numbers, not class labels. Numerical regression technique forecasts continuous-valued functions [29]. Numeric class-label predictions. Classification/regression predicts Y = f (X) for X input variables and Y output variables.

4.2. Unsupervised Learning

Unsupervised learning requires no labels for training tuples. Unsupervised learning generates less data. Unsupervised learning makes point estimates/desired output for essential input vectors harder. Discovers data patterns. Unsupervised learning helps grouping and dimensionality reduction. Clustering divides items into "clusters" that are more similar. [23] explains. KNN, K-means, hierarchical, and agglomerative clustering are frequent [24].

4.3. Semi-supervised Learning

Semi-supervised learning is supervised and unsupervised. It helps when training data has few labeled and many unlabeled samples. It labels data automatically. It often outperforms supervised/unsupervised learning. Algorithms model unlabeled data with pseudo labels after training with minimal labeled data. Labeled data with pseudo labels and unlabeled data increases accuracy [25, 26]. Combining both semi-supervised techniques components takes more work in sophisticated situations.

4.4. Deep Learning

Deep learning is a subfield of machine learning that lends itself particularly well to the analysis of large amounts of data. Through the use of deep learning, a computer is able to construct complex notions from simple ones [28]. A mathematical function that maps a set of

ISBN: 978-81-965283-7-9

input values to a set of output values is called an MLP. Composing a large number of smaller functions together results in the formation of the function. A neural network (NN) with only one or two hidden layers is referred to as a shallow neural network (SNN). Deep neural networks, often known as DNNs, are networks that have tens to hundreds of these types of layers. DNNs are able to identify distributed expressions of data because they are able to extract features layer by layer and integrate low-level features to produce high-level features [29].

5. AI AT THE CIRCUIT SIMULATION

Early modeling of functional and electrical performance differences enhances IC yield. AI/ML algorithms in E-CAD tools improve chip turnaround and performance with less design work. Researchers estimated leakage power, total power, dynamic power, propagation latency, and IRdrop using stack-level transistor and subsystem models [18]. Circuit modeling has used LR, PR, RSM, SVM, ensembled approaches, Bayes theorem, ANNs, and pattern recognition models at various abstraction levels [19]. Subsections describe VLSI device/circuit characterization learning approaches.

6. AI IN ARCHITECTURES

AI/ML redesigned VLSI [13]. High-bandwidth, high-performance semiconductors and NN algorithms simplify real-time application hardware implementations. Architectures influenced VLSI for decades. Edge applications with fast processing, dependability, low implementation cost, and short time-to-market windows improve design most. Images, voice, IoT, and cars dominate architectural design literature.

7. AI IN PHYSICAL DESIGN

VLSI Combinatorial physical design issues need several iterations. Advanced design rules and DFM limitations have made optimum solutions difficult due to semiconductor technology improvement [18]. These violations are fixed. Advanced node manual design closure misses market windows. Early alterations harm design quality and manufacturing, delaying design closing. Current technology requires early design foresight. Machine learning and pattern-matching enhance abstraction and physical design processes. They link stages and accelerate design closing.

8. AI AT TESTING

VLSI testing finds IC problems after chip manufacturing. VLSI design's most crucial phase. Early defect detection saves money. Rule of 10 [23] states that IC design cycle fault detection expenses increase by a factor of 10. Yield must be increased to avoid tarnishment [28]. VLSI testing consumes 70% of design development time and resources. Testing methods vary per design phase. Functional verification, acceptance, manufacturing, wafer-level, packaging-level, etc. AI/ML's testing contributions are emphasized.

9. CHALLENGES AND OPPORTUNITIES FOR AI/ML IN VLSI

Leakage rises, gain decreases, and manufacturing process variances grow when devices reach atomic dimensions [7]. Yield drops. Nanometer technology's rising process variability and environmental variety are decreasing circuit performance. Most parameters fluctuate statistically, invalidating worst-case process corner modeling. Many characteristics are interconnected and vary greatly.

10. CONCLUSION

Algorithms and architectures must balance training data with model limitations. Another major challenge is integrating these innovative methods into the VLSI pipeline. Another issue is standardized, licensed, debuggable ML algorithms. Critically channeling ML designers' subject expertise with CAD designers delivers high-yield implementations. Capturing and exploring data flow throughout the design cycle maximizes training data availability. Systematic creation, data capture, distributed bid data systems for semiconductor operations, and data-driven optimizations can increase quality, cost, and time for chip designers. Billion-device chips create huge data . AI/ML algorithms can use simulation/analysis data to study complex I/O interactions between components, processes, and abstraction layers at each abstraction level. AI/ML can optimize VLSI–CAD design-flow optimization and file operation data stream clustering. Differential programming and quantum ML may change EDA.

REFERENCES

1. A. B. Kahng, Machine learning applications in physical design: Recent results and directions, in: Proceedings of the 2018 International Symposium on Physical Design, 2018, pp. 68–73.
2. A. R. Brown, N. Daval, K. K. Bourdelle, B. Nguyen, A. Asenov, Comparative simulation analysis of process-induced variability in nanoscale soi and bulk trigate finfets, IEEE Transactions on Electron Devices 60 (2013) 3611–3617.
3. B. H. Calhoun, Y. Cao, X. Li, K. Mai, L. T. Pileggi, R. A. Rutenbar, K. L. Shepard, Digital circuit design challenges and opportunities in the era of nanoscale cmos, Proceedings of the IEEE 96 (2008) 343–365.
4. B. Khailany, H. Ren, S. Dai, S. Godil, B. Keller, R. Kirby, A. Klinefelter, R. Venkatesan, Y. Zhang, B. Catanzaro, et al., Accelerating chip design with machine learning, IEEE Micro 40 (2020) 23–32.
5. C. K. C. Lee, Deep learning creativity in eda, in: 2020 International Symposium on VLSI Design, Automation and Test (VLSI-DAT), 2020, pp. 1–1.
6. C. Schuermyer, Deploying new nodes faster with machine learning for ic design and manufacturing, in: 2019 International Symposium on VLSI Technology, Systems and Application (VLSI-TSA), 2019, pp. 1–3. doi:10.1109/VLSI-TSA.2019.8804650.
7. D. Sylvester, H. Kaul, Power-driven challenges in nanometer design, IEEE Des. Test 18 (2001) 12–22.
8. D. S. Lopera, L. Servadei, G. N. Kiprit, S. Hazra, R. Wille, W. Ecker, A survey of graph neural networks for electronic design automation, in: 2021 ACM/IEEE 3rd Workshop on Machine Learning for CAD (MLCAD), 2021, pp. 1–6. doi:10.1109/MLCAD52597.2021.9531070.
9. G. E. Moore, Cramming more components onto integrated circuits, reprinted from electronics, volume 38, number 8, april 19, 1965, pp.114 ff., IEEE Solid-State Circuits Society Newsletter 11 (2006) 33–35.
10. G. Rabbat, Vlsi and ai are getting closer, IEEE Circuits and Devices Magazine 4 (1988) 15–18.
11. H.-G. Stratigopoulos, Machine learning applications in ic testing, in: 2018 IEEE 23rd European Test Symposium (ETS), IEEE, 2018, pp. 1–10.
12. H. Iwai, Logic lsi technology roadmap for 22 nm and beyond, in: 2009 16th IEEE International Symposium on the Physical and Failure Analysis of Integrated Circuits, 2009, pp. 7–10. doi:10.1109/ IPFA.2009.5232710.
13. H. . P. Wong, D. J. Frank, P. M. Solomon, C. H. J. Wann, J. J. Welser, Nanoscale cmos, Proceedings of the IEEE 87 (1999) 537–570.
14. I. A. Elfadel, D. Boning, X. Li, Machine Learning in VLSI Computer-Aided Design, 2019. doi:10.1007/978-3-030-04666-8.
15. J. Carballo, W. J. Chan, P. A. Gargini, A. B. Kahng, S. Nath, Itrs 2.0: Toward a re-framing of the semiconductor technology roadmap, in: 2014 IEEE 32nd International Conference on Computer Design (ICCD), 2014, pp. 139–146. doi:10.1109/ICCD.2014.6974673.
16. J. Delgado-Frias, W. Moore, VLSI for Artificial Intelligence, volume 68, 1989. doi:10.1007/978-1-4613-1619-0.
17. L. Capodieci, Data analytics and machine learning for designprocess-yield optimization in electronic design automation and ic semiconductor manufacturing, in: 2017 China Semiconductor Technology International Conference (CSTIC), 2017, pp. 1–3.
18. L. Wang, M. Luo, Machine learning applications and opportunities in ic design flow, in: 2019 International Symposium on VLSI Design, Automation and Test (VLSI-DAT), 2019, pp. 1–3.
19. M. Belleville, O. Thomas, A. Valentian, F. Clermidy, Designing digital circuits with nano-scale devices: Challenges and opportunities, Solid-State Electronics 84 (2013) 38–45. Selected Papers from the ESSDERC 2012 Conference.
20. M. H. Abu-Rahma, M. Anis, Variability in vlsi circuits: Sources and design considerations, in: 2007 IEEE International Symposium on Circuits and Systems, 2007, pp. 3215–3218. doi:10.1109/ISCAS. 2007.378156.
21. M. Rapp, H. Amrouch, Y. Lin, B. Yu, D. Z. Pan, M. Wolf, J. Henkel, Mlcad: A survey of research in machine learning for cad keynote paper, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (2021) 1–1.
22. M. Z. A. Khan, H. Saleem, S. Afzal, Application of vlsi in artificial intelligence (2012).
23. P. A. Beerel, M. Pedram, Opportunities

for machine learning in electronic design automation, in: 2018 IEEE International Symposium on Circuits and Systems (ISCAS), 2018, pp. 1–5.

24. R. S. Kirk, The impact of ai technology on vlsi design, in: Managing Requirements Knowledge, International Workshop on, IEEE Computer Society, 1985, pp. 125–125.
25. R. S. Rathore, A. K. Rana, R. Sharma, Threshold voltage variability induced by statistical parameters fluctuations in nanoscale bulk and soi finfets, in: 2017 4th International Conference on Signal Processing, Computing and Control (ISPCC), 2017, pp. 377–380. doi:10.1109/ISPCC.2017.8269707.
26. R. Vaddi, S. Dasgupta, R. P. Agarwal, Device and circuit design challenges in the digital subthreshold region for ultralow-power applications, VLSI Des. 2009 (2009).
27. S. Chaudhuri, N. K. Jha, Finfet logic circuit optimization with different finfet styles: Lower power possible at higher supply voltage, in: 2014 27th International Conference on VLSI Design and 2014 13th International Conference on Embedded Systems, 2014, pp. 476–482. doi:10.1109/VLSID.2014.89.
28. V. Hamolia, V. Melnyk, A survey of machine learning methods and applications in electronic design automation, in: 2021 11th International Conference on Advanced Computer Information Technologies (ACIT), 2021, pp. 757–760. doi:10.1109/ACIT52158.2021.9548117.
29. Y. Ma, Z. He, W. Li, L. Zhang, B. Yu, Understanding Graphs in EDA: From Shallow to Deep Learning, Association for Computing Machinery, New York, NY, USA, 2020, p. 119–
30. URL: https: //doi.org/10.1145/3372780.3378173.
31. https://www.vlsi-expert.com/2012/04/hierarchical-design-flow-part-1.html

ID: 29

NAVIGATING THE ALGORITHMIC LANDSCAPE: A COMPARATIVE ANALYSIS OF KEY MACHINE LEARNING STRATEGIES

Dr.G.NIRMALA
Assistant Professor,
Department of ECE,
Mahendra Institute of Technology,
Namakkal, Tamil Nadu, India.

Ms.V.ANNAPOORANI
Assistant Professor,
Department of ECE,
Mahendra Institute of Technology,
Namakkal, Tamil Nadu, India.

ABSTRACT:
Massive amounts of data are produced by Industry 5.0. Despite its limitations in data analysis and interpretation, machine learning excels in a number of other areas, including intelligent control, decision making, speech recognition, natural language processing, computer graphics, and computer vision. Both Deep Learning and Machine Learning perform well in engineering applications that need real-time processing. Machine learning is necessary for the development of intelligent and automated programs in the areas of health, cyber security, and transportation. Reinforcement learning, semi-supervised learning, unsupervised learning, and supervised learning are the several types of algorithms that are used in machine learning. The use of machine learning is improved here to better serve real-time engineering applications. This research demonstrates the potential of machine learning in a variety of fields, including intelligent transportation, healthcare, and cyber security. Research on machine learning and application management are both topics that are covered in this article. This study will give technical benchmarks for a variety of applications and scenarios that are found in the actual world to those who are decision-makers in the business, as well as academics.

Keywords: *Machine Learning, Industry 5.0, Data Analysis, Intelligent Control, Decision Making, Cyber Security.*

1. INTRODUCTION

1.1 Machine Learning Evolution

Real-time engineering creates tons of digital data [1, 2]. Semi- and unstructured data exist. Data insights make clever apps in many domains. Author-created data-driven cybersecurity [3]. Author designs context-aware smart app [1]. Data management speeds real-time engineering app insight.AI's machine learning smartens computers and data analysis [4]. Industry 4.0's machine learning allows apps learn from experience instead of programming [1, 3]. Industrial 4.0 automates tasks with machine learning [5]. Data-analyzing machine learning algorithms provide real-time engineering solutions. ML classification. Reinforcement, unsupervised, semi-supervised, and supervised learning. Algorithms and data power machine learning. Data-driven systems utilize reinforcement learning, association rule learning, reduction of dimensionality, feature engineering, data clustering, regression, and classification analysis [7, 8]. Deep Learning intelligently examines data [9]. Machine learning techniques have goals. Data confirms [10]. Selecting target domain solution learning methods is tricky. Real-time engineering requires ML. Improve application intelligence via machine learning. Machine learning will assist business and academia analyze and build data-driven intelligent systems for real-time engineering.

1.2. Types of Machine Learning Techniques.

Figure 1 displays a timeline of machine learning. Figure 2 illustrates the many types of machine learning, including reinforcement learning, semi-supervised learning, unsupervised learning, and supervised learning. Real-time engineering encompasses the utilization of all four machine learning (ML) approaches. In the following discourse, we provide a concise overview of these methods.

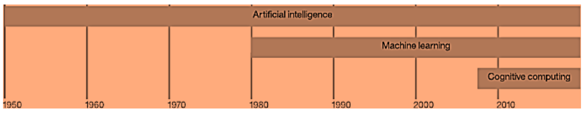


Figure 1 Timeline of machine learning [16]

Reinforcement learning is a computational approach that focuses on the interaction between an agent and its environment, with the aim of enabling the agent to learn and make decisions based on the feedback it receives from the environment. Reinforcement learning (RL) facilitates the automated analysis of optimum behavior in machines and software agents, hence enhancing efficiency [11]. The utilization of environmental data aids in the avoidance of penalties and the enhancement of incentives for RL. The user has provided a numerical reference, indicating the presence of a citation or source. The utilization of trained AI models for operational optimization or automation has the potential to boost intricate systems such as supply chain logistics, manufacturing, autonomous driving, robotics, and other related domains.The semi-supervised approach is a hybrid technique that utilizes both labeled and unlabeled data (3, 7).Given the presence of a substantial amount of unlabeled data and a limited amount of labeled data across different situations, the utilization of a semi-supervised technique appears to be appropriate for real-time applications (12).The performance of semi-supervised predictions surpasses that of predictions made only using labeled data. Activities such as text analysis, data processing, fraud detection, and machine translation are very prevalent.

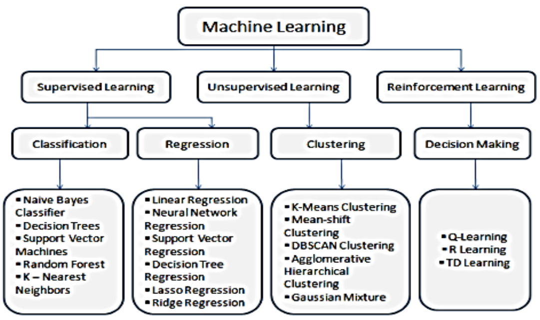


Figure.2 Types of Machine learning,[17]

In the context of machine learning, the supervised approach, similar to the one described in reference [7], involves the utilization of a function that maps input data to corresponding output values. The function is inferred through the utilization of labeled training samples and data. Supervised learning commences upon the identification of certain inputs [3].

2. STATE OF THE ART

Timeline of Machine Learning. Figure 3 illustrates the many types of machine learning approaches, including reinforcement learning, semi-supervised learning, unsupervised learning, and supervised learning. Real-Time Engineering incorporates all four machine learning methodologies.

The concept of environment-driven reinforcement refers to the phenomenon in which an individual's behavior is influenced and shaped by the surrounding environment. Reinforcement learning (RL) has been shown to enhance the behavior of machine and software agents [11]. The utilization of environmental data enables RL to mitigate the risk of incurring financial penalties. The user has provided a numerical reference. The utilization of trained AI models for operational optimization or automation has the potential to boost intricate systems such as supply chain logistics, manufacturing, autonomous driving, robotics, and other related domains.

The semi-supervised approach makes use of both labeled and unlabeled data (3, 7).Real-time semi-supervised learning has high efficacy when applied in scenarios where there is an abundance of unlabeled data and a scarcity of labeled data. According to reference [12].Semi-supervised predictions exhibit superior performance compared to predictions made only using labeled data. Data, fraud detection, machine translation, and text are often encountered phenomena in several domains.

Unsupervised learning refers to the process of doing data-driven analysis on datasets that lack labeled information, with minimum human intervention. The user has provided a numerical reference. Unsupervised algorithms perform the tasks of categorization, pattern identification, and extraction of generic properties. Unsupervised tasks encompass a range of activities such as identifying anomalies, establishing association rules, reducing dimensionality, learning features, analyzing density, and performing grouping.

Similar to the concept discussed in reference [7], the supervised approach requires the presence of an input-output function. The function is inferred based on the labeled training samples and data. The process of identifying inputs initiates the process of supervised learning [3].Practical challenges encountered in everyday life. Machine learning leverages client data for its operations. The educational program adheres to the prescribed curriculum and adapts accordingly. Data-adaptive algorithms that exhibit unpredictability. Digital assistants equipped with context-reading capabilities has the ability to peruse emails and extract pertinent information. This study aims to forecast client behavior. It facilitates comprehension for customers and enables a proactive approach to addressing their needs. Machine learning is utilized by several sectors.

2.1 An Introduction to Fundamentals of Cybersecurity.

The majority of individuals utilize the internet. According to article [13], as of 2017, over 48% of the global population consumes the Internet as a source of knowledge. According to the cited source [14], industrialized nations have reported an 82% statistic. The user's text is already academic.The Internet facilitates the interchange of data between devices, networks, and computers. The proliferation of mobile and computer system developments has significantly contributed to the increased use of the internet. The prevalence of online information consumption has led to a heightened focus by hackers on targeting such platforms. The user has provided a numerical reference. The implementation of stable computer systems is crucial for safeguarding data. The article [15] asserts that unauthorized network access has a detrimental impact on regular operations and compromises the integrity and security of computer systems. Cybersecurity measures are implemented to protect user assets and maintain the security of cyberspace. The protection of data is ensured by the implementation of cybersecurity measures, as referenced in [13].

2.2 An Overview of ml IN Health.

The use of Deep Learning and Machine Learning has significant implications for the domains of government, transportation, and business. The field of Deep Learning has had significant growth during the past decade. The field of deep learning has significantly improved the capabilities of speech recognition, natural language processing, and computer vision. In recent times, scholars have employed Deep Learning/Machine Learning methodologies in the field of healthcare [12]. The researchers demonstrated proficiency in many tasks, such as brain tumor segmentation [13], medical picture reconstruction [20, 21], lung nodule detection [22], lung illness categorization [14], and body component recognition [12]. The integration of deep learning and machine learning techniques is expected to significantly improve the performance of second-opinion computer-aided diagnosis (CAD) systems and other radiology support systems [13, 14]. The utilization of big data, mobile connectivity, edge computing, and cloud computing enables healthcare applications to leverage deep learning and machine learning techniques [27]. According to the cited source [28], the utilization of these technologies enhances the precision of predictions and facilitates the development of intelligent solutions that prioritize human needs and preferences.

Intelligent Transportation Systems. Intelligent transportation systems incorporate the utilization of sensor, communication, and information technologies [15]. Smart cities effectively oversee and regulate several aspects of urban transportation, including road traffic, public transit systems, traveler information services, and the integration of autonomous autos [10].These services are expected to enhance societal well-being, promote energy efficiency, and mitigate pollution.

3. Recent Works on Real-Time Engineering Applications

Machine learning algorithms and models let computers learn and decide. Urbanites endure traffic, air pollution, road injuries, and delays. Rapid car expansion, poor transportation infrastructure, and minimal road safety laws are producing serious environmental and quality-of-life challenges in metropolitan areas. Big automobiles clog many cities. Bicyclists' clothing, posture, partial occlusions, and viewing angles impair ML systems' detection rates, causing many near misses. Large datasets from numerous sources are increasingly analyzed and visualized using machine learning and deep learning. This trend increases pedestrian, bicycle, special vehicle (emergency vehicles, huge trucks), and License Plate Recognition (LPR) categorisation and recognition. This growing desire seeks safer and more sustainable environments. Deep models must reflect several visual appearances. Artificial neural networks mimic the brain and do deep learning through linked nodes. Other-level weighted connections will link close nodes.Input and weight parameters determine node activation function output values.

Cybersecurity protects computers, networks, and data. Machine learning studies techniques and models that let computers learn and adapt. Cybersecurity, design, manufacturing, medical, education, and finance use AI and ML. Machine learning detects black web/deep web sites, phishing, viruses, fraud, and spam. Cybersecurity must evolve. Machine-learning suits evolutionary assaults. Machine learning predicts cyberthreats on the black web, a hacker-dominated social network. This study's [11]. A research found that machine learning and social network data could anticipate corporate hacks. 53 darkweb forums estimated this. 135–138 advanced the topic. Antivirus, firewall, unified threat management, intrusion prevention, and SIEM protect cyberspace. Post-cyberattack responsiveness, performance, and mistake rate are better with AI-driven cyber security [14]. According to study, 60% of hacks are detected afterward [13]. ML powers cybersecurity and cyberattacks. ML prevents, detects, and improves cyber security [14, 15]. ML helps hackers bypass firewalls and locate vulnerabilities.

4. Current Challenges on Machine Learning Technology

Despite challenges, machine learning benefits businesses worldwide. Pattern-finding machine learning cannot generalize.User "algorithm weariness" is another issue. GPUs feed ML models. Real-time engineering requires sophisticated ML algorithms. One model cannot manage all jobs in all domains in real-time engineering applications. Most real-time engineering applications bypass early concerns with ML. ML predicts disease and terrorism. as in [13]. ML techniques are employed in numerous sectors, some as an alternative to accuracy and speed . ML demands precision. Avoid data leaks to create confidence. [15]. GPS and phones must track ML moving items. Article [13]compares object location data from several sources. According to article [14], network delays make location data from many sources confusing and require ML verification. A consumer-provider web trust ontology is suggested [248]. Trustworthiness classifies text. [15] advocates merging trustworthiness with semantic and practical text meanings. Software trustworthiness metrics [250]. Power-aware ML can reduce power use in enterprises and data centers. [15]. Dynamic machine shutdown eliminates use. Forecasts control machine shutdowns.

5. MACHINE LEARNING APPLICATIONS

The preference for machine learning in the industry stems from its ability to make intelligent decisions and acquire knowledge from previous experiences. In this context, we conduct evaluations of machine learning applications.

5.1. The Role of Intelligent Decision-Making and Predictive Analytics in Academic Research.

Data-driven predictive analytics and machine learning enable smart judgments [2]. Post-crime predictive analytics can identify credit card theft and offenders. Recording predicted variable-explanatory factor associations does this [7]. Predictive analytics and intelligent decision-making may assist retailers eliminate out-of-stocks, manage inventory, monitor consumer behavior, and optimize logistics and warehouse operations. These disciplines employ SVMs, Decision Trees, and ANNs [8, 10]. Social media, transportation, sales and marketing, healthcare, financial services, banking, telecommunications, e-commerce, and others benefit from accurate prediction. Transportation, cybersecurity, and prediction. Every nation requires mobility. Traffic in various regions globally causes crises, accidents, CO_2 emissions, increased fuel costs, traffic congestion, and delays [8]. Thus, a smart city needs an ITS to predict and manage traffic. Deep learning and machine learning improve traffic forecasts. Machine learning may assist transportation companies identify route concerns and recommend routes based on prior travel data. Forecasting and visualization may enhance traffic management, sustainable transportation, and real-world disruptions.

5.2. The Analysis of Sentiment and Natural Language Processing (NLP).

Natural Language Processing (NLP) refers to the computational techniques employed to enable computers to comprehend and interpret human language [2]. Natural Language Processing (NLP) facilitates the ability of computers to perceive, interpret, and comprehend textual and auditory information, enabling them to assess sentiment and prioritize tasks through the utilization of machine learning techniques. Natural Language Processing (NLP) encompasses several applications such as language translation, document summarization, speech recognition, conversational agents, and virtual personal assistants. Sentiment analysis, a subfield of natural language processing (NLP), collects public sentiment from many sources such as news articles, online forums, social media platforms, product evaluations, and blogs [6]. Sentiment analysis employs machine learning techniques to ascertain the polarity (neutral, negative, positive) and emotions (disinterest, interest, anger, extreme sadness, sadness, pleasure, and great satisfaction) associated with a given text.

5.3. The Recognition of Images, Speech, and Patterns.

The utilization of machine learning is prevalent in the field of image recognition [7] for the purpose of object identification. Image recognition encompasses several applications such as tagging recommendations on social media, detecting faces, recognizing characters, and designating cancerous areas in X-ray images. Popular speech recognition models include Alexa, Siri, Cortana, Google Assistant, and several others [3]. Pattern recognition refers to the process of automatically identifying and detecting patterns and regularities within data, such as the study of images [2]. This discipline employs techniques like as classification, feature selection, clustering, and sequence labeling.

5.4. The Field of Agriculture.

Agriculture is a fundamental requirement for all human activities . Sustainable agriculture has been found to enhance productivity levels while concurrently mitigating the adverse effects on the environment [279–281]. Sustainable agriculture employs supply networks that are reliant on knowledge-intensive practices. Machine learning has the potential to be applied at several stages of agricultural processing, production, preproduction, and distribution. These stages encompass consumer analysis, inventory management, production planning, and demand forecasting. Mobile information systems refer to the technological infrastructure and applications that enable the access, storage, and retrieval of information using mobile devices. These systems play a crucial role in facilitating communication The user mentions a range of topics related to agriculture, including livestock estimation, soil nutrient management, weed identification, disease detection, weather forecasting, irrigation requirements, soil characteristics, and agricultural productivity.

6. CONCLUSION AND FUTURE SCOPRE

This research on machine learning algorithms encompasses several applications and the analysis of data. In this discussion, we will provide a concise examination of the practical challenges encountered in real-world scenarios, as well as the manner in which learning algorithms effectively address these obstacles.

The efficacy of the model is contingent upon the utilization of machine learning methodologies and the inherent characteristics of the data.

In order to generate informed conclusions, machine learning algorithms require knowledge in the target application and access to real-world data. This paper examines the suitability of machine learning (ML) techniques for addressing practical challenges in several sectors of application. In conclusion, this study emphasizes the identification of research paths and the identification of difficulties.

In order to be deemed effective, solutions must successfully address all challenges pertaining to the target applications. This study offers a comprehensive evaluation for decision makers across many application domains and real-world contexts, catering to both industry practitioners and academic researchers.

Machine learning encompasses a wide range of applications. The phenomenon is seeing growth across several sectors including banking, finance, information technology, media, entertainment, gaming, and the automotive industry. Due of its expansive nature, Machine Learning is now being pursued by researchers with the aim of revolutionizing several domains.

REFERENCES

1. Sarker, I. H., Hoque, M. M., Uddin, M. D. K., & Tawfeeq, A. (2020). Mobile data science and intelligent apps: concepts, AI-based modeling and research directions. *Mobile Networks and Applications, 26*(1), 1–19.
2. Mahdavinejad, M. S., Rezvan, M., Barekatain, M., Adibi, P., Barnaghi, P., & Sheth, A. P. (2018). Machine learning for Internet of Things data analysis: a survey. *Digital Communications and Networks, 4*(3), 161–175.
3. Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: an overview from machine learning perspective. *Journal of Big Data, 7*(1), 29.
4. Sarker, I. H., & FurhadNowrozy, R. (2021). AI-driven cybersecurity: an overview, security intelligence modeling and research directions. *SN Computer Science, 2*(3), 173.
5. Slusarczyk, B. (2018). Industry 4.0 - are we ready? *Polish Journal of Management Studies, 17*(1), 232–248.
6. Google. (2019). Trends In. https://trends.google.com/trends/
7. Han, J., Pei, J., & Kamber, M. (2011). *Data Mining: Concepts and Techniques*. Elsevier.
8. Witten, I. H., & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.
9. Sarker, I. H. (2021). Deep cybersecurity: a comprehensive overview from neural network and deep learning perspective. *SN Computer Science, 2*(3), 154.
10. Sarker, I. H., Watters, P., & Kayes, A. S. M. (2019). Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage. *Journal of Big Data, 6*(1), 28.
11. Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: a survey. *Journal of Artificial Intelligence Research, 4*, 237–285.
12. Jhaveri, R., Ramani, S., Srivastava, G., Gadekallu, T. R., & Aggarwal, V. (2021). Fault-Resilience for Bandwidth Management in Industrial Software-Defined Networks. *IEEE Transactions on Network Science and Engineering, 8*(4), 3129–3139.
13. ICT-Fact. (2020). ICT-Fact and Figures 2017 Accessed. https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017.pdf
14. ICT-Facts. (2017). Telecommunication Union. https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx
15. Bhattacharyya, D. K., & Kalita, J. K. (2013). *Network Anomaly Detection: A Machine Learning Perspective*. Chapman & Hall.
16. IBM Developer. (n.d.). Beginner's Guide to Machine Learning, AI, and Cognitive Computing. https://developer.ibm.com/articles/cc-beginner-guide-machine-learning-ai-cognitive/
17. Analytics Vidhya. (2021). Everything You Need to Know About Machine Learning. https://www.analyticsvidhya.com/blog/2021/03/everything-you-need-to-know-about-machine-learning/

ID: 30

NAVIGATING THE DIGITAL LANDSCAPE: A MACHINE LEARNING DRIVEN NETWORKING PARADIGMS

Ms.J.BANUPRIYA
Assistant Professor,
Department of ECE,
Mahendra Institute of Technology,
Namakkal, Tamil Nadu, India.

Mr.S.DINESH KUMAR
Assistant Professor,
Department of ECE,
Mahendra Institute of Technology,
Namakkal, Tamil Nadu, India.

ABSTRACT:
Practical implementations of machine learning (ML) have solved many problems and enabled automated operations in many fields. Data accessibility, machine learning advancements, and computer capacity development promote this phenomenon. Machine learning (ML) has successfully addressed several common and challenging network administration and management concerns. Machine learning has been used to numerous networking fields and specialized network technologies. The investigation examined several machine learning techniques used in important networking domains across various network technologies. This approach allows readers to learn about various learning paradigms and machine learning methods used to solve networking problems. This domain has challenges in traffic prediction, routing and classification, congestion control, resource and fault management, QoS and QoE management, and network security. This paper also examines limits, provides valuable views, suggests research topics, and highlights the future of machine learning in networking. Thus, this study examines ML's effects on networking. This paper analyzes how machine learning is expanding autonomic network operation and administration.

KEYWORDS: *Machine Learning, Traffic Prediction, Traffic Classification, Traffic Routing, Congestion Control, Resource Management, Fault Management*

I. INTRODUCTION

Machine learning examines data. It involves learning and applying. ML uses "training" data's hidden patterns. Patterns identify unknown data. Reprogramming automation. ML builds data-fitting algorithms. Reborn ML. ML algorithms were rigid [1]. Recent ML breakthroughs have made these strategies flexible and adaptable to many real-world scenarios, from remarkable to routine. ML enhanced medical imaging and computer-aided diagnosis. Standard ML. ML helps search engines rank, crawl, suggest queries, and fix spelling. ML approaches will become more relevant as home automation and autonomous autos increase. Others revive ML. ML success requires data[2].IoT's billions of linked devices will boost data [3].ML understands data generation and hidden patterns. Modern computers store and test ML models.

Received large data. Cloud, GPUs, and TPUs speed big data training and inference. Smartphones infer from ML models. Despite these developments, network operations and administration are difficult, and human error causes most network faults [4]. Network failures cost providers money and reputation. Thus, powerful autonomic networks are sought [5]. ML has unique difficulties, but network operation and administration need cognitive control [6]. Each network is unique. Corporate networks differ. Thus, network patterns may vary. Network dynamics prevent set operating and management patterns. Network applications and devices have expanded, making manual network administration nearly impossible.

1.1 A holistic view of machine learning in networking

We examine 20 years of high-impact,

peer-reviewed literature. This survey covers networking developments. Citations, merit, and publication year decide selection. The same-year A and B have citation counts x and y. x > y. B can be addressed if it gives new ideas, insights, or lessons. ML networking. ML networking benefits and drawbacks. We integrate networking regardless of technology. This broadens readers' networking problem-solving across technologies. – Challenges and research. ML-based networking solutions reveal research barriers to network operation and administration. Discussing these chances boosts work and networking.

II. NETWORKING MACHINE LEARNING

Arthur Samuel called "Machine Learning" "the field of study that gives computers the ability to learn without being explicitly programmed" in 1959. ML clusters, classifies, regressions, and extracts rules [7]. Clustering isolates similar data, while classification and regression handle discrete or continuous output. Extracting rules detects statistical relationships. ML techniques apply to diverse domains. Data mining in huge databases [8] requires careful analysis of characteristics, variables, invariants, and transformations. ML anticipates events with big datasets.

Machine learning (ML) encompasses several learning models, including semi-supervised learning, unsupervised learning, and reinforcement learning. These paradigms have had an impact on various aspects of research, including data collection, the objective is to do dataset inference. When the user possesses comprehension of the data description, the labels get associated with the training data. A class of interest is typically encountered. There is evidence of generative and discriminative learning. Learning approaches are based on Bayes' theorem for conditional probability and the basic rule that links joint probability to conditional probability. The networking machine learning algorithm is highlighted in table.

Table 1 Machine learning algorithm used in networking

Sl.No.	Algorithm	Highlights
1	A Survey of Machine Learning in Big Data Networking"	Focuses on applying machine learning to big data networking challenges. Discusses key ML techniques, applications, and their impact on network optimization, security, and management. Explores open research issues and future directions.
2	"Machine Learning in Wireless Sensor Networks"	Surveys the application of machine learning in wireless sensor networks (WSNs). Discusses ML algorithms for WSNs, including classification, clustering, and regression. Explores energy efficiency, fault detection, and data aggregation.
3	"Machine Learning Techniques in Network Traffic Data Analysis"	Reviews the use of machine learning for network traffic analysis. Covers traffic classification, anomaly detection, and intrusion detection. Discusses various ML algorithms, feature selection, and challenges in real-time analysis.
4	"Machine Learning in Software-Defined Networking"	Explores the integration of machine learning with software-defined networking (SDN). Discusses ML techniques for SDN management, traffic engineering, resource allocation, and security. Highlights opportunities and challenges in this emerging field.
5	"Deep Learning Techniques for Network Traffic Feature Selection and Classification"	Focuses on deep learning techniques for network traffic analysis. Surveys methods for feature selection, classification, and intrusion detection using deep neural networks. Discusses the potential of deep learning in improving network security.

2.1 Data Collection

Data collection is the process of collecting and documenting research data. In order to effectively network machine learning models, it is imperative to utilize data that is representative and, ideally, devoid of bias. The collecting of data is of utmost importance due to the presence of datasets that are distinctive to subjects and periods. The collecting of data using online and offline methods is a topic of interest in the field of research, with a total of 460 studies conducted on this subject. The process of training and testing models might involve the utilization of extensive historical data sets that have been obtained in an offline manner. The utilization of online real-time network data has the capability to provide input to models or facilitate their retraining. Monitoring and measuring technologies are utilized to collect data from both offline and online sources. Utilize the provided tools to regulate the data sample rate, monitoring duration, and placement (inside the network core or network edge). CiscoNetFlow, IPFIX, and SNMP are network monitoring protocols. The monitoring process may be categorized into two main approaches: passive monitoring and active monitoring [12]. Active monitoring involves the transmission of probe packets for the purpose of data collection. Passive monitoring is employed to observe network traffic. The presence of injected traffic significantly increases the need for bandwidth in active monitoring processes. The elimination of this cost is achieved by passive monitoring, which necessitates the examination of network traffic by devices in order to extract valuable data.

2.2 Feature Engineering

The raw data exhibits a high level of noise or incompleteness. The process of acquiring knowledge necessitates the use of data preparation techniques. Feature extraction is a fundamental step in the modelling process. These two processes, learning and inference, are distinct from one other. There is a plethora of networking features available. Granularity is a process of classification. Packet-level characteristics like as mean, root mean square (RMS), variance, and Hurst are derived or generated from the packets that have been gathered. The act of packet sampling has an effect on the characteristics of features, but it does not alter the statistical qualities at the level of individual packets [10]. Flow-level characteristics are determined by computing the average flow duration, the number of packets per flow, and the number of bytes per flow [11]. Transport layer connection-level characteristics are responsible for providing connection-oriented information. Flow-level features can utilize two important parameters from the TCP connection header, namely the throughput and the advertised window size. Significant-quality data is generated, but, there are notable computational expenses and asymmetries in sampling and routing [12]. Machine learning (ML) encompasses the process of feature engineering, which involves the selection and extraction of relevant features. The proposed approach aims to reduce the dimensionality of massive data and identify differentiating features, hence mitigating computational expenses and enhancing the accuracy of machine learning models. The process of selection serves to reduce unnecessary elements [13]. Redundant features lead to overfitting, whereas irrelevant features result in an increase in computational cost without providing any improvement in accuracy. The computing cost of feature extraction can be significantly increased by the utilization of entropy, Fourier transform, and principal component analysis (PCA).

2.3 Performance Metrics and Model Validation

Machine learning is a multidisciplinary domain that encompasses several academic fields, such as computer science, mathematics, philosophy, economics, neurology, psychology, and control theory [14]. Throughout the past 75 years, several machine learning approaches have been developed (15, 16). The concept of the artificial neuron has been crucial in current research on neural networks. The inclusion of weighted human brain connections is crucial for the improvement of this prototype. In 1949, the utilization of Hebbian learning [18] was implemented to modify the synaptic weights inside the neural network model. The utilization of Hebbian learning principles and neuron models has significantly helped to the progress and development of neural networks. The initial development of the neural network computer, referred to as SNARC, took place in 1950 and incorporated two fundamental principles [17]. Alan Turing devised a method to assess the cognitive capacities of a computer by formulating a test in which the computer attempts to replicate human behavior.For a thorough examination of the difficulties encountered by the author in connection with their "learning machine" concept, please refer to the supplied source. Following these developments, a period of twenty years was devoted to the scientific exploration of machine learning.

III. TRAFFIC PREDICTION

Traffic prediction is a crucial aspect of transportation planning and management. It involves forecasting the future traffic conditions on road networks, which helps in the prediction of network traffic is of significant importance in the operations and administration of modern networks, which are characterized by their growing complexity and diversity. The task involves predicting future traffic patterns and has historically been approached using time series forecasting (TSF) methods. The primary goal of Time Series Forecasting (TSF) is to develop a regression model that can effectively establish a precise link between projected traffic volume and past traffic levels. The current methods for traffic prediction may be categorized into two main types: statistical analysis models and supervised machine learning models. The construction of statistical analysis models often relies on the generalized autoregressive integrated moving average (ARIMA) model, although the bulk of learning for traffic prediction is accomplished using supervised neural networks (NNs). The ARIMA model is commonly utilized for time series forecasting (TSF), employing a combination of autoregressive (AR) and moving average (MA) models to conduct auto-regression on the differenced and "stationarized" data. Nevertheless, due to the exponential expansion of networks and the escalating intricacy of network traffic, conventional TSF models appear to be compromised, leading to the emergence of more sophisticated ML models. In recent times, there has been a focus on minimizing unnecessary costs and enhancing the precision of predicting intra-traffic by using information derived from flows, apart from only considering traffic volume.

3.1 Traffic prediction as a pure TSF problem

According to existing information, Yu et al. [18] were the initial researchers to utilize machine learning in traffic prediction by employing a multilayer perceptron neural network (MLP-NN). The main objective of the researchers was to enhance the precision of conventional augmented reality (AR) techniques. The above examples showed that Cybenko and Hornik's SLP-NN approximators, which use a sufficient number of neurons with continuous sigmoidal activation, are universal. This allows them to approximate any continuous function with any precision.

3.2 Traffic prediction as a non-time series forecasting issue

Different approaches and supplementary data can be employed to predict network traffic instead of relying just on time series forecasting. Li et al. conduct a study on network traffic flows in the frequency-domain, as an alternative to assessing traffic volume. In order to approximate the volume of incoming and outgoing traffic through an inter-data centre connection characterized by elephant flows, or substantial data transfers, it is necessary to do an estimation. The researchers utilized Feedforward Neural Networks that were trained via BP (Backpropagation) in their inquiry. The training process employs fundamental gradient descent and wavelet transform techniques to capture the temporal and frequency characteristics of traffic time series data. The prediction model utilizes many feature dimensions in order to forecast the movement patterns of elephants. The cost of collecting elephant flows with a high frequency is higher than the cost of collecting byte count for traffic volume. Therefore, the acquisition of elephant flow data is conducted at more frequent intervals and afterwards interpolated to address missing values, thereby facilitating the process of data gathering. The dataset consists of 30-second intervals of incoming and outgoing traffic data from the SNMP counters of Baidu's data center (DC) edge routers, as well as statistics on inter-DC connections. The data was collected over a period of six weeks. Data pertaining to both incoming and outgoing traffic is collected at regular intervals of 5 minutes. The five most popular applications account for 80% of the total traffic. Interpolation is a method that provides more accurate estimations for missing values by rescaling the data to a 30-second interval. The wavelet transform at level 10 partitions the time series into 120 distinct features for each individual timestamp.

IV. TRAFFIC CLASSIFICATION

In this section, we will discuss the concept of traffic classification. Internet traffic may be classified through the use of port numbers, payload analysis, and host-based techniques. Port-based techniques are often regarded as antiquated and unreliable by network security professionals. The primary contributing factors to this phenomenon include dynamic port negotiation, tunnelling, and the improper utilization of well recognized application port numbers. These methodologies obscure network traffic and circumvent firewalls [18]. Nevertheless, payload-based approaches evaluate the payloads of applications. The accuracy and computational complexity of traffic categorization algorithms in supervised and unsupervised machine learning are hindered by the presence of encryption. In the context of network traffic classification, it is commonly observed that handshake payloads tend to lack encryption. The utilization of high data rate lines renders this approach unfeasible. Nevertheless, the monitoring of long-lived UDP traffic may be conducted by payload analysis. The methodology employed in this study involves the utilization of a random cargo inspection technique within an observation window, as outlined in reference [19]. The value of this tool is constrained and its effectiveness is contingent upon the size of the observation window. Similar to host-based traffic categorization, routing asymmetries can have an impact on it. Traffic categorization algorithms that utilize flow characteristics examine the entirety of the communication session, in contrast to previous perspectives. Every network packet is assessed in a one-way route. The field of traffic categorization in academia has extensively investigated the use of supervised and unsupervised machine learning techniques. Supervised learning has witnessed remarkable accuracy through the utilization of kernel estimation, neural network (NN), and support vector machine (SVM) machine learning techniques. Traditional kernel estimating techniques are commonly used and have proven to be efficient. However, it is worth noting that these approaches may have assumptions that are overly stringent when applied to traffic categorization. The National Bureau of Knowledge Exploration (NBKE) has been the subject of extensive study in the field of traffic classification. The utilization of probabilistic or Bayesian weights in neural network (NN) traffic categorization has been found to enhance accuracy. In the initial investigations, researchers utilized traditional and multi-class methodologies. The utilization of Support Vector Machines (SVM) has been proposed as a means to enhance the accuracy and applicability of traffic classification for large datasets. Network operators seldom possess complete application data. Hence, it is unfeasible to attain comprehensive knowledge of all traffic categorization applications in advance. Flow data-based unsupervised machine learning (ML) techniques have been extensively investigated in the realm of traffic classification. The field of unsupervised machine learning in traffic categorization has explored both hard and soft clustering algorithms. The presence of similarities in application flow poses challenges in achieving precise traffic categorization through the use of hard clustering techniques. Density-based clustering algorithms exhibit faster training times compared to EM-based approaches while still achieving the desired level of granularity in soft clustering.

V. MACHINE LEARNING IN NETWORK SECURITY

Machine learning in network security focused on network-based intrusion detection. The study classified network intrusion detection systems as misuse-, anomaly-, and hybrid. We illustrate machine learning methods in each area using the latest deep learning (DL) and reinforcement learning implementations. One key result is machine learning improving misuse-based intrusion detection. With improved capabilities, rule-based systems may analyze audit data to discover complex attack patterns. The program may spot attack pattern irregularities. Because they can extract decision rules, white-box models like decision trees (DT) are better for abuse detection than neural networks. Ensemble-based approaches trained numerous machine learning models utilizing many dataset subsets or feature sets. Ensemble-based approaches reduce training time. Machine learning (ML) in intrusion detection systems (IDS) has definite benefits, but considerable guesswork. Despite much machine learning anomaly detection research, its use is scarce [15]. Snort, a misuse-based IDS, is popular [10].The possibility of false positives (FPs) and major misclassification make anomaly detection unpopular.

VI. CONCLUSION

The preceding two decades have witnessed a favourable trajectory for machine learning in the field of networking.

Machine learning (ML) methodologies used to network operation and management primarily concentrate on the areas of traffic engineering, performance optimization, and network security. The present study involved a comprehensive review of pertinent scholarly works and an evaluation of the feasibility and applicability of suggested machine learning (ML) approaches for the forthcoming autonomous operation and administration of networks. The future networks will require a rise in traffic and the integration of connected gadgets in order to provide unparalleled access to and exchange of information. The complexity of traffic engineering operations, encompassing congestion control, prediction, classification, routing, faults, and security assaults, is expected to increase as a result of heightened uncertainty. Machine learning (ML) solutions have demonstrated promise in addressing traffic engineering challenges. Nevertheless, while scaling these solutions, it is crucial to take into account factors such as data volume, device count, and application use. The predominant focus of fault and security management systems based on machine learning is directed towards networks that are characterized by a single tenant and a single layer. In order to establish a fault and security management framework for forthcoming networks, it is imperative to enhance or restructure existing machine learning methodologies to accommodate multi-tenancy inside multi-layer networks. This survey encompasses a range of topics and potentialities, among others. The results of our study indicate that further investigation is required in order to progress the field of autonomic networking.

VII. REFERENCES

1. Aben E (2010) NLANR PMA data. https://labs.ripe.net/datarepository/data-sets/nlanr-pma-data, Accessed 27 Dec 2017
2. AckleyDH, HintonGE, SejnowskiTJ (1985) Alearning algorithm for boltzmann machines. Cognitive science 9(1):147–169
3. ACM Special Interest Group on KDD (2016) KDD cup archives. http://www.kdd.org/kdd-cup, Accessed 22 Nov 2017 4. AdamsR (2013) Active queue management: A survey. IEEECommunicationsSurveysTutorials 15(3):1425– 1476
4. Bakhshi T, Ghita B (2016) on internet traffic classification: A two-phased machine learning approach. Journal of Computer Networks and Communications 2016 32. Bakre A, Badrinath BR (1995) I-tcp: Indirect tcp for mobile hosts. In: Proceedings of the 15th International Conference on Distributed Computing Systems, IEEE Computer Society, Washington, DC, USA, ICDCS '95, pp 136–
5. Chakrabarti S, Chakraborty M, Mukhopadhyay I (2010) Study of snort-based ids. In: Proceedings of the International Conference and Workshop on Emerging Trends in Technology, ACM, pp 43–47
6. Chang CC, Lin CJ (2011) Libsvm: a library for support vector machines. ACM transactions on intelligent systems and technology (TIST) 2(3):27
7. Forster A, Murphy AL (2007) Froms: Feedback routing for optimizing multiple sinks in wsn with reinforcement learning. In: Intelligent Sensors, Sensor Networks and Information, 2007. ISSNIP 2007. 3rd International Conference on, IEEE, pp 371–376
8. Fraleigh C, Diot C, Lyles B, Moon S, Owezarski P, Papagiannaki D, Tobagi F (2001) Design and deployment of a passive monitoring infrastructure. In: Thyrrhenian Internatinal Workshop on Digital Communications, Springer, pp 556–575
9. Goetz P, Kumar S, Miikkulainen R (1996) On-line adaptation of a signal predistorter through dual reinforcement learning. In: ICML, pp 175–181
10. Hochreiter S, Schmidhuber J (1997) long short-term memory. Neural Comput 9(8):1735–1780
11. Kelley HJ (1960) Gradient theory of optimal flight paths. ARS Journal 30(10):947–954
12. Leela-Amornsin L, Esaki H (2010) Heuristic congestion control for message deletion in delay tolerant network. In: Proceedings of the Third Conference on Smart Spaces and Next Generation Wired, and 10th International Conference on Wireless Networking, Springer-Verlag, Berlin, Heidelberg, ruSMART/NEW2AN'10, pp 287–298
13. Mahmoud Q (2007) Cognitive Networks: Towards Self-Aware Networks. Wiley-Interscience 292.
14. Nguyen T, Armitage G (2006) Synthetic sub-flow pairs for timely and stable ip

traffic identification. In: Proc. Australian Telecommunication Networks and Application Conference

15. Nguyen TT, Armitage G (2006) Training on multiple sub-flows to optimise the use of machine learning classifiers in real-world ip networks. In: Local Computer Networks, Proceedings 2006 31st IEEE Conference on, IEEE, pp 369–376 Malware-Traffic-Analysisnet (2017) A source for pcap files and malware samples. http://www.malware-t raffic-analysis.net, Accessed 15 Dec 2017
16. Roughan M, Sen S, Spatscheck O, Duffield N (2004) Class-of-service mapping for qos: a statistical signature-based approach to ip traffic classification. In: Proceedings of the 4th ACM SIGCOMM conference on Internet measurement, ACM, pp 135–148
17. Sutton RS, Barto AG (1987) A temporal-difference model of classical conditioning. In: Proceedings of the ninth annual conference of the cognitive science society, Seattle, WA, pp 355–378
18. Wang Z, Zhang M, Wang D, Song C, Liu M, Li J, Lou L, Liu Z (2017) Failure prediction using machine learning and time series in optical network. Optics Express 25(16):18,553–18,565
19. Zineb AB, Ayadi M, Tabbane S (2015) Cognitive radio networks management using an anfis approach with qos/qoe mapping scheme. In: Networks, Computers and Communications (ISNCC), 2015 International Symposium on, IEEE, pp 1–6

ID: 31

SAFEGUARDING DATA IN THE DIGITAL AGE: UNVEILING THE NEXT WAVE OF SECURITY TRENDS

Ms.ANIT SEBI
Research Scholar,
R.V.S College of Arts and Science,
Coimbatore, Tamil Nadu, India.

ABSTRACT:
This Chapter investigates arising patterns in information security, featuring creative systems to address developing difficulties. The ascent of quantum cryptography and homomorphic encryption offers arrangements against quantum dangers while handling encoded information. Computer based intelligence driven danger recognition upgrades ongoing network safety, and the zero-trust engineering rethinks access control. Biometric confirmation and security safeguarding advancements guarantee strong distinguishing proof and consistence with information assurance guidelines. The theoretical embodies quantum cryptography, homomorphic encryption, computer based intelligence danger recognition, zero-trust design, biometric confirmation, and protection safeguarding advances as crucial patterns in information security.

KEYWORDS: *Homomorphic Encryption, Biometric Authentication, Quantum Cryptography, Zero-Trust Architecture, Privacy Preserving Technology.*

I. INTRODUCTION

In an undeniably interconnected world, information security has turned into a central worry for people, organizations, and legislatures the same. The quick development of innovation carries with it new difficulties and potential open doors, as well as a requirement for novel ways to deal with defending delicate data. This section investigates the arising patterns in information security, revealing insight into the imaginative techniques that are molding the fate of this basic field.

1. QUANTUM CRYPTOGRAPHY: DEFENDING AGAINST QUANTUM THREATS

The appearance of quantum processing presents both a commitment and a hazard to information security. While quantum PCs can possibly break a considerable lot of the present encryption strategies, they likewise offer an answer: quantum cryptography. Utilizing the standards of quantum mechanics, this approach guarantees the security of information by utilizing quantum key circulation (QKC) to empower secure correspondence channels. As quantum PCs become more refined, incorporating quantum-safe calculations into existing frameworks becomes basic to upset expected assaults.

2. HOMOMORPHIC ENCRYPTION

Homomorphic encryption addresses a forward leap in information security, empowering calculations to be performed on encoded information without the requirement for decoding. This development has sweeping ramifications, especially in situations where delicate information should be handled while keeping up with security. Businesses, for example, medical care and money stand to profit from this innovation, as it empowers secure information examination and sharing without compromising privacy.

3. AI AND MACHINE LEARNING FOR THREAT DETECTION

Man-made reasoning (simulated intelligence) and AI are becoming critical in recognizing and alleviating security dangers progressively. These innovations dissect tremendous datasets to distinguish designs characteristic of cyber-attacks, extortion, or other malevolent exercises. High level calculations might adjust and develop to perceive new and advancing dangers, making a powerful safeguard component against steadily changing digital dangers.

4. ZERO-TRUST ARCHITECTURE: VERIFYING EVERY USER AND DEVICE

The conventional edge based security model is demonstrating deficient in the present complex organization conditions. Zero-trust engineering advocates for a "never trust, consistently check" approach, where clients and gadgets are ceaselessly verified and approved prior to getting to assets. This pattern requires a shift from the customary palace and-canal way to deal with a more granular and versatile security system that better lines up with the real factors of distributed computing and remote work.

5. BIOMETRIC AUTHENTICATION AND BEHAVIORAL BIOMETRICS

Biometric confirmation, including unique finger impression, facial acknowledgment, and iris checking, is acquiring conspicuousness as a safe option in contrast to passwords. Moreover, the idea of conduct biometrics dissects client ways of behaving, like composing velocity and mouse developments, to lay out special computerized profiles. Consolidating these methodologies offers a diverse validation system that upgrades security while limiting client grinding.

6. PRIVACY - PRESERVING TECHNOLOGIES

As information protection guidelines become more severe, security safeguarding advancements are on the ascent. Methods, for example, differential protection add a layer of commotion to information, forestalling individual recognizable proof while keeping up with the general trustworthiness of factual investigations. Unified learning, then again, empowers AI models to be prepared across decentralized gadgets without the need to move crude information, in this manner protecting client security.

CONCLUSION

The scene of information security is developing at an exceptional speed, driven by arising advancements and the developing complexity of digital dangers. Quantum cryptography, homomorphic encryption, artificial intelligence driven danger location, zero-trust design, biometric validation, and protection saving innovations are reshaping the manner in which we approach information security. Remaining ahead in this unique climate requires a mix of imaginative systems, cooperative endeavours, and a pledge to versatility. By embracing these patterns, people and associations can explore the intricacies of information security and shield their computerized future.

ID: 32

EMERGING HORIZONS: UNVEILING THE FUTURE OF BLOCKCHAIN TECHNOLOGY

Ms.POURNAMI.S.LAL
Student,
B.A Economics,
Miranda House College,
New Delhi, India.

ABSTRACT:
This Chapter investigates the developing scene of block chain innovation, zeroing in on arising patterns that are forming its applications and effect. The conversation incorporates five key regions: interoperability, decentralized finance (DeFi), non-fungible tokens (NFTs), versatility arrangements, and the maintainability of block chain networks. The section additionally features the developing revenue in sovereign advanced monetary standards and national bank computerized monetary standards (CBDCs) as states and national banks investigate block chain's capability to modernize monetary frameworks. By looking at these patterns, the part gives understanding into the ground-breaking capability of block chain innovation across different areas.

Keywords: *Block chain, Emerging Trends, Interoperability, Decentralized Finance (DeFi), Non-Fungible Tokens (NFTs), Scalability Solutions, Sustainability, Central Bank Digital Currencies (CBDCs).*

INTRODUCTION

In Present days, block chain innovation has developed past its underlying application in cryptographic forms of money, like Bitcoin. Its capability to upset different ventures has started revenue and speculation from organizations, state run administrations, and tech lovers around the world. This section digs into the arising patterns in block chain innovation, featuring its extending use cases and the ground-breaking effect it is ready to bring.

1. INTEROPERABILITY: BRIDGING BLOCKCHAINS

One of the vital difficulties in the block chain biological system has been the absence of interoperability between various block chain networks. As the quantity of block chain stages develops, the need to interface and offer information across these organizations becomes apparent. Arising arrangements, for example, cross-chain conventions and interoperability structures, mean to address this test. By empowering consistent correspondence between different block chains, these arrangements improve the general utility of block chain innovation, opening ways to complex use cases that require coordinated effort between various organizations.

2. DECENTRALIZED FINANCE (DEFI): BEYOND THE HYPE

Decentralized Money, or DeFi, arose as one of the most encouraging patterns inside the block chain space. While it acquired consideration for its capability to upset conventional monetary frameworks, it likewise confronted analysis because of administrative worries and security issues. Notwithstanding, DeFi has kept on advancing, with projects zeroing in on upgrading security, further developing client experience, and integrating genuine resources into the decentralized biological system. The pattern has extended to incorporate decentralized protection, forecast markets, and loaning stages, indicating a future where conventional monetary mediators may be minimized.

3. NON-FUNGIBLE TOKENS (NFTS): DIGITAL OWNERSHIP AND CREATIVITY

Non-Fungible Tokens (NFTs) burst into the standard, altering how we see advanced proprietorship and credibility. At first promoted inside the craftsmanship and collectibles space, NFTs have since extended to areas like music, gaming, and, surprisingly, land. The innovation empowers makers to tokenize their computerized resources, demonstrating their possession on the block chain. As this pattern proceeds, we are probably going to see more inventive applications that influence NFTs to change businesses and empower new adaptation models for advanced content.

4. SCALABILITY SOLUTIONS: ENABLING MASS ADOPTION

Block chain's versatility issues, especially apparent out in the open organizations like Bitcoin and Ethereum, have frustrated its standard reception. Arising versatility arrangements, including layer-2 arrangements like Lightning Organization and sharding procedures, try to address these limits. These arrangements plan to upgrade exchange throughput and decrease inactivity, making block chain innovation more common-sense for regular use. As these arrangements mature, they could prepare for huge scope applications, for example, production network the executives, personality check, and Web of Things (IoT) gadgets.

5. SUSTAINABILITY AND ENERGY EFFICIENCY

The ecological effect of block chain innovation, especially Evidence of Work (PoW) agreement systems, has raised worries about its drawn out feasibility. Arising patterns are zeroing in on more energy-proficient agreement components, like Confirmation of Stake (PoS) and Verification of Power (PoA), which require fundamentally less energy to approve exchanges. Moreover, projects are investigating approaches to reuse block chain squander heat for viable applications, displaying a guarantee to maintainability inside the business.

6. SOVEREIGN DIGITAL CURRENCIES AND CENTRAL BANK DIGITAL CURRENCIES (CBDCS)

Legislatures and national banks overall are investigating the capability of block chain innovation to give sovereign advanced monetary standards. National Bank Computerized Monetary forms (CBDCs) mean to modernize and smooth out monetary frameworks, offering advantages, for example, further developed straightforwardness, decreased exchange costs, and improved financial arrangement execution. As CBDCs keep on getting forward momentum, they could reshape the worldwide monetary scene, affecting cross-line instalments, settlements, and monetary incorporation.

CONCLUSION

In Conclusion, the block chain scene is consistently developing, with arising patterns reshaping its applications and possible effect. Interoperability, DeFi, NFTs, versatility arrangements, supportability endeavours, and the coming of CBDCs are only a couple of instances of the unique movements happening inside the block chain environment. As these patterns mature and collaborate with one another, they hold the commitment of changing businesses, making new monetary models, and driving mechanical advancement in exceptional ways. Keeping up to date with these patterns will be vital for organizations, policymakers, and people hoping to explore the developing block chain scene.

ID: 33

NAVIGATING THE FUTURE: EMERGING TRENDS IN EXPERT SYSTEMS AND BUSINESS INTELLIGENCE

Ms.POURNAMI.S.LAL
Student,
B.A Economics,
Miranda House College,
New Delhi, India.

ABSTRACT:
This Chapter features the extraordinary patterns forming the fields of master frameworks and business intelligence (BI), driving associations towards information driven direction and development. The talked about patterns include expanded examination, offering mental experiences and mechanizing investigation; logical computer based intelligence, zeroing in on straightforward direction; half and half cloud and edge examination, upgrading information handling; Natural language Processing (NLP) in BI, democratizing information investigation; also, moral and dependable artificial intelligence, guaranteeing fair and consistent computer based intelligence frameworks. These patterns unite to rethink how organizations remove experiences, decide, and keep up with moral norms in the time of cutting edge innovation.

***KEYWORDS**: Augmented Analytics, Explainable AI, Hybrid Cloud, Natural Language Processing (NLP), Ethical AI.*

INTRODUCTION

In the steadily developing scene of innovation and business, master frameworks and business Intelligence (BI) have arisen as basic apparatuses for associations to acquire experiences, pursue informed choices, and accomplish upper hands. This section dives into the state of the art drifts that are moulding the fate of master frameworks and business knowledge, investigating their expected effect on organizations across ventures.

1. AUGMENTED ANALYTICS AND COGNITIVE INSIGHTS

Quite possibly of the main pattern in business knowledge is the ascent of expanded examination. Customary BI stages have principally centered on information representation and detailing, expecting clients to have a strong comprehension of information examination. Increased investigation, then again, use AI and regular language handling to robotize information arrangement, examination, and even give relevant experiences. This pattern not just engages non-specialized clients to investigate information yet additionally upgrades independent direction by uncovering stowed away examples and relationships that could slip through the cracks.

Mental experiences make increased investigation a stride further by integrating progressed computer based intelligence procedures. These frameworks can perceive patterns, irregularities, and suggest activities in light of verifiable information, accordingly assisting organizations with expecting market moves and pursue proactive decisions.

2. EXPLAINABLE AI AND TRANSPARENT DECISION-MAKING

As artificial intelligence and AI calculations become necessary to business processes, the interest for straightforwardness and responsibility in navigation has flooded. The idea of logical simulated intelligence intends to demystify the "discovery" nature of intricate calculations. This pattern centers around creating simulated intelligence models that can give exact expectations as well as make sense of how they showed up at those forecasts. This is especially vital in exceptionally managed businesses where choices should be auditable and reasonable.

3. HYBRID CLOUD AND EDGE ANALYTICS

With the rising volume of information being produced at the edge (IoT gadgets, sensors, and so forth.), associations are embracing edge investigation to process and dissect information nearer to the source. This pattern diminishes idleness and upgrades constant direction.

Simultaneously, there's a developing acknowledgment of the advantages of crossover cloud methodologies, permitting organizations to adjust their information handling needs between on-premises foundation and public or confidential cloud conditions. This mix empowers associations to improve execution, adaptability, and cost-adequacy

4. NATURAL LANGUAGE PROCESSING (NLP) IN BI

Natural language Processing has taken huge steps as of late, empowering frameworks to comprehend and decipher human language. In BI, NLP is changing the way in which clients connect with information. Rather than questioning data sets utilizing complex SQL proclamations, clients can now pose inquiries in regular language and get precise experiences. This pattern upgrades client openness and democratizes information investigation, empowering chiefs at all levels to separate important data without particular preparation.

5. ETHICAL AND RESPONSIBLE AI

The combination of artificial intelligence and master frameworks into business processes accompanies moral contemplations. The pattern of moral and capable simulated intelligence underlines the significance of decency, predisposition relief, and information protection. Associations are progressively putting resources into man-made intelligence administration structures to guarantee that their computer based intelligence frameworks are straightforward, unprejudiced, and agreeable with guidelines. This pattern mitigates possible dangers as well as upgrades trust among clients, partners, and representatives.

CONCLUSION

The scene of master frameworks and business knowledge is quickly developing, driven by headways in man-made intelligence, AI, and information examination. As associations explore these arising patterns, they have the amazing chance to open new degrees of proficiency, development, and key independent direction. By embracing expanded examination, straightforward artificial intelligence, edge examination, NLP, and moral contemplations, organizations can situate themselves at the very front of the computerized change venture, guaranteeing their supportability and progress in an undeniably information driven world.

ID: 34

BEYOND TEXT: NAVIGATING NEW HORIZONS IN SOCIAL SENSING AND ANALYSIS

Dr.V.AYSHA
Principal,
College of Applied Science
Ayalur, Kerala, India.

ABSTRACT:
This Chapter investigates the developing scene of social detecting and social examination, driven by mechanical progressions and creative approaches. Five key patterns are featured: multi-modular information combination, moral and protection contemplations, continuous occasion discovery, feeling investigation 2.0, and network elements. These patterns are moulding the way in which we figure out human way of behaving, cultural occasions, and connections in a comprehensive way. The mix of mental processing and logical investigation, alongside cross-space cooperation, further improves the field's true capacity. Moral ramifications and protection concerns are tended to close by the astonishing possibilities of utilizing multi-modular information for continuous bits of knowledge into opinion and occasion elements. Through these patterns, we gain a more profound comprehension of the many-sided texture of human culture, informed by multidisciplinary points of view.

Keywords: *Social Sensing, Multi-Modal Data Fusion, Real-Time Event Detection, Sentiment Analysis, Network Dynamics, Contextual Analysis, Ethical Considerations, Privacy, Cross-Domain Collaboration.*

I. INTRODUCTION

In Present Days, the scene of social detecting and social examination has seen critical development because of the fast headways in innovation, information assortment, and scientific strategies. This part digs into the arising patterns that are forming the field and changing how we comprehend and cooperate with the social world.

1. MULTI-MODAL DATA FUSION

Conventional social detecting techniques basically depended on printed information from web-based entertainment stages. Nonetheless, the development of multi-modular information, including pictures, recordings, brief snippets, and sensor information, has improved the data accessible for examination. The combination of these different information types permits specialists to acquire a more all-encompassing and nuanced comprehension of get-togethers and ways of behaving. For example, breaking down both text and pictures from a dissent occasion can give bits of knowledge into what individuals are talking about as well as the viewable signs and feelings they express.

2. ETHICAL AND PRIVACY CONSIDERATIONS

As the volume and granularity of information gathered through friendly detecting increment, so do concerns connected with morals and protection. Arising patterns underline the significance of integrating moral contemplations into the plan and execution of social investigation projects. Finding some kind of harmony between information utility and individual security is a key test. New strategies, like differential protection and united learning, are being investigated to empower significant examination while defending delicate data.

3. REAL-TIME AND EVENT DETECTION

Ongoing social detecting is picking up speed, empowering the location and observing of situation as they develop. From cataclysmic events to social developments, the capacity to quickly recognize and answer occasions utilizing virtual entertainment information can possibly save lives and drive more successful emergency the board. AI calculations, especially profound learning models, are being outfit to distinguish and order occasions from the ceaseless stream of social information consequently.

4. SENTIMENT ANALYSIS 2.0

Sentiment Analysis has developed past essential good/pessimistic arrangement. The field currently consolidates progressed methods to perceive feelings, mockery, incongruity, and social subtleties present in web-based entertainment posts. This more profound comprehension of feeling upgrades the exactness of measuring general assessment, which is significant for organizations, states, and associations expecting to tailor their methodologies and strategies.

5. NETWORK DYNAMICS AND INFLUENCE PROPAGATION

Social cooperation's happen in complex organizations, and understanding the elements of these organizations is significant for anticipating data dispersion, pattern engendering, and impact. Arising patterns center on network examination to reveal stowed away examples of impact, recognize key hubs, and anticipate how data spreads through interconnected networks. Such experiences track down applications in advertising, political missions, and the investigation of social peculiarities.

6. COGNITIVE COMPUTING AND CONTEXTUAL ANALYSIS

The mix of mental figuring, which incorporates normal language handling and AI, has raised the abilities of social examination frameworks. Relevant investigation, which considers the more extensive setting of a web-based entertainment post or communication, empowers a more exact understanding of importance and purpose. This pattern is driving the improvement of Chabot's, remote helpers, and suggestion frameworks that get it and answer clients in a more human-like way.

7. CROSS-DOMAIN COLLABORATION

Arising patterns in friendly detecting support cooperation across different areas, like social science, software engineering, brain research, and metropolitan preparation. Cooperative endeavours cultivate a multidisciplinary way to deal with critical thinking, prompting more exhaustive and significant bits of knowledge. By consolidating mastery from different fields, specialists can address complex cultural difficulties and foster arrangements that take care of a more extensive scope of viewpoints.

CONCLUSION

In Conclusion, the field of social detecting and social examination is quickly developing, filled by propels in innovation and information examination methods. These arising patterns are reshaping the way that we grasp human way of behaving, social associations, and worldwide occasions. As we explore the moral contemplations, tackle ongoing abilities, and extend our scientific strategies, what's to come guarantees invigorating chances to acquire extraordinary experiences into the unpredictable embroidered artwork of human culture?

ID: 35

INVESTIGATING IMAGINATIVE OUTSKIRTS IN BIOMEDICAL AND CLINICAL IMAGING

Dr.AYSHA.V
Principal,
College of Applied Science,
Ayalur, Kerala, India.

ABSTRACT:
The field of biomedical and clinical imaging is going through fast change because of mechanical progressions and imaginative exploration. This part investigates arising patterns like nanotechnology, computerized reasoning (simulated intelligence), expanded reality (AR), practical imaging, and quantum imaging. Nanotechnology empowers exact infection identification and designated treatments, while artificial intelligence improves diagnostics and prescient bits of knowledge. AR upsets careful route and preparing, and utilitarian imaging extends comprehension of neurological issues. Quantum imaging guarantees phenomenal awareness and goal. Moral and administrative contemplations are indispensable as these patterns reshape medical services. Interdisciplinary coordinated effort is vital to saddling these advancements for worked on persistent results.

KEYWORDS: *Biomedical Imaging, Nanotechnology, Artificial Intelligence, Augmented Reality, Quantum Imaging.*

INTRODUCTION

The field of biomedical and clinical imaging has been going through quick changes lately, determined by noteworthy mechanical progressions and inventive examination. From early recognition to customized treatment draws near, these arising patterns are ready to upset medical services as far as we might be concerned. In this part, we dive into the absolute most thrilling improvements in the domain of biomedical and clinical imaging.

1. NANOTECHNOLOGY AND TARGETED THERAPIES

Nanotechnology has arisen as a unique advantage in the domain of biomedical imaging and treatment. Specialists have saddled the one of a kind properties of nanoparticles to make contrast specialists that give unmatched clearness in imaging, considering prior and more exact sickness discovery. These nanoparticles can be designed to target unhealthy cells, empowering exact medication conveyance and limiting inadvertent blow-back to sound tissues explicitly. The coordination of nanotechnology with clinical imaging methods holds the commitment of individualized medicines, where treatments are custom fitted to a patient's sub-atomic profile.

2. ARTIFICIAL INTELLIGENCE AND RADIOMICS

Man-made consciousness (computer based intelligence) has taken critical steps in breaking down complex clinical pictures. Radiomics, a field that extricates an immense measure of quantitative information from clinical pictures, is currently enabled by man-made intelligence calculations equipped for distinguishing inconspicuous examples and connections that are frequently indistinct to the natural eye. Computer based intelligence driven diagnostics improve the precision of sickness location as well as give prescient experiences into patient results. The capacity to quickly process and decipher enormous volumes of imaging information is upsetting dynamic cycles in clinical settings.

3. AUGMENTED REALITY AND SURGICAL NAVIGATION

The reconciliation of expanded reality (AR) into clinical imaging has changed surgeries. Specialists can now overlay ongoing imaging information onto a patient's life systems, furnishing them with improved representation and direction during complex tasks. This innovation empowers negligibly obtrusive medical procedures with phenomenal accuracy, lessening gambles and working on persistent results. Besides, AR-controlled careful preparation stages are engaging the up and coming age of clinical experts with vivid growth opportunities.

4. FUNCTIONAL IMAGING AND CONNECTOMICS

Customary clinical imaging generally centered on physical designs, yet ongoing patterns are moving towards useful imaging and connectomics. Practical imaging strategies, for example, utilitarian attractive reverberation imaging (fMRI) and positron discharge tomography (PET), give bits of knowledge into mind movement, brain availability, and metabolic cycles. These strategies are vital for understanding complex neurological issues and propelling our insight into mind capability. Connectomics, then again, includes planning the complicated brain networks inside the cerebrum, preparing for a more profound comprehension of cognizance and conduct.

5. QUANTUM IMAGING AND BEYOND

The appearance of quantum imaging innovations holds extraordinary commitment for upsetting clinical imaging's awareness and exactness. Quantum sensors and imaging gadgets offer unrivalled accuracy in catching signs, empowering the identification of weak signals and improving the goal of pictures. Quantum-improved imaging is ready to significantly affect early sickness recognition, empowering medical care experts to recognize oddities at their incipient stages.

6. ETHICAL AND REGULATORY CONSIDERATIONS

As these arising patterns reshape the scene of biomedical and clinical imaging, moral and administrative contemplations become foremost. Guaranteeing patient security, informed assent, and the mindful utilization of computer based intelligence and nanotechnology are basic viewpoints that require cautious consideration. Joint efforts between scientists, clinicians, ethicists, and policymakers are fundamental for work out some kind of harmony among development and protecting patient privileges.

CONCLUSION

All in all, the field of biomedical and clinical imaging is at the cusp of another period set apart by pivotal advances and creative exploration. Nanotechnology, simulated intelligence, increased reality, practical imaging, quantum advances, and moral contemplations are driving these changes. The intermingling of these patterns vows to change medical services, empowering prior infection identification, exact therapies, and worked on persistent results. As these developments keep on developing, interdisciplinary coordinated effort will assume an essential part in moulding the eventual fate of clinical imaging and its significant effect on worldwide medical care.

ID: 36

VISUALIZING COLLABORATION: UNVEILING THE EVOLVING LANDSCAPE OF IMAGE MULTI-AGENT SYSTEMS

Ms.A.RENUKA
Head of the Department,
Department of Computer Science,
Nethaji Memorial Arts & Science College,
Kerala, India.

ABSTRACT:

This Chapter investigates arising patterns in Picture Multi-Specialist Frameworks (MAS) and their extraordinary effect on cooperative direction. Key patterns including profound support learning, united learning, reasonable artificial intelligence, moral contemplations, and decentralization are analysed. These patterns engage specialists to decipher visual information, guarantee security, improve straightforwardness, address inclination, and advance moral use. The interchange of these patterns reshapes the scene of picture MAS, empowering applications across assorted spaces while cultivating mindful and even-handed innovation headway.

Keywords: *Multi-Agent Systems, Image Processing, Deep Reinforcement Learning, Federated Learning, Explainable AI.*

I. INTRODUCTION

The field of Multi-Agent Systems (MAS) has encountered a change in perspective with the coordination of picture based advances. As specialists become progressively equipped for seeing and handling visual information, additional opportunities have arisen for cooperative critical thinking and direction. This part presents an outline of arising patterns in picture MAS, revealing insight into their importance and suggestions.

1. DEEP REINFORCEMENT LEARNING FOR IMAGE MAS

Deep Reinforcement Learning (DRL) has acquired unmistakable quality as a strong method for preparing specialists to pursue successive choices. With regards to picture MAS, DRL empowers specialists to learn ideal activities by associating with their current circumstance and getting visual input. This pattern has prompted forward leaps in different areas, including independent vehicles, automated route, and computer game man-made intelligence.

One remarkable methodology is the reconciliation of convolutional Neural Networks (CNNs) with support learning calculations. CNNs succeed at removing significant highlights from pictures, empowering specialists to decipher complex obvious signals. Combined with support learning, these specialists can learn mind boggling undertakings like agreeable control of items or facilitated observation.

2. FEDERATED LEARNING AND DECENTRALIZATION

Federated Learning has arisen as a distinct advantage in the domain of security saving AI. In picture MAS, where various specialists team up while keeping up with information protection, combined learning permits specialists to advance all in all without sharing crude information. Every specialist prepares its model locally on its information and afterward shares model updates with a focal server, which totals information from all specialists. This pattern guarantees information protection, making it appropriate for applications like clinical imaging and observation.

Decentralization, an idea firmly connected with united learning, has likewise built up momentum. In this specific situation, every specialist has its own learning abilities, and cooperation happens through data trade as opposed to unified control. This approach advances heartiness and adaptation to non-critical failure, making picture MAS more versatile to dynamic conditions.

3. EXPLAINABLE AI IN IMAGE MAS

As Image MAS turns out to be more complex, the test of understanding specialist conduct and choices strengthens. Reasonable man-made intelligence (XAI) resolves this issue by giving straightforwardness into the dynamic cycles of computer based intelligence specialists. In picture MAS, XAI methods permit partners to appreciate how specialists decipher visual info and show up at explicit activities.

Interpretable profound learning designs, for example, consideration systems and element perception, assume an essential part in making specialist conduct justifiable. This pattern is especially urgent in security basic applications like clinical analysis and catastrophe reaction, where the reasoning behind specialist choices should be fathomable to human specialists.

4. ETHICAL CONSIDERATIONS AND BIAS MITIGATION

As Image MAS turns out to be profoundly incorporated into society, moral worries emerge. The information used to prepare and test these frameworks frequently reflect cultural predispositions, which can prompt out of line or biased results. Tending to predisposition in picture MAS is an advancing pattern that requests consideration.

Scientists are effectively creating strategies to alleviate predispositions in preparing information and calculations. This incorporates differentiating preparing datasets, reclassifying decency measurements, and integrating moral rules into the improvement cycle. Guaranteeing that picture MAS works morally and reasonably is urgent for its acknowledgment and joining into different areas.

CONCLUSION

The field of Image Multi-Agent Systems is quickly advancing, driven by forward leaps in profound support learning, combined learning, reasonable computer based intelligence, and moral contemplations. These patterns are forming the capacities and uses of picture MAS, from independent frameworks to clinical imaging and then some. As these patterns keep on creating, specialists and experts should stay watchful, guaranteeing that picture MAS benefits society while sticking to moral guidelines.

ID: 37

EXPLORING PARADIGM SHIFTS: EMERGING TRENDS IN PARALLEL COMPUTING

Mr.AKSHAY.S.LAL
M.Tech Student,
BITS Pilani,
Rajasthan, India.

ABSTRACT:
This Chapter investigates arising patterns in equal figuring, featuring how they are reshaping the scene of computational systems. Quantum parallelism, heterogeneous figuring, brain network parallelism, in-memory registering, and estimated processing are examined as key patterns that hold ground-breaking potential. These patterns are driving progressions in different spaces, from quantum processing's progressive critical thinking capacities to the advancement advantages of cloud-based parallelism. This theoretical gives a succinct outline of the section's items, offering bits of knowledge into the developing idea of equal registering.

KEYWORDS: *Quantum Parallelism, Heterogeneous Computing, Neural Network Parallelism, In-Memory Computing, Approximate Computing, Cloud-Based Parallelism, Computational Methodologies.*

I. INTRODUCTION

Parallel Computing has seen noteworthy development throughout the long term, and as innovation keeps on progressing, recent fads are arising that guarantee to reshape the scene of equal handling. This part dives into the absolute most conspicuous patterns in the field, investigating how they are changing the manner in which we approach calculation.

1. QUANTUM PARALLELISM

Quantum Parallelism has arisen as a progressive worldview, utilizing the standards of quantum mechanics to perform calculations in manners that traditional PCs can't. Quantum parallelism is at the center of this headway, permitting quantum bits or qubits to exist in a superposition of states and execute different calculations all the while. As specialists make progress in building more steady and mistake lenient quantum frameworks, quantum parallelism holds the possibility to take care of mind boggling issues that were once remembered to be unconquerable, like cryptography, advancement, and atomic recreations.

2. HETEROGENEOUS COMPUTING

In Present day, Registering structures progressively contain an assortment of handling units, including central processors, GPUs, FPGAs, and specific gas pedals. Heterogeneous figuring use these different assets to streamline execution for explicit errands. The test lies in actually disseminating and overseeing jobs across these units. Arising patterns center around creating programming models and runtime frameworks that work with consistent mix of heterogeneous parts, empowering productive use of equipment assets and accomplishing execution gains.

3. NEURAL NETWORK PARALLELISM

With the fast development of computerized reasoning and profound learning, brain network models have expanded and more perplexing. Preparing these models requests enormous computational power. Brain network parallelism investigates strategies to disseminate the preparation cycle across different gadgets or hubs, speeding up assembly and empowering preparing of much more mind boggling designs. Model parallelism, information parallelism, and half and half methodologies are being refined to advance the preparation of profound brain organizations, making computer based intelligence applications more open and productive.

4. IN-MEMORY COMPUTING

Customary figuring frameworks frequently experience the ill effects of memory and capacity bottlenecks that hinder execution. In-memory registering addresses this test by performing calculations straightforwardly inside memory, diminishing information development among memory and processors. With progressions in memory advancements like High Transmission capacity Memory (HBM) and determined memory, in-memory registering is getting some forward momentum. New calculations and information structures are being created to take advantage of this worldview, bringing about significant execution enhancements for information serious applications.

5. APPROXIMATE COMPUTING

Not all applications require exact, deterministic outcomes. Surmised processing use the possibility that compromising exactness for speed or productivity can yield huge increases. By permitting calculations to deliver inexact results that meet specific quality edges, designers can radically diminish computational prerequisites. This pattern is especially applicable in spaces like logical re-enactments, signal handling, and AI, where slight deviations from precise arrangements are satisfactory.

6. CLOUD-BASED PARALLELISM

Distributed computing has democratized admittance to strong registering assets, and parallelism is assuming an urgent part in improving cloud-based jobs. Arising patterns centre on empowering proficient scaling of uses in the cloud, auto-scaling in view of interest, and coordinating equal execution across disseminated virtual machines. Server less figuring, which abstracts foundation the board, is likewise incorporating parallelism to guarantee consistent execution while letting designers free from asset provisioning concerns.

7. BIOLOGICALLY - INSPIRED PARALLELISM

Drawing motivation from organic frameworks, analysts are investigating novel ways to deal with equal figuring. Brain organizations, hereditary calculations, and multitude insight are instances of organically enlivened standards that benefit from developing way of behaving and conveyed critical thinking. These methodologies hold guarantee for settling complex enhancement and example acknowledgment assignments, possibly prompting more energy-proficient and versatile processing frameworks.

CONCLUSION

All in all, the field of equal registering is going through a significant change with the development of these patterns. Quantum processing, heterogeneous designs, brain network parallelism, in-memory figuring, rough registering, cloud-based parallelism, and organically enlivened approaches all in all highlight a fate of remarkable computational capacities. As scientists and specialists keep on improving, the limits of what is attainable in equal registering are ceaselessly pushed, changing enterprises and molding the mechanical scene.

ID: 38

EXPLORING THE FRONTIERS OF ADAPTIVE ALGORITHMS: TRENDS, APPLICATIONS, AND CHALLENGES

Mr.AKSHAY.S.LAL
M.Tech Student,
BITS Pilani,
Rajasthan, India.

ABSTRACT:
This Chapter gives a compact outline of the arising patterns in versatile calculations, featuring their applications and suggestions for different ventures. Versatile calculations have developed to progressively change frameworks, applications, and gadgets in view of evolving information. This part investigates patterns like ongoing learning, interdisciplinary collaboration, united learning, moral variation, and difficulties looked by these calculations. As innovation propels, the fate of versatile calculations holds guarantee in fields like medication disclosure, environment displaying, and materials science.

KEYWORDS: *Adaptive Algorithms, Real-Time Learning, Interdisciplinary Synergy, Federated Learning, Ethical Adaptation.*

I. INTRODUCTION

In the present quickly developing mechanical scene, the field of versatile calculations has collected huge consideration. These calculations assume a crucial part in empowering frameworks, applications, and gadgets to gain from information and powerfully change their way of behaving to evolving conditions. This section digs into the arising patterns in versatile calculations, investigating their applications, challenges, and the potential they hold for molding the eventual fate of different ventures.

1. ADAPTIVE ALGORITHMS: AN OVERVIEW

Versatile calculations, established in the domain of AI and man-made reasoning, engage frameworks to adjust and upgrade their exhibition in light of accessible information. These calculations are based upon the underpinning of factual models, gaining from information to make forecasts, choices, or even control activities. Throughout the long term, versatile calculations have advanced from straightforward rule-based components to complex, information driven frameworks fit for handling complicated true issues.

2. REAL-TIME LEARNING AND DECISION-MAKING

Perhaps of the most unmistakable pattern in versatile calculations is the move towards constant learning and navigation. Customary calculations frequently require broad disconnected preparing, restricting their relevance in powerful conditions. Be that as it may, present day versatile calculations influence strategies like web based learning and support figuring out how to adjust continuously. This pattern 2has changed enterprises like money, where algorithmic exchanging frameworks go with split-subsequent options in view of market variances.

3. INTERDISCIPLINARY SYNERGY

Another arising pattern is the assembly of versatile calculations with different disciplines. By consolidating bits of knowledge from fields like neuroscience, mental science, and transformative science, specialists are creating calculations that emulate the flexibility and learning components tracked down in normal frameworks. This interdisciplinary methodology has prompted leap forwards in regions like multitude knowledge, where calculations imitate the aggregate way of behaving of social creatures like subterranean insects or honey bees.

4. FEDERATED LEARNING AND PRIVACY

As information protection concerns heighten, unified learning has arisen as a promising arrangement. This approach permits models to be prepared cooperatively across decentralized gadgets while keeping delicate information confined. Versatile calculations utilized in united learning adjust to assorted information disseminations across gadgets, giving customized encounters without compromising protection. This pattern has built up forward movement in applications going from medical care to customized proposal frameworks.

5. ETHICAL AND FAIR ADAPTATION

With the rising impact of calculations in dynamic cycles, guaranteeing moral and fair transformation has become critical. Algorithmic predispositions and unseen side-effects can emerge while adjusting to information, possibly sustaining disparities. As a reaction, the pattern of integrating decency mindful and moral contemplations into versatile calculations has acquired noticeable quality. Analysts are creating calculations that adjust in view of information as well as focus on reasonableness and value, adding to additional equitable results.

6. CHALLENGES AND FUTURE DIRECTIONS

While versatile calculations hold monstrous potential, they additionally face difficulties. The scourge of dimensionality, information shortage, and ill-disposed assaults are among the obstacles that specialists endeavour to survive. Moreover, finding some kind of harmony between quick transformation and solidness stays a test, especially in security basic applications.

Looking forward, the fate of versatile calculations is splendid. As quantum figuring and neuromorphic processing innovations advance, versatile calculations are ready to investigate new wildernesses. The capacity to adjust and learn in profoundly mind boggling and nonlinear conditions could reform ventures like medication revelation, environment displaying, and materials science.

CONCLUSION

The development of versatile calculations reflects the speed of innovative headways. These calculations have advanced from simple rule-based frameworks to keen, information driven leaders. As they keep on forming enterprises and impact day to day existence, their capability to drive advancement stays tremendous. With interdisciplinary joint efforts, moral contemplations, and headways in registering innovations, versatile calculations are set to lead us into a future where clever variation is at the core of progress.

ID: 39

ADVANCING IMAGE RETRIEVAL AND VISUAL SEARCH: EMERGING TRENDS AND APPLICATIONS

Mr.S.SANATHKUMAR
Research Scholar,
Rathinam College of Arts and Science,
Coimbatore, Tamil Nadu, India.

ABSTRACT:
This Chapter dives into the powerful domain of picture recovery and visual inquiry, investigating the most recent patterns that are reshaping the field. Utilizing state of the art innovations like profound learning and brain structures, cross-modular recovery, self-regulated learning, hardly any shot and zero-shot recovery, reasonable visual pursuit, and security safeguarding recovery, the part gives an exhaustive outline of the developing scene. The patterns talked about find applications in areas like web based business, medical services, and social legacy safeguarding. The section underlines the ground-breaking effect of these patterns on how we draw in with visual information and presents a forward-looking viewpoint on the field's direction.

KEYWORDS: *Image Retrieval, Visual Search, Deep Learning, Cross-Modal Retrieval, Privacy-Preserving Retrieval.*

I. INTRODUCTION

In Modern Days, the field of picture recovery and visual pursuit has seen quick progressions driven by leap forwards in PC vision, AI, and information examination. These advances have reformed the manner in which we associate with visual information, empowering applications going from web based business item search to content association in computerized libraries. This part investigates the absolute most encouraging arising patterns in this unique field.

1. DEEP LEARNING AND NEURAL ARCHITECTURES

Deep learning has pushed the abilities of picture recovery and visual inquiry by empowering the advancement of complicated brain designs fit for separating complex elements from pictures. Convolutional Brain Organizations and Transformer-based models certainly stand out for their capacity to learn rich portrayals of pictures, upgrading the exactness and proficiency of recovery frameworks. Start to finish learning approaches have additionally arisen, permitting organizations to upgrade for explicit recovery errands straightforwardly.

2. CROSS-MODAL RETRIEVAL

Cross-modular recovery has arisen as an entrancing region, empowering the quest for pictures utilizing various modalities like text or sound. This pattern has tracked down applications in different areas, for example, looking for pictures in light of regular language depictions or recovering music tracks in view of pictures. High level models that span the semantic hole between various modalities are being created, working with more natural and far reaching search encounters.

3. SELF-SUPERVISED LEARNING

Naming tremendous measures of information for preparing can be costly and tedious. Self-regulated learning has gotten momentum as an approach to pre-train models on huge datasets without the requirement for express marks. These models gain from intrinsic examples in the information, which can then be tweaked for explicit recovery errands. This approach has shown guarantee in working on the presentation of picture recovery frameworks while limiting the human exertion expected for manual explanation.

4. FEW-SHOT AND ZERO-SHOT RETRIEVAL

Few shot and zero-shot recovery methods address the test of restricted marked information. Hardly any shot recovery includes preparing models to recover pictures with just few models for every class, while zero-shot recovery stretches out this plan to classes that were absent in the preparation information by any means. These strategies are critical for situations where getting broad marked information is unrealistic, for example, looking for uncommon or recently arising ideas.

5. EXPLAINABLE VISUAL SEARCH

As Image recovery frameworks become more intricate, the requirement for straightforwardness and interpretability emerges. Reasonable computer based intelligence methods intend to give bits of knowledge into how recovery choices are made, providing clients with a more profound comprehension of the indexed lists. Strategies, for example, consideration perception and saliency maps assist clients with grasping the reason why certain pictures were recovered, upgrading trust and ease of use.

6. PRIVACY-PRESERVING RETRIEVAL

With developing worries about information security, specialists are investigating strategies to direct picture recovery without compromising client protection. Strategies like united learning permit models to be prepared across various gadgets without sharing crude information, while cryptographic conventions empower secure inquiries over scrambled pictures. Security safeguarding recovery guarantees that delicate data stays safeguarded while as yet empowering powerful pursuit capacities.

7. REAL-WORLD APPLICATIONS

The arising patterns in picture recovery and visual pursuit are tracking down applications in different certifiable situations. Online business stages are utilizing visual pursuit to upgrade client experience by permitting clients to find items utilizing pictures. Medical care experts can profit from picture recovery frameworks that guide in diagnosing ailments. Social legacy foundations are utilizing these patterns to coordinate and look through tremendous files of visual substance, saving and sharing information.

CONCLUSION

In Conclusion, the scene of picture recovery and visual pursuit is developing quickly because of headways in profound learning, cross-modular recovery, self-managed learning, not many shot recovery, reasonableness, protection conservation, and their horde applications. As these patterns keep on developing, they will shape the manner in which we communicate with visual information, preparing for more clever and effective recovery frameworks that take care of the assorted necessities of clients across various spaces.

ID: 40

A NOVEL METHOD IN CLOUD COMPUTATION STORAGE BY USING FIRMNESS AND CRYPTO MECHANISM

Mr.P.MUTHUKUMAR
Ph.D Research Scholar,
Department of Computer Science,
Rathinam College of Arts and Science,
Coimbatore, Tamil Nadu, India.

Dr.T.VELUMANI
Assistant Professor,
Department of Computer Science,
Rathinam College of Arts and Science (Autonomous),
Coimbatore, Tamil Nadu, India.

ABSTRACT:
In order to communicate the compressed secure multimedia data over the internet by using the mechanism of compression and crypto algorithm is a modern and innovative idea in cloud computing. The cloud computing is an inevitable field of data communication and resource sharing in the modern era. The essence of its functioning is the boundaries, the development of new applications, becoming increasingly agile and collaborative, inspiring subjects for research. In the new century, it appears that the ability of data centers is limited and runs out. The economy in sequences the trend of technological development and the solution is the adoption of grid services and/or utility computing as well as the use of virtualization to maximize the available resources. Existing services and applications become more distributed, paradigms like Service-Oriented Architecture emerge in response to integration and service orchestration, and the organization and technologies used in data centers evolve. In this research paper is mainly focus on the role of compression and crypto algorithm in cloud data storage mechanism.

KEYWORDS: *Cloud, Compression, Storage, Secure and Data.*

I. INTRODUCTION

In general the cloud applications are categorized based On-demand self-service - A consumer can unilaterally provide computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider. Broad network access and the capabilities are available over the network and accessed through standard mechanisms that promote the use by heterogeneous thin or thick client platforms (e.g., mobile phones, tablets, laptops, and workstations).

Resource is pooling and the providers computing resources are combined to serve multiple consumers using a multi-client model, with different physical and virtual resources, dynamically assigned and reassigned according to consumer demand.

There is a sense of location independence since, the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify the location at a higher level of abstraction (e.g., country, state, or data center). Examples of resources include storage, processing, memory, and network bandwidth. Rapid elasticity - Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time.

Measured service and the cloud systems automatically control and optimize resource use, by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g. storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the used service.

Software as a Service (SaaS) - The capability provided to the consumer, ability to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through either a thin client interface, such as a web browser (e.g., web-based email), or a program interface. The consumer does not manage or control the underlying cloud infrastructure, including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user specific application configuration settings.

Platform as a Service (PaaS) - The capability provided to the consumer is to deploy onto the cloud infrastructure, consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly configuration settings for the application-hosting environment.

Infrastructure as a Service (IaaS) - The capability provided to the consumer is the provision of processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications, and possibly limited control of select networking components (e.g., host firewalls). From the figure 1.1 illustrates the principles behind the proposed approach,

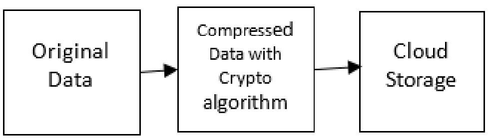


Figure 1.1 Layout for Cloud data storage modern approach

II. RELATED WORK

Data compression technique is used for saving disk space, reducing the time needed for communication or the time needed for data transfer and more. Data to be handled as well as software has been growing, and the amount of information communicated between systems has also been constantly increasing. In these circumstances, data compression technology is regarded as an important factor to support the information infrastructure. Data compression technology is not only used in PCs, but also in many fields like modems, routers, digital cameras, facsimiles, CDs, MDs(Mini disk), video on demand(VOD), TV conference system, DVDs, Digital telephones and other fields yet the data compression technique generally refer to data stored in 'files' and file transfer over phone lines.

In MS-DOS, used programs like ARC and PKZIP and ubiquitous. UNIX used has the COMPRESS and COMPACT utilities and WINDOWS used have WINZIP utilities. Types Compression: There are two types of data compression, "lossless" data compression and the "lossy" compression. Lossless data compression is used when the data has to be uncompressed exactly as it was before compression. This type of compression used when storing database records, spreadsheets and word processing files. Text files are stored using lossless techniques, since losing a single character can in the worst case, make the text dangerously misleading. Archival storage of master sources for images, videos data and audio data generally needs to be lossless as well. However there are strict limits to the amount of compression that can be obtained with lossless compression. Lossless compression ratio is generally in the range of 2:1 to 8:1.

Lossy compression, in contrast, works on the assumption, that the data does not have to be stored perfectly. Much information can be simply through away from images, video data and audio data, and when decompressed, the data will still be of acceptable quality. Lossy data compression concedes a certain loss of accuracy in exchange for greatly increased compression. Lossy methods can be used on speech and graphics, and they are capable of acting dramatically higher compression ratios. The question of which is a 'better', lossless or lossy technique is pointless. Each has its own uses, with lossless techniques better in some cases and lossy techniques better in others. In fact, lossless and lossy technique are used together to obtain the higher compression ratio.

III. PROPOSED WORK

In general, cloud data compression /data compression consists of talking a stream of symbols and transforming them into codes. If the compression is effective, the resulting stream of codes will be smaller than the original symbols. The decision is output a certain code for a symbol or set of symbols is based on a model. So, data compression consists of two components, modeling and coding. According to the recent applications of cloud computing, it is the next big thing.

Cloud computing promises reduced cost, flexibility; improve the time to market, higher availability, and more focus on the core business of an organization. Virtually all players of the IT industry are jumping on the cloud computing. The issue that seems to be able to stop cloud computing are security concerns. One of the most prominent issues for cloud computing is privacy; that is, the protection of sensitive data from unauthorized users. Privacy is defined is the way to secure data protection and the control is outsourced to the cloud provider.Coding is frequently used to refer to the entire data compression process instead of just a single component of that process. For example, the phrase "Huffman coding" describes a data compression technique, which actually mean the coding method used in conjunction with a model to compress data. In data compression metaphor, coding would be the wheels, but modeling would be the engine. Information theory uses the term entropy as a measure of how much information is encoded in a message. The higher the entropy of a message, the more information it contain. The entropy of a symbol is defined as the negative logarithm of its probability. To determine the information content of a message in bits, the base 2 logarithm as follows:

Number of bits = - Log base 2 (probability)

The entropy of an entire message is simply the sum of the entropy of all individual symbols. For example, if the probability of the character 'e' is 1/16, the information content of the character is four bits. So the character string 'eeeee' has a total content of 20 bits. If this string is encoded in the standard 8-bit ASCII, 40 bits are required to encode this message. Encoding characters using EBCDIC or ASCII require every character to be encoded using the same number of bits. To compress data, we need to encode symbols with exactly the number of bits of information the symbol contains. If the character 'e' only gives us four bits of information, then it should be coded with twelve bits. But this will introduce lots of error, with most of the codes in a message being too long and some being too short. This coding problem can be solved by Shannon-Fano coding and Huffman coding two different ways of generating variable length codes when gives a probability table for a given set of symbols.

The Encryption is a possible way to fulfill the confidentiality requirement of data that is being stored in an untrusted cloud database. The request data are used to retrieve data from a cloud database. After the completion of the framework, we proceeded to tests in order to validate experimentally the suitable operation of the framework and to evaluate different parameters associated with the compression, decompression, encryption and decryption of multimedia contents. As may be seen in this table, the obtained results show that compression followed by encryption may be effective for efficient and secure storage of multimedia content. On the other hand, if we apply first encryption followed by compression, the compression rate is very small due to the higher degree of randomness introduced by encryption. These results confirm the unsuitability of encryption followed by compression for efficient multimedia storage. The properties of an encryption function determine which kinds of method can be applied on the encrypted data without decrypting them and, consequently, how it will be stored in to the cloud service provider. The interface for the cloud data access is illustrated by the figure 1.2.

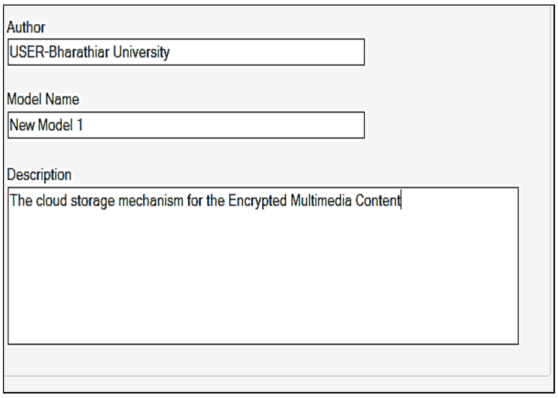


Figure 1.2 Interfaces

The data for the transmission is used for the encryption by using any one of the crypto algorithm. A deterministic encryption scheme (as opposed to a probabilistic encryption scheme) is a cryptosystem which always produces the same ciphertext for a given plaintext and key. The property of indistinguishability under chosen plaintext attack is considered a basic requirement for most provably secure public key crypto systems. The Cloud Encryption Component is shown by the following figure 1.3.

The complexity in virtual machine environments can also be more challenging than their traditional counterparts, giving rise to conditions that undermine security. For example, paging, check pointing, and migration of virtual machines can leak sensitive data to persistent storage, subverting protection mechanisms in the hosted operating system intended to prevent such occurrences. Moreover, the hypervisor itself can potentially be compromised. For instance, a vulnerability that allowed specially crafted File Transfer Protocol (FTP) requests to corrupt a heap buffer in the hypervisor, which could allow the execution of arbitrary code at the host, was discovered in a widely used virtualization software product, in a routine for Network Address Translation (NAT).

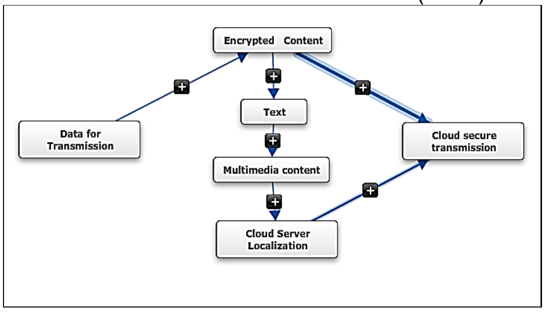


Figure 1.3 Cloud Encryption Components

Most virtualization platforms have the ability to create software-based switches and network configurations as part of the virtual environment to allow virtual machines on the same host to communicate more directly and efficiently. For example, for virtual machines requiring no external network access, the virtual networking architectures of most virtualization software products support same-host networking, in which a private subnet is created for intra-host communications. Traffic over virtual networks may not be visible to security protection devices on the physical network, such as network-based intrusion detection and prevention systems

IV. CONCLUSION

Determining the security of complex computer systems composed together is also a long-standing security issue that plagues large-scale computing in general, and cloud computing in particular. Attaining high-assurance qualities in implementations has been an elusive goal of computer security researchers and practitioners and, as demonstrated in the examples given in this report, is also a work in progress for cloud computing. Nevertheless, public cloud computing is a compelling computing paradigm that agencies need to incorporate as part their information technology solution set.

V. REFERENCES

1. Yared Keleta, J.H.P. Eloff, H.S. Venter, Proposing a Secure XACML Architecture Ensuring Privacy and Trust, Research in Progress Paper, University of Pretoria, 2005.
2. Brian Krebs, Amazon: Hey Spammers, Get Off My Cloud! The Washington Post, July 1, 2008.
3. Eileen Kowalski et al., Insider Threat Study: Illicit Cyber Activity in the Government Sector, U.S. Secret Service and Carnegie Mellon University, Software Engineering Institute, January 2008.
4. Michael Krigsma, Amazon S3 Web Services Down. Bad, Bad News for Customers, ZDNET, February 15, 2008.
5. Sushil Kumar, Oracle Database Backup in the Cloud, White Paper, Oracle Corporation, September 2008. [Lab95] Stephen Labaton, 2 Men Held in Attempt to Bomb I.R.S. Office, New York Times, December 29, 1995.
6. 20-Year Term in Plot to Bomb IRS Offices, Nation In Brief, Los Angeles Times, August 10, 1996.
7. Neal Leavitt. Is Cloud Computing Really Ready for Prime Time?, IEEE Computer, January 2009.
8. James Maguire, How Cloud Computing Security Resembles the Financial Meltdown, Datamation, internet.com, April 27, 2010

ID: 41

SECURED ONLINE PAYMENTS WITH DISTRIBUTED LEDGER USING BLOCKCHAIN

Ms.S.SABEENA
Assistant Professor,
Department of Computer Science,
Sri Krishna Arts and Science College,
Coimbatore, Tamil Nadu, India.

Ms.C.SATHVIGA
Student,
Department of Computer Science,
Sri Krishna Arts and Science College,
Coimbatore, Tamil Nadu, India.

ABSTRACT:
Each entry in the growing list of records on the blockchain is linked together via encryption. A timestamp, transaction details, and a cryptographic hash of the preceding block are all included in each block of the chain. With this secure one, we want to substitute the present system for online payment. A modern internet payment gateway is vulnerable to attackers who can tamper with the network and cause financial loss. Additionally, the transaction must pass through several payment systems, which takes time and raises chances of a failed transaction. Our solution would employ blockchain, which permits online transactions, to enable online payments to be delivered straight from a party to a different without involving a financial institution in a safe manner. This kind of technology enables two parties to conduct online transactions without depending on or having faith in a third party using cryptographic evidence. We utilize a proof-of-work mechanism to keep track of transactions, making it very difficult for an attacker to alter them. Digital signatures contribute to the solution in guaranteeing the confidentiality and integrity of the data recorded on a blockchain.

I. INTRODUCTION

A blockchain is a network that stands on its own on the web and holds an accessible archive of the information gathered from that network. The revolutionary potential of blockchain is due to the way this data is recorded. The term "blockchain" describes a network of interlinked blocks. Both the terms "chain" (data stored in an open database) and "block" (digital info units). Every single block in the network that's carrying the data is secured and connected to one another with the aid of cryptography principles. Once recorded in a block, data cannot be edited or updated, making it impossible to do so without the other network users seeing it. A computer, cellphone, printers, or possibly a refrigerator that is connected to the internet is capable of serving as a node. On a blockchain, every node is equally important, yet various nodes might play different roles in how a blockchain functions. The blockchain network is fundamentally distinct from conventional networks since there is no one location where data is managed and stored. Instead, many networked devices—known as peers—share responsibility for upkeep and data storage. A network of blockchain connections is referred to as a network of peers as a result. Instead, all of the network's participants are constantly storing and exchanging information.

Blockchain relies on cryptography to secure all data and value storages and ensure that transactions get carried out safely. As a result, everyone who utilizes blockchain may be confident that all data saved on it has done so carefully and lawfully. The primary component of data recorded in blockchain is the digital signature, which protects the data's security. The Genesis block, the initial block in the blockchain, comprises all of the transactions that later aggregate and are verified to generate a distinct hash. The subsequent block in the chain uses this hash as input to generate a distinctive hash using all of the newly executed transactions. As a result, each block is guaranteed to have a hash connection to the one before it, building a chain back to the genesis block and giving rise to the name blockchain. The market for online transactions has grown significantly, therefore it is crucial to consider both their advantages and disadvantages. Presently, payment processing systems use several types of outside systems, which makes it time-consuming as the payment process must frequently transit through multiple parties and runs a chance of failure. The second major issue is protection, and the current method falls short of what the client expects. This is because there have been numerous occasions where an attacker has undermined the safety of a transaction, leading to financial losses and a decrease in client trust. Then there are more transactional fees, which should be considered from the perspective of the client and can be decreased using blockchain. A decentralized application can be developed using blockchain technology to improve financial administration, which is still another necessity. Blockchain facilitates transaction maintenance generally and accelerates processes where present technology falls short, that is, where it is far slower than blockchain. Therefore, we intend to implement blockchain for payments made online. In the blockchain, each node, or customer-side transaction, would be documented. Ethereum, Bitcoin, and other public blockchains are currently accessible. Although Ethereum provides a foundation for us to build our personal smart contracts, as it's an open network, all users can see the details of transactions instead of having access to them only via a private sub-network.

II. LITERATURE SURVEY

Nakamoto, Satoshi. The entire blockchain-based mechanism for an electronic cash system is provided in this paper, which effectively makes it possible for payments made online to be transmitted straight across participants without being processed by a banking institution. This provides an explanation of a distributed network system, or among peers system, which encouraged the creation of the Proof of Work method and a fix for double spending. From a technological standpoint, Judmayer, Aljosha et al. offered an introduction to blockchain technologies as well as the concepts of distributed ledgers and crypto currency. This article argues that science is currently responding remarkably slowly to this novel and rapidly evolving area of blockchain technology because there aren't enough resources accessible outside of bitcoin. This article primarily focuses on Bitcoin's crypto currency. It went into great detail regarding bitcoin as well as the reasons it has sparked interest in contemporary technology and such a sizable industry. It also highlighted the difficulties of handling online assets as well as bitcoin's ease of use, secrecy, and safety issues through an individual's standpoint as well as the idea, traits, and necessity of blockchain technology, and how bitcoin functions. It makes an effort to demonstrate how Blockchain will influence banking and financial organizations in the future. First, Zibin Zheng et al. offered a summary of the blockchain architecture before contrasting a few common consensus methods applied to various blockchains. Also covered were different blockchain-based applications that cover a wide range of industries, including financial services, reputation management, IOT, and more. The technical difficulties with blockchain technology, including its scalability, security issues that need to be solved, recent advancements, and potential future trends, are also briefly discussed.

III. ISSUES IN REGULAR BANKING SYSTEM

The Indian financial system has greatly deepened and expanded in the last few years. Numerous pressures have been put on our banks as a result of the growing importance of the Indian economy's banking sector, more deregulation, and increased competition. A failure of the banking system could have greater detrimental effects than in the past. As a result, the RBI, which oversees and regulates the Indian banking sector, is putting more of an emphasis on ensuring more stable finances. The banking industry is susceptible to a variety of dangers and difficulties when working in this extremely demanding environment, some of which are covered here:

IV. IMPROVING RISK MANAGEMENT SYSTEM

The RBI delivered suggestions on credit-related risk management, market threat management, and operational risk management in October 2002 and 2005, respectively, after publishing guidelines on asset liability management and systems for managing risks in banks in 1999. Since

Basel II places a lot of emphasis on risks, implementing it shouldn't be seen as a goal in and of itself. The contemporary corporate climate calls for an integrated risk management strategy. A unique strategy for every threat is no more adequate. Indian banks are making the switch from a segmented system to an all-encompassing Risk Management System. This has raised the requirement for capacity building and increased demand for risk management capabilities in banks. The major objective is to integrate risk across the entire bank, but it will also be important to examine if risk aggregation across the Group is desirable. Banks would need to invest a lot of money in this endeavour during the years that follow.

1. TECHNOLOGICAL PROBLEMS

It's true that Indian banks have started using computers for operations, and there have been numerous other technological breakthroughs, but these are insufficient. However, the full network of branches in other towns and villages cannot support and be equivalent to the good similar technology used by Indian local banks in metro areas.

2. TRANSPARENCY AND DISCLOSURES

The RBI took a variety of steps to ensure that banks' true financial status was meaningfully disclosed, enabling users of financial statements to assess and contrast their positions. It included a wide range of subjects, including related party disclosures, asset quality, profitability, segment reporting, nation risk exposure, and exposure to derivative risk. The RBI has proposed better exposures of several quality factors in order to comply with Basel II Pillar 3's disclosure obligations as well as global best practices and global accounting standards. Banks must have a comprehensive disclosing strategy covering internal controls over the disclosure process as well as the bank's approach to deciding what disclosures to make.

3. LACK OF ACCESSIBILITY

Rural and underprivileged regions frequently have limited access to traditional banking services because traditional banks are frequently concentrated in urban areas. People are unable to access necessary financial services due to this lack of accessibility, which limits financial inclusion.

4. MIDDLEMEN AND INTERMEDIARIES

Numerous intermediates, including correspondent banks, clearinghouses, and payment processors, are frequently used in traditional banking procedures. Each intermediary increases transaction complexity and expense.

V. BENEFITS

Through the use of technology, blockchain payments eliminate the need for banks and are therefore free of bank fees.

The payments are secure because they can be tracked from start to finish using blockchain technology, which also protects user privacy by notifying only network members of transactions while hiding the identities of those who carried them out.

1. ENHANCED SECURITY

Your data is confidential and crucial, so the blockchain system has the power to drastically alter the way others see your vital details. Blockchain eliminates fraud and unlawful conduct by generating a record which cannot be altered that is entirety encrypted. By leveraging permissions to limit access and anonymizing user data, blockchain can address privacy concerns. Attackers have trouble obtaining data since it is kept over a group of machines instead of on one server.

2. GREATER TRANSPARENCY

Prior to blockchain, each company would need to manage its own database. Blockchain involves usage of a distributed ledger, which enables the synchronization of exact transaction as well as data records over a number of sites. The concurrent viewing of identical information by every user with the relevant rights ensures the network's complete transparency. A date and time label and a permanent record are present for every transaction. Users can view the entire history of the transaction, effectively removing the danger of cheating.

3. INSTANT TRACEABILITY

Blockchain establishes a verification track that documents an asset's origins at each stage of its travel. This helps by offering evidence in industries where clients worry about how a product will affect the environment or human rights, or in industries where scam and counterfeit are rife. With relation to the source, blockchain enables straight exchange of information with customers. Any supply chain can experience issues, such as when products are kept on a dock for loading while being shipped, which can be highlighted by traceability data.

4. DECENTRALIZATION

Blockchain decentralization distributes power throughout a network of nodes, promoting security, resiliency, and transparency by removing single points of failure. It improves data integrity and lowers the possibility of illegal changes, strengthening the system's defences against intrusions. As transactions are clearly recorded and verified by the entire network rather than depending on a single authority, this structure also promotes confidence among participants.

5. SMART CONTRACTS

Smart contracts are autonomous agreements that run on the blockchain and automate tasks according to established criteria. These programmable contracts do away with middlemen, increasing efficiency and lowering the possibility of human error. When conditions are met, they operate autonomously and transparently carry out actions, enabling tamper-proof and trustless execution. Smart contracts transform industries by offering transparent, automated, and secure solutions that expedite procedures and eliminate dependence on centralized institutions. Their applications range from financial transactions to supply chain management.

VI. EXISTING SYSTEM

The current traditional banking system has a centralized structure where established entities, particularly banks and financial intermediaries, manage and support financial transactions and services. Customers in this system depend on banks to handle transactions, maintain their accounts, and offer a variety of financial services. When intermediaries like clearinghouses, payment processors, and correspondent banks are involved in a transaction, there are frequently delays, costs, and complications.

This centralized approach, though, has drawbacks. Due to the participation of intermediaries and different time zones, transactions can be delayed, particularly for cross-border payments. Customers frequently have little knowledge of transaction processes, which can cause disagreements and confusion among users when there is a lack of openness. Furthermore, this system is vulnerable to data intrusions and security breaches, which compromise critical consumer information. Users must also contend with accessibility issues because typical banking services may not be available in distant or underdeveloped areas, preventing some sections of the population from participating in the financial system. High transaction fees, strict working hours, and high transaction expenses all add to the annoyance.

These shortcomings have increased interest in blockchain technology and it's potential to completely alter the way that finance operates today. Many of the flaws in the present system are addressed by the decentralized, open, and automated nature of blockchain, which offers improved security, effectiveness, and inclusivity in financial transactions.

VII. PROPOSED OUTCOME

Blockchain is something we want to use for payments made online. Each node, or each transaction on the customer's end, would be recorded on the blockchain. There are currently public blockchains for Ethereum, Bitcoin, and other cryptocurrencies. Because Ethereum is an open network, anybody can see the details of a transaction as opposed to being limited to a subnetwork in which just the appropriate parties may see their transactions. We may develop our personal smart contracts using the Ethereum platform as a foundation. Before being stored in the database as a chain, all transactions would be verified by miners. A transaction attempt is originally shared with the miners each time a user makes one. Before being added to the chain, the transaction will be confirmed by a specific amount of miners by proof of work, and their findings will be shared.

VIII. ALGORITHM

The following phases are commonly included in the blockchain transaction algorithm:

- Creating a Transaction
- Digital Signature
- Verification
- Mining
- Confirmation
- Updating the Ledger

These features guarantee the security, openness, and irrevocability of blockchain transactions.

Blockchain technology relies heavily on consensus algorithms since they enable decentralized networks to concur on the ledger's current state. Blockchain utilizes a variety of consensus techniques, including:

1. Proof of Work (PoW): The Bitcoin network employs this method of consensus. The first miner to finish a difficult mathematical puzzle adds a block to the blockchain and is rewarded with newly created bitcoin. Miners compete to solve the puzzle.
2. Proof of Stake (PoS): Used by several blockchain networks, notably Ethereum, A substitute to Proof of Work (PoW) is Proof of Stake. In a PoS system, validators are selected to add blocks to the blockchain depending on how much bitcoin they own.
3. Delegated Proof of Stake (DPoS): Networks like EOS and BitShares use DPoS as a form of PoS. In a DPoS system, token holders elect the delegates in charge of creating new blocks for the blockchain.
4. Byzantine Fault Tolerance (BFT): As a consensus technique, BFT is used by some blockchain networks, such as Hyperledger Fabric. In a BFT system, a predetermined number of nodes must concur on the ledger's status for a transaction to be regarded as legitimate.
5. Proof of Authority (PoA): Based on their identification and reputation, validators are permitted to contribute blocks to some private and consortium blockchains using the Proof of Authority (PoA) consensus process.

Every consensus method has advantages and disadvantages, and different blockchain networks use various algorithms depending on their particular needs. The objective of a consensus algorithm is to guarantee decentralized, secure, and transparent network consensus on the ledger's current state.

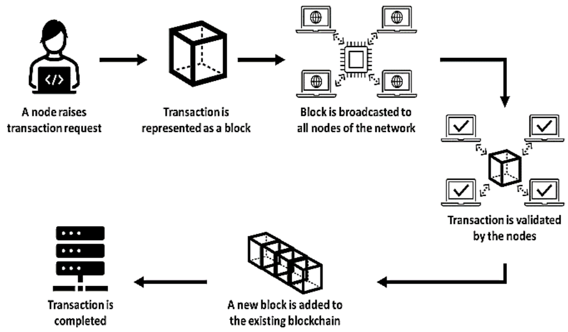


Fig.1 How Blockchain Works

IX. RESULT

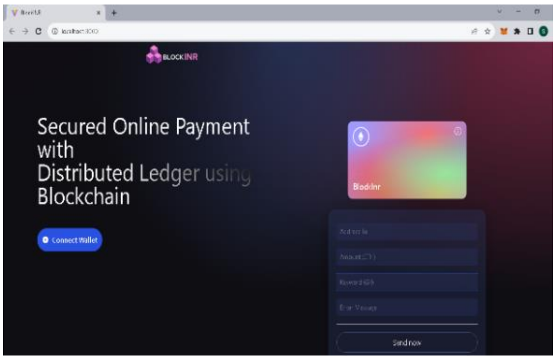


Fig.2 HOME PAGE

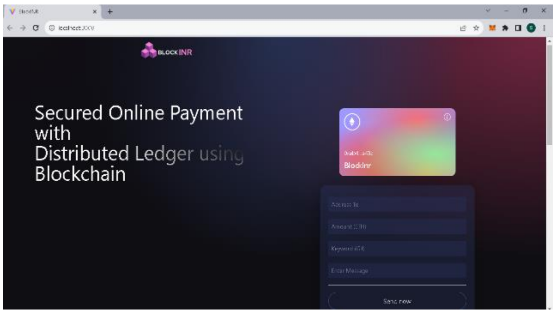


Fig. 3 WALLET CONNECTION

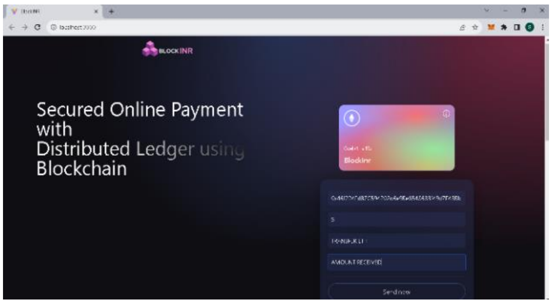


Fig. 4 TRANSACTION DETAILS UPDATION

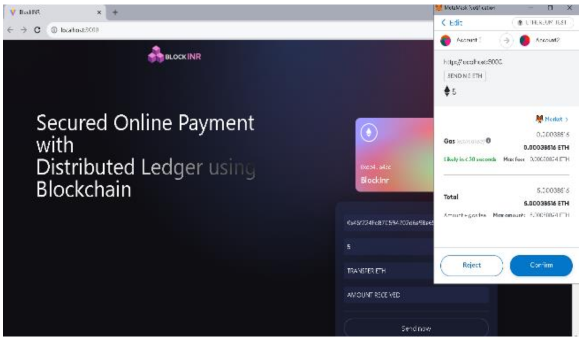


Fig. 5 WALLET TRANSFER USING METAMASK

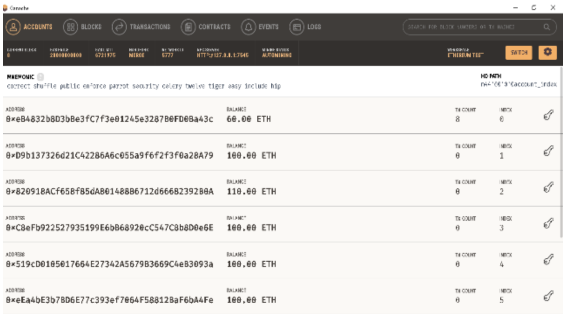


Fig. 6 GANACHE TEST ENVIRONMENT

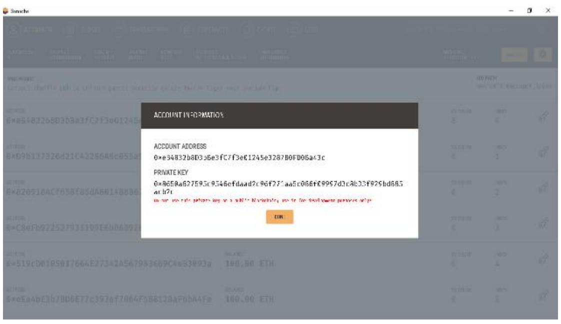


Fig. 7 PRIVATE KEY

X. CONCLUSION

Therefore, using blockchain to implement online payment transactions has been a success. This secures the entire procedure. A one-way hashing algorithm is used to transfer data to miners in a secure manner. The transactions are then verified by the miners using a proof-of-work technique utilizing the received hash value. Once the transaction have been placed in the blockchain and have been verified, they cannot be changed after that point. Thus, the application's goal is to provide an online transaction procedure that is secure by preventing assaults like man-in-the-middle attacks and by doing away with third-party gateways that makes the entire online money transfer process quicker. By automating agreements, smart contracts reduce the need for human participation and the likelihood of mistakes and disagreements. Faster cross-border transactions, lower prices, and greater accessibility are all potential benefits of this ground-breaking strategy, especially for underprivileged communities. Although there are still issues with scalability, regulatory alignment, and user adoption, it is clear that blockchain-based technology has the capacity to revolutionize the world of online payments. Collaboration between inventors, regulators, and stakeholders will be essential as we negotiate this technical transition to fully realize the advantages that blockchain technology brings to the world of secure and efficient online payments.

XI. REFERENCES

1. What is Blockchain Technology? A Step-by-Step Guide for Beginners. Available Online: https://blockgeeks.com/guides/ is-blockchaintechnology/
2. Nakamoto, Satoshi. ”Bitcoin: A peer-to-peer electronic cash system.” (2008). Available Online: https://bitcoin.org/bitcoi
3. Judmayer, Aljosha, Nicholas Stifter, Katharina Krombholz, and Edgar Weippl. ”Blocks and chains: introduction to bitcoin, cryptocurrencies, and their consensus mechanisms.” Synthesis Lectures on Information Security, Privacy, & Trust 9, no. 1 (2017): 1-123.
4. Zheng, Zibin, Shaoan Xie, Hongning Dai, Xiangping Chen, and Huaimin Wang. “An overview of blockchain technology: Architecture, consensus, and future trends.” In 2017 IEEE International Congress on Big Data (BigData Congress), pp. 557-564. IEEE, 2017.
5. Kishigami J, Fujimura S, Watanabe H, Nakadaira A, Akutsu A. The Blockchain-Based Digital Content Distribution System. In: Big Data and Cloud Computing (BDCloud), 2015 IEEE Fifth International Conference on; 2015. [6] Double-spending; 2016. Accessed: 24/3/2016. https://en.bitcoin.it/wiki/ Double-spending.

ID: 42

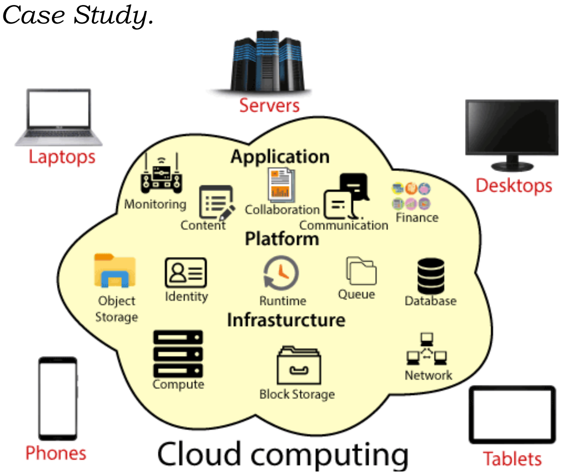
CLOUD DATABASE

Mr.S.SYED THOUFEEQ
II-BCA,
Sri Krishna Adithya College of Arts and Science,
Kovai Pudur, Coimbatore, Tamil Nadu, India.

Mr.S.SHARAN
II-BCA,
Sri Krishna Adithya College of Arts and Science,
Kovai Pudur, Coimbatore, Tamil Nadu, India.

ABSTRACT:
An introduction to cloud databases sets the foundation for understanding this vital component of modern computing. Cloud databases represent a paradigm shift in data management, offering businesses and individuals the ability to store, access, and manipulate data in a flexible, scalable, and cost-effective manner. In traditional database systems, data is stored on physical servers or on-premises infrastructure. However, with the advent of cloud computing, databases have moved to the cloud, allowing data to be stored and managed remotely on the servers of cloud service providers. This remote accessibility opens up new possibilities for organizations to store massive amounts of data without the burden of managing complex infrastructure. Cloud databases come in different models, including SQL, NoSQL, and New SQL, each tailored to specific use cases and requirements. Structured Query Language (SQL) databases are suitable for structured data and maintain strong data consistency, while NoSQL databases cater to unstructured and semi-structured data with flexible schema designs. New SQL databases combine elements of both SQL and NoSQL databases to offer scalability and ACID (Atomicity, Consistency, Isolation, Durability) properties. The advantages of cloud databases are manifold. They allow for easy scalability, enabling businesses to adapt to changing workloads without upfront investments in hardware. Moreover, cloud databases provide a pay-as-you-go model, reducing capital expenses and making them financially accessible to businesses of all sizes. However, with these advantages come challenges and concerns. Data security and privacy are major issues as data resides on external servers, raising questions about data breaches and unauthorized access. Additionally, data governance and compliance with regulations become paramount to ensure data integrity and legal adherence. Leading cloud service providers offer various cloud database solutions, each with unique features and performance capabilities

KEYWORDS: *Cloud Database Models, Cloud Database Architecture, Benefits Of Cloud Data Base, Challenges And Concerns, Cloud Database Providers, Migrating To The Cloud, Case Study.*



CLOUD DATABASE MODELS:

Cloud databases come in various models, each designed to cater to specific data management requirements and use cases. The three main cloud database models are:

1. SQL (Structured Query Language) Databases:

SQL databases are based on the traditional relational database management system (RDBMS) model. They use SQL as the query language to manage and manipulate structured data organized into tables with predefined schemas. Key characteristics of SQL databases include:

- Data integrity and consistency through ACID (Atomicity, Consistency, Isolation, Durability) properties.

- Support for complex queries using SQL, making them suitable for transactional and analytical workloads.
- Well-suited for applications with structured data and established data relationships.

2. NoSQL (Not Only SQL) Databases:

NoSQL databases are designed to handle unstructured or semi-structured data, which may not fit neatly into rigid table-based structures. They offer more flexibility and horizontal scalability compared to SQL databases. Key characteristics of NoSQL databases include:

- Schema-less or flexible schema design, allowing easy adaptation to changing data models.
- Horizontal scalability, making them ideal for handling large volumes of data and distributed architectures.
- Types of NoSQL databases include document stores, key-value stores, column- family stores, and graph databases.

3. New SQL Databases:

New SQL databases aim to combine the best of both SQL and NoSQL databases, offering the benefits of traditional ACID transactions while also providing the scalability and flexibility of NoSQL databases. Key characteristics of New SQL databases include:

- Horizontal scaling capabilities, similar to NoSQL databases.
- ACID-compliant transactions, ensuring data consistency and reliability for critical applications.
- Suitable for use cases where both scalability and data consistency are essential.

Choosing the appropriate cloud database model depends on the nature of the data, the specific application requirements, and the expected scale of the workload. SQL databases are well-suited for applications with structured data and complex queries, while NoSQL databases excel in handling unstructured and semi-structured data with dynamic schemas. New SQL databases are a compelling choice when there's a need for both scalability and transactional consistency.

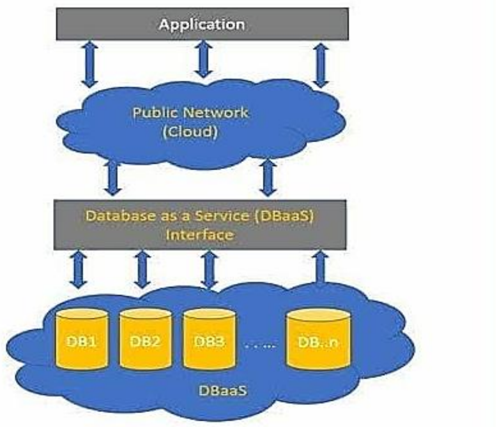
By understanding the characteristics of each cloud database model, organizations can make informed decisions about which model best aligns with their data management needs and can effectively support their applications in the cloud environment

CLOUD DATABASE ARCHITECTURE:

Cloud database architecture refers to the design and arrangement of components that enable the storage, management, and retrieval of data in a cloud computing environment. This architecture is specifically optimized to leverage the benefits of cloud computing, such as scalability, availability, and cost-effectiveness. Here are the key components of cloud database architecture:

1. Data Storage:

The foundation of cloud database architecture is data storage. Cloud service providers offer scalable and distributed storage systems to handle large volumes of data. The data can be organized into different storage tiers based on access patterns and performance requirements.



Cloudbase Database Architecture (DBaaS)

2. Database Management System (DBMS):

The DBMS is the software responsible for managing the database, handling data queries, and ensuring data integrity. It includes various components like the query engine, transaction manager, and data storage engine. In a cloud database, the DBMS is optimized for distributed environments to handle data across multiple nodes.

3. Replication and Data Distribution:

Cloud databases use replication and data distribution techniques to ensure high availability and fault tolerance. Data is replicated across multiple servers in different

ISBN: 978-81-965283-7-9

geographical locations to reduce the risk of data loss and improve performance.

4. Load Balancing:

Load balancing distributes incoming requests across multiple servers to optimize resource utilization and prevent overloading any single node. This ensures that the database can handle varying workloads efficiently.

5. Caching Mechanism:

Caching is employed to improve database performance by storing frequently accessed data in memory, reducing the need for repeated data retrieval from the underlying storage.

6. Security and Access Control:

Cloud database architecture incorporates robust security measures to protect sensitive data. Access control mechanisms are implemented to restrict unauthorized access and ensure data privacy.

7. Monitoring and Analytics:

Cloud database architecture often includes monitoring and analytics components to track database performance, detect anomalies, and optimize resource utilization.

8. Backup and Disaster Recovery:

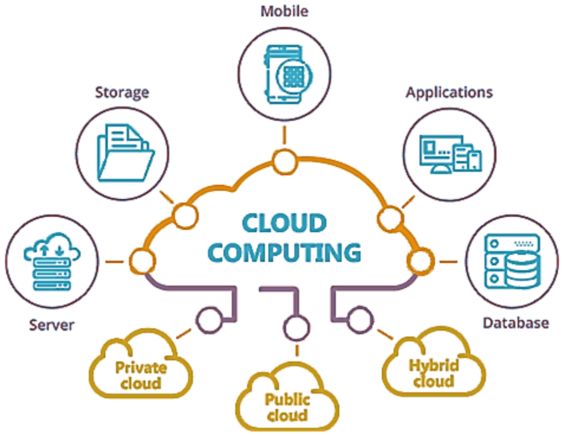
Cloud databases implement backup and disaster recovery strategies to safeguard data against unforeseen events. Regular backups are taken, and disaster recovery mechanisms are in place to recover data in case of failures.

9. Integration with Cloud Services:

Cloud database architecture seamlessly integrates with other cloud services, such as authentication, messaging, and analytics, to create a comprehensive cloud ecosystem.
The specific architecture may vary based on the cloud database model (SQL, NoSQL, or New SQL) and the chosen cloud service provider. Each provider offers its own set of features and tools for database management, so organizations must carefully design their cloud database architecture to meet their specific requirements while taking advantage of the cloud's capabilities.

BENEFITS OF CLOUD DATABASES:

Cloud databases offer numerous benefits that have revolutionized the way data is managed and accessed. Some of the key advantages of using cloud databases include:



1. **Scalability**: Cloud databases are highly scalable, allowing businesses to expand them data storage and processing capabilities dynamically. As data needs grow, cloud databases can easily accommodate increased workloads without the need for significant upfront investments in hardware.
2. **Cost-Effectiveness**: Cloud databases operate on a pay-as-you-go model, where organizations only pay for the resources they consume. This cost-effective approach eliminates the need for purchasing and maintaining physical infrastructure, reducing capital expenses and operational costs.
3. **Accessibility and** Flexibility: Cloud databases provide easy access to data from any location with an internet connection. This accessibility allows teams and stakeholders to collaborate seamlessly and work with real- time data, enhancing productivity and decision-making.
4. **Data Replication and High Availability**: Cloud databases often employ data replication techniques, storing copies of data across multiple servers in different locations. This redundancy ensures high availability, minimizing the risk of data loss due to hardware failures or disasters.
5. **Automated Backups and Disaster Recovery**: Cloud database services typically offer automated backups and disaster recovery options. Regular backups ensure data is protected, and

ISBN: 978-81-965283-7-9

in the event of a disaster, recovery mechanisms can swiftly restore data to a previous state.

6. **Global Reach:** Cloud databases are hosted on servers located in various regions worldwide. This global reach reduces data access latency for users in different geographical locations, ensuring a better user experience.
7. **Security**: Cloud service providers invest heavily in security measures to protect data. They employ encryption, access controls, and other security features to safeguard data from unauthorized access and breaches.
8. **Automatic Updates and Maintenance**: Cloud database providers handle maintenance tasks, including software updates and patches. This ensures that the database is running on the latest version, with improved features and security enhancements.
9. **Resource Optimization**: Cloud databases allow organizations to optimize resource allocation, automatically scaling up or down based on demand. This flexibility ensures optimal performance and resource utilization while minimizing costs.
10. **Integration with Cloud Ecosystem**: Cloud databases seamlessly integrate with other cloud services and tools, such as analytics, machine learning, and data warehousing, enabling organizations to build comprehensive data-driven solutions.

These benefits make cloud databases an attractive choice for businesses of all sizes, empowering them to manage data more efficiently, focus on their core competencies, and adapt to changing business requirements with ease.

CHALLENGES AND CONCERNS:

While cloud databases offer various benefits, they also present specific challenges and concerns that organizations need to address to ensure the security, integrity, and reliability of their data. Some of the key challenges and concerns with cloud databases include:

1. **Data Security and Privacy**: One of the most significant concerns is data security and privacy. Storing data in the cloud means relying on a third-party service provider to manage and protect sensitive information. Organizations must ensure that robust security measures, such as encryption, access controls, and authentication mechanisms, are in place to safeguard against data breaches and unauthorized access.
2. **Data Governance and Compliance**: Compliance with industry regulations and data governance standards is critical when handling sensitive data. Organizations must verify that their cloud database service provider complies with relevant regulations and offers tools to help them maintain data compliance.
3. **Data Loss and Recovery**: While cloud databases often have data replication and backup features, there is always a risk of data loss due to technical failures, human errors, or natural disasters. Adequate backup and disaster recovery strategies are essential to ensure data can be restored in case of any unforeseen events.
4. **Performance and Latency**: The performance of a cloud database heavily depends on the internet connection and network latency. Data access speed may vary based on the user's location and network quality. Organizations must consider latency and bandwidth requirements when selecting a cloud database provider.
5. **Vendor Lock-In**: Migrating data between different cloud database providers can be complex and costly. Organizations may face vendor lock-in if they heavily rely on proprietary features or services offered by a specific provider, making it challenging to switch to an alternative provider.
6. **Downtime and Availability**: Despite the high availability features of cloud databases, they are not immune to service outages or downtime. Organizations should inquire about the provider's uptime guarantees and explore strategies to mitigate the impact of potential downtime.
7. **Cost Management**: While cloud databases can offer cost savings, they can also lead to unexpected costs if not managed properly. Monitoring and optimizing resource usage are crucial to prevent over-provisioning and excessive expenses.
8. **Data Portability**: Ensuring data portability and the ability to retrieve data

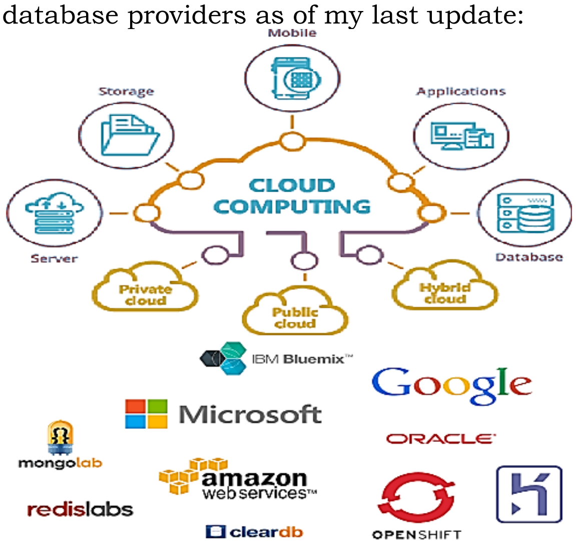
easily from a cloud database is essential. Organizations should consider how easily they can move their data from one cloud provider to another or bring it back on-premises if required.

9. **Compliance with Industry Standards**: Cloud databases must comply with industry- specific standards and certifications, such as HIPAA for healthcare or GDPR for data privacy in the European Union. Organizations must verify that their chosen provider adheres to these standards.

Addressing these challenges and concerns involves careful planning, choosing the right cloud database provider, implementing robust security measures, and having well- defined data management policies. By proactively addressing these issues, organizations can take full advantage of cloud databases while ensuring the safety and integrity of their data.

CLOUD DATABASE PROVIDERS:

As of my last update in September 2021, there were several major cloud database providers, each offering a range of database services to suit various application needs. Keep in mind that the cloud computing landscape is constantly evolving, and new providers or services may have emerged since then. Here are some prominent cloud database providers as of my last update:



1. **Amazon Web Services (AWS):**
   - Amazon Relational Database Service (RDS) for managed relational databases (MySQL, PostgreSQL, Oracle, SQL Server, etc.).
   - Amazon Dynamo DB for managed NoSQL databases.
   - Amazon Neptune for managed graph databases.

2. **Microsoft Azure:**
   - Azure SQL Database for managed SQL databases.
   - Azure Cosmos DB for globally distributed NoSQL databases.
   - Azure Database for PostgreSQL and Azure Database for MySQL for managed open- source databases.

3. **Google Cloud Platform (GCP):**
   - Cloud SQL for managed relational databases (MySQL, PostgreSQL, SQL Server).
   - Cloud Fire store and Cloud Data store for NoSQL databases.
   - Cloud Spanner for globally distributed horizontally scalable SQL databases.

4. **IBM Cloud:**
   - Db2 on Cloud for managed relational databases.
   - Cloudant for scalable NoSQL databases.
   - Compose Databases for various database offerings like MongoDB, Elastic search, etc.

5. **Oracle Cloud:**
   - Oracle Autonomous Database for automated and self-securing databases.
   - Oracle NoSQL Database Cloud Service for scalable NoSQL databases.

6. **Alibaba Cloud:**
   - *ApsaraDB for RDS for managed relational databases.*
   - ApsaraDB for MongoDB for managed MongoDB databases.
   - ApsaraDB for Redis for managed Redis databases.

7. **DigitalOcean:**
   - Managed Databases for managed PostgreSQL, MySQL, and Redis databases.

8. **MongoDB Atlas:**
   - A fully-managed, global cloud database service for MongoDB.

These are just a few examples of cloud database providers, and there are many other providers and specialized database services available in the market. When selecting a cloud database provider, it's essential to consider factors such as the specific database model needed (SQL, NoSQL, etc.), data requirements, performance, scalability, security, pricing, and the provider's global reach and compliance with industry regulations.

MIGRATING TO THE CLOUD:

Migrating to the cloud can bring significant benefits, but it requires careful planning and execution to ensure a smooth and successful transition. Here are the key steps involved in migrating to the cloud:

1. **Assess Current Infrastructure**: Start by analyzing your existing on-premises infrastructure, applications, and data. Identify the systems and workloads that are suitable for migration to the cloud.
2. **Set Goals and Objectives**: Define clear goals and objectives for the cloud migration. Determine what you want to achieve with the migration, such as cost savings, scalability, or improved performance.
3. **Choose the Right Cloud Provider**: Research and compare different cloud providers based on your requirements, budget, and desired features. Select a provider that best aligns with your needs and offers the services you require, such as cloud databases, compute instances, storage options, etc.
4. **Develop a Migration Plan**: Create a detailed migration plan outlining the sequence of steps and tasks required for each workload. Consider any dependencies between applications and services and plan for minimal disruption during the migration.
5. **Data Migration**: Determine the best approach to migrate your data to the cloud. This can involve using data transfer services, backup and restore, or database replication, depending on your data volume and sensitivity.
6. **Application Migration**: Assess your applications for cloud compatibility and make any necessary adjustments or modifications to ensure they run smoothly in the cloud environment. This may involve containerization, refactoring, or rearchitecting certain components.
7. **Testing and Validation:** Perform thorough testing and validation of your migrated applications and data to ensure they function correctly in the cloud. This includes testing for performance, security, and compliance.
8. **Implement Security Measures**: Implement robust security measures to protect your data and applications in the cloud. Use encryption, access controls, and other security features offered by your cloud provider.
9. **Train Your Team**: Provide training to your IT team and staff to ensure they are familiar with the cloud environment and can effectively manage and operate the cloud services.
10. **Monitor and Optimize**: Continuously monitor your cloud environment for performance, cost, and security. Optimize resource usage to ensure cost-effectiveness and adjust configurations as needed.
11. **Plan for Disaster Recovery**: Develop a comprehensive disaster recovery plan to ensure business continuity in case of any unforeseen events.
12. **Gradual Migration**: Consider a phased approach to migration, starting with non- critical workloads and gradually moving more critical applications as your team gains confidence and experience in the cloud environment.

Remember that cloud migration is a continuous process, and you may need to adapt your strategies as your business requirements evolve. By following a well-defined migration plan and leveraging the resources and expertise of your chosen cloud provider, you can successfully transition to the cloud and take advantage of its numerous benefits

CASE STUDIES:

Certainly! Cloud databases have been successfully adopted by numerous organizations, and several case studies demonstrate their effectiveness and impact. Here are a few notable case studies showcasing how businesses have benefited from using cloud databases:

1. Netflix:

- Netflix, a leading global streaming platform, migrated its massive video library and user data to Amazon Web Services (AWS) cloud database services.
- The move to the cloud allowed Netflix to scale its infrastructure rapidly to accommodate its growing user base and handle spikes in demand during

peak hours.

- With AWS's elasticity, Netflix optimized resource usage and improved the streaming experience for millions of users worldwide.

2. Airbnb:

- Airbnb, an online marketplace for lodging and travel experiences, shifted its data storage and management to Amazon Dynamo DB, a NoSQL database service on AWS.
- By using a NoSQL database, Airbnb could handle its rapidly increasing amount of unstructured user data, such as user reviews and messages.
- The cloud database provided flexibility and scalability, enabling Airbnb to maintain a seamless user experience during periods of high traffic and dynamic data growth.

NASA Jet Propulsion Laboratory (JPL):

- JPL, a NASA center responsible for space exploration and missions, leveraged Google Cloud's Big table to store and process large volumes of satellite imagery and scientific data.
- By using Big table, JPL could manage petabytes of data efficiently and access data in real-time, enabling faster analysis and decision-making for space missions.

General Electric (GE):

- GE, a multinational conglomerate, employed Microsoft Azure's cloud database services to support its Predix platform for Industrial Internet of Things (IIoT) applications.
- Azure Cosmos DB, a globally distributed NoSQL database, facilitated GE's IIoT data collection, analytics, and predictive maintenance capabilities across its industrial machinery and equipment.

3. Capital One:

- Capital One, a financial services company, adopted AWS's cloud database solutions to enhance its mobile banking application.
- The migration to the cloud enabled Capital One to deliver real-time account information to customers, improve application responsiveness, and provide a better user experience. These case studies illustrate the versatility and advantages of cloud databases across diverse industries. Organizations like Netflix, Airbnb, NASA JPL, General Electric, and Capital One have experienced improved performance, scalability, and cost-effectiveness by leveraging cloud database solutions, driving innovation, and enhancing customer experiences.

Future Trends in CloudDatabases:

As technology continues to evolve, cloud databases are expected to undergo several transformative trends in the future. Some of the key trends in cloud databases are:

1. **Server less Database Architecture**:

Server less database architecture is gaining popularity, where the cloud provider manages all aspects of the database, including scaling, provisioning, and backups. This approach allows developers to focus solely on application logic without worrying about database management tasks.

2. **Edge Computing and Cloud Databases**:

The rise of edge computing will lead to closer integration between edge devices and cloud databases. Edge computing reduces latency and bandwidth requirements by processing data locally on edge devices before selectively uploading it to the cloud for storage and further analysis.

3. **Multi-Cloud and Hybrid Cloud Databases**:

Organizations will increasingly adopt multi-cloud and hybrid cloud strategies to avoid vendor lock-in, improve redundancy, and optimize cost and performance by leveraging multiple cloud providers for different workloads.

4. **Advanced Data Analytics**:

Cloud databases will be further integrated with advanced data analytics and machine learning tools, enabling real-time data analysis and predictive capabilities directly on the cloud data.

5. **Block chain Integration**:
Block chain technology will likely be integrated into cloud databases to enhance data security, transparency, and traceability for critical applications like supply chain management and financial transactions.

6. **Increased Emphasis on Data Privacy and Compliance**:
As data privacy regulations continue to evolve globally, cloud database providers will need to enhance their security and compliance features to meet stringent data protection requirements.

7. **Quantum Computing Impact**:
The advent of quantum computing may lead to new possibilities and challenges for cloud databases. Quantum-resistant encryption and quantum database algorithms may become essential to secure sensitive data.

8. **Performance Improvements**:
Cloud database providers will continue to enhance the performance of their offerings, reducing data access latency and optimizing query response times to handle massive data volumes more efficiently.

9. **Focus on Sustainability**:
Cloud database providers may emphasize sustainability by adopting greener data centers and implementing energy-efficient technologies to reduce their environmental impact.

10. **Enhanced Data Mobility**:
Cloud database services will likely provide improved data portability and migration tools, making it easier for organizations to move their data between cloud providers or between the cloud and on-premises environments. Overall, the future of cloud databases will revolve around empowering businesses with more streamlined, secure, and intelligent data management capabilities, enabling them to extract greater value from their data assets while meeting the demands of a rapidly evolving digital landscape.

CONCLUSION:

In conclusion, cloud databases have emerged as a game-changing technology, revolutionizing the way organizations manage and access their data. The migration of databases to the cloud has unlocked numerous benefits, enabling businesses to scale their operations, optimize costs, and improve data accessibility. As cloud computing continues to evolve, cloud databases will play a pivotal role in shaping the future of data management.

The advantages of cloud databases include unparalleled scalability, cost-effectiveness, and global accessibility, making them a compelling choice for businesses of all sizes. Organizations can focus on their core competencies while leaving the complexities of database management to cloud service providers. However, with these advantages come challenges and concerns. Data security and privacy remain critical considerations, necessitating robust security measures and compliance with industry regulations. Effective data governance and backup strategies are vital to ensure data integrity and business continuity. Looking ahead, cloud databases are poised for transformative trends, such as server less architecture, edge computing integration, and advanced data analytics. The rise of multi- cloud and hybrid cloud strategies will provide organizations with greater flexibility and redundancy, reducing vendor lock-in risks. As cloud database technology advances, data- driven organizations must stay informed about emerging trends, implement best practices, and continuously optimize their cloud database strategies to leverage its full potential. With careful planning, adherence to security best practices, and collaboration with reputable cloud providers, businesses can harness the power of cloud databases to drive innovation, enhance customer experiences, and thrive in an ever-evolving digital landscape. The future of data management belongs to the cloud, and organizations that embrace this transformative technology will be at the forefront of success in the digital era.

REFERENCE:

1. https://www.ibm.com/cloud/learn/what-is-cloud-database
2. https://www.techtarget.com/searchcloudcomputing/definition/cloud-database
3. https://www.oracle.com/in/database/what-is-a-cloud-database/
4. https://www.mongodb.com/cloud-database
5. https://www.simplilearn.com/cloud-databases-across-the-globe-article

ID: 43

ASPECT EYE ON BIG DATA & HADOOP

Mr.YASWANTH MAHAA
II BCA,
Sri Krishna Adithya College of Arts & Science,
Coimbatore, Tamil Nadu, India.

Dr.K.GEETHA
Head,
Department of BCA, B.Sc AI&ML, B.Sc Data Science,
Sri Krishna Adithya College of Arts & Science,
Coimbatore, Tamil Nadu, India.

ABSTRACT

Computing technology has revolutionized how we work, learn, and live. One of the hottest topics in the IT world is distributed data processing technology. It offers a straightforward and centralized computing platform by decreasing the cost of hardware. The features of distributed data processing technology have revolutionized the entire industry. Hadoop is the most popular open-source project of the Apache foundation. The most representative platform of distributed big data processing is Hadoop. The distributed framework of Hadoop has provided a secure and fast big data processing structure. Users can design distributed applications without knowing the details of the system at the bottom layer. In this thesis, we will briefly introduce you to the Hadoop platform. Because of the complexity of the platform, we will only focus on the key technologies of Hadoop, such as HDFS, Map Reduce and Hbase.

What is Big Data?

Here Comes the Question as Big Data the Big Data is a large and complex set of data that is difficult to manage, store, and analysed within a reasonable timeframe. This necessitates the development of a new data processing model that has the capacity to store, process, and analyses the data more efficiently. This is the driving force behind the emergence of Big Data technology. This technology provides a new approach to extracting, interacting, integrating, and analysing Big Data. The goal of the Big Data strategy is to extract the significant and valuable data information behind the Big Data through specialized processing. To compare the Big Data with an industry, it is the industry's key to creating the data value by increasing the processing power of the data.

Big Data manages an Extremely Large Data:

What is Hadoop?

The Apache Software Foundation (ASF) developed the open-source framework, Hadoop, to facilitate the storage and processing of large datasets, making it a popular choice for applications and operations in the fields of big data analytics and data mining, as well as predictive analytics, machine learning, and more.

Hadoop works by aggregating computers and servers together, called commodity hardware, to create a distributed model of data processing, allowing for faster, more efficient, and more reliable data analysis.

Additionally, Hadoop is capable of working with a wide range of data formats, sizes, and data types, including both structured and non-structured data. It is also supported by a variety of additional applications and projects.

What are the benefits of Hadoop?

Hadoop is capable of storing, handling, and processing petabytes of data. It has revolutionized big data, social networks, and the Internet of Things (IoT). It even supports Yahoo search engine. Hadoop is notable for its ability to run parallel tasks and jobs, scalability, and fault tolerance and high availability.

- Data is protected from hardware loss and failure due to distributed computing and storage in Hadoop.
- It is also flexible, allowing you to store a wide range of data types without pre-processing. Unlike traditional relational databases, Hadoop does not require any pre-processing. Cost-effective and free,

- Hadoop is widely used in the financial sector, healthcare, and beyond. Cost-effectiveness. Hadoop's open-source framework is free, and its use cases span the financial sector to healthcare and beyond.

Where Do We Use Hadoop?

How does Hadoop work?
Hadoop processes data using a few key steps:

1. Applications collect data submitted by the client.
2. The data is stored on a Hadoop cluster called a Data Node, and is maintained and managed by the Name Node server – effectively the master node – via an application programming interface (API). This Name Node will oversee multiple Data Nodes.
3. The data is then processed.

It is written in Java, but Hadoop programmes can also be coded in Python or C++ programming languages.

HADOOP'S MAIN MODULES
Hadoop has four core components:

Hadoop Common
Hadoop Common provides the common Java libraries that support Hadoop's other modules.

Hadoop Distributed File System (HDFS)
HDFS is the distributed file system that provides better data throughput – the amount of data successfully moved from one place to another during a given time period – than traditional systems, as well as higher fault tolerance. Through metadata, HDFS structures directories and files in a tree, and attributes information such as permissions, ownership, and replication factors.

Hadoop YARN (Yet another Resource Negotiator)
YARN is Hadoop's framework for job scheduling and cluster resource management. It manages and monitors cluster nodes and resource usage, and schedules jobs and tasks.

Hadoop Map Reduce
Map Reduce is a YARN-based framework that enables the parallel processing of large data sets.

Map Reduce is a programming model in which input data is transformed into data sets that can be calculated in the form of key value pairs (KVPs). The output of a Map Reduce job is then aggregated. The Hadoop ecosystem.

There are a number of different applications, technologies, and tools that sit within what's known as the Hadoop ecosystem.

All of these applications use Hadoop technologies to collect, store, process, analyse, and manage data:

- Spark is a super-fast computing engine that can handle a lot of big data. It's easy to use and has a programming model that makes it easy to work with big data. Plus, it runs in-memory and has optimised execution, so you can get the most out of it. It can handle a bunch of different applications, like machine learning, stream processing, analytics, graph computing, databases, batch processing, ad hoc queries, and more.
- Ambari is a web tool that helps you manage and keep an eye on your Hadoop cluster. It's used to manage and monitor your cluster, and it supports a bunch of different apps, like HDFS and Map Reduce. It also has dashboards to help you keep an eye on the health of your cluster.
- Avro is a database that can be used to serialize large amounts of data. It's scalable and has no single point of failure.
- What is Cassandra?
- Cassandra is a database that can be used with multiple servers. It's super easy to use and doesn't have any major issues.
- Chukwa is a distributed data collection platform designed to facilitate the management of large-scale distributed systems.
- What iso Flume? What is Flume? Flume is a software package that facilitates the efficient collection, consolidation, and transfer of large volumes of log data.
- Apache HBase is an all-in-one database and data store that can store and store data in a way that's not relational. It's also scalable and can be versioned and distributed, and it can store structured data. Plus, you can access huge tables in real-time, even if

they have billions of rows or millions of columns.

- Hive is a data warehouse that can be distributed and fault tolerant. It's great for aggregating data, doing ad-hoc queries, and doing analytics on a huge scale.
- Impala is an SQL query engine developed by Apache that is part of the Apache MSP (Multi-Processed) family. It is supported by the Clouderar platform...
- Kafka If you're looking for a way to store and process events in a distributed way, then look no further than Apache Kafka. It's the perfect solution for stream-processing and event storage.
- Mahout is a great tool for anyone who wants to use machine learning or data mining. It's super easy to use and can be scaled up or down as needed. Ozie. What is Oozie? Oozie is a scheduling tool that helps you keep track of your Hadoop jobs.
- Ozone is a distributed object store that can be scaled up or down as needed.
- Pig: If you're using Hadoop, you'll need to know about Apache Pig. It's the language and execution language that powers the data-flow in Hadoop. Framework for parallel computation.
- Presto is a SQL query engine that helps you analyse data quickly and easily. It's designed to be fast and easy to use, and it can handle data from multiple sources like HDFS or Amazon S3.
- Presto (formerly known as SQL Query Engine) is an SQL query engine designed to facilitate the rapid and user-friendly analysis of data. It is capable of processing data from a variety of sources, such as HDFS (High-Level File System) and Amazon S3 (Amazon S3). Sqoop (formerly known as Sqoop) is a software interface application that facilitates the transfer of data from relational databases to Hadoop.
- Storm is a tool that helps you process streams in a distributed way.
- Submarine is a unified AI platform designed to facilitate the execution of machine learning (ML) and deep learning (DL) workloads across distributed clusters by engineers and data scientists.
- Tez is a YARN-based data-flow programming language that is being adopted by many applications as an alternative to Map Reduce as the main executive engine. Zeppelin. What is Zeppelin? Zeppelin is an open-source, interactive notebook that allows you to explore data in a Zookeeper is a service that helps you keep track of all your shared apps.

Explore and Apply Hadoop Technologies:

Gain an in-depth understanding of the programming language Apache Hadoop, as well as develop other essential computer science competencies, through an MSc Computer Science course delivered entirely online. Developed for ambitious individuals with no prior experience in the field of computer science, this flexible Master's degree is ideal for those who wish to pursue a career in the field.

The degree contains a core module in Database Systems and Security, which allows the student to investigate the use of Hadoop technology in conjunction with other data sources and database systems in order to gain an understanding of their applicability. This includes topics such as:

- Database concepts and techniques
- Database design (Oracle Data Modeller)
- Relational model concepts working with database technologies
- SQL and Oracle
- Non-relational systems such as NoSQL, MongoDB, and Oracle NoSQL
- Spark environments
- Database management and administration management of data
- Data governance
- Data security.

This module will provide you with the foundational competencies required to design and deploy database systems that enable the processing of data and ensure its security. You will gain a fundamental understanding of and demonstrate proficiency in the tools and methods employed in developing database applications, as well as the fundamentals of data management and security.

REFERENCES:

1. https://core.ac.uk/
2. University of Wolverhampton, United Kingdom
3. https://online.wlv.ac.uk/what-is-hadoop/#:~:text=Apache%20Hadoop%20is%20an%20open,predictive%20analytics%2C%20and%20machine%20learning.
4. Acharjya and Ahmed, 2016aD. Acharjya, K.P. Ahmed, Survey on big data analytics: challenges, open research issues and tools, Int. J. Adv. Computer. Sci. App., 7 (2016), pp. 511-518.
5. Aher and Kulkarni, 2015, S.B. Aher, A.R. Kulkarni Hadoop mapreduce: a programming model for large scale data processing, Am. J. Comput. Sci. Eng. Surv. (AJCSES), 3 (2015), pp. 01-10.

ID: 44

BIO-MEDICAL AND MEDICAL IMAGING

Mr.S.BALASUBRAMANI
Assistant Professor,
Deprtment of Data Science,
Sri Krishna Adithya College of Arts and Science,
Coimbatore, Tamil Nadu, India.

Mr.S.GOKULAKRISHNAN
I - B.Sc. Data Science,
Department of Data Science,
Sri Krishnaadithya College of RTS and Science,
Coimbatore, Tamil Nadu, India.

ABSTRACT: -

Medical imaging, on the other hand, focuses specifically on imaging techniques used for diagnosing and monitoring medical conditions in humans. It encompasses non-invasive and minimally invasive modalities such as X-ray imaging, computed tomography (CT), magnetic resonance imaging (MRI), ultrasound imaging, and nuclear medicine imaging. These technologies enable healthcare professionals to visualize anatomical structures, physiological functions, and pathological changes within the human body.

Biomedical and medical imaging are advanced technologies that provide non-invasive or minimally invasive visualization of the human body's internal structures and functions. From cellular research to clinical diagnosis, these imaging modalities play a pivotal role in modern healthcare. This abstract highlights the significance and applications of biomedical and medical imaging, offering a glimpse into their essential role in advancing medical practice and improving patient care.

KEYWORDS: *Biomedical imaging, Medical imaging, Non-invasive, Minimally invasive, Cellular research, Clinical diagnosis, Advanced technologies, Patient care, Modern medicine, Visualization, Imaging modalities, Early detection.*

History of Bio Medical: -

The history of biomedical research and innovation is vast and complex, with significant milestones spanning centuries. Here's an overview of some key events and developments in the history of biomedical science:

1. **Ancient Times:**
   - Early civilizations, such as the Egyptians and Greeks, made observations about anatomy and medicinal plants.
   - The Hippocratic Corpus, attributed to the Greek physician Hippocrates, laid the foundation for systematic medical practice and ethics.

2. **Middle Ages:**
   - Arab scholars preserved and advanced medical knowledge, translating Greek texts and making significant contributions to medicine.
   - The development of hospitals and medical schools in the Islamic world promoted the dissemination of medical knowledge.

3. **Renaissance and Early Modern Era:**
   - The Renaissance witnessed renewed interest in anatomy and dissection, with Andreas Vesalius revolutionizing anatomical studies with his work "De humani corporis fabrica."
   - The invention of the microscope by Antonie van Leeuwenhoek opened new avenues for studying microorganisms.

4. **19th Century:**
   - The discovery of vaccination by Edward Jenner in 1796 marked a significant breakthrough in preventing infectious diseases.
   - Louis Pasteur's germ theory laid the foundation for understanding the role of microorganisms in disease.

- The development of anaesthesia by William Morton and others revolutionized surgery.

5. **20th Century:**
   - The discovery of penicillin by Alexander Fleming in 1928 heralded the era of antibiotics and revolutionized the treatment of infectious diseases.
   - The Human Genome Project, initiated in 1990, aimed to map the entire human genome, leading to ground breaking advances in genetics and personalized medicine.
   - The development of imaging technologies, such as X-ray, MRI, and PET, transformed medical diagnosis and research.

6. **Contemporary Biomedical Research:**
   - Advancements in molecular biology, genetic engineering, and stem cell research have led to new therapeutic possibilities and gene editing technologies like CRISPR-Cas9.
   - The emergence of personalized medicine aims to tailor treatments to an individual's genetic makeup and specific health needs.
   - The field of biotechnology has seen the development of novel drugs, vaccines, and diagnostic tools.

INTRODUCTION:

Biomedical and medical imaging play a crucial role in modern healthcare, providing valuable insights into the human body's structure, function, and disease processes. These advanced imaging techniques have revolutionized medical diagnosis, treatment planning, and research. Here's an overview of biomedical and medical imaging:

1. Biomedical Imaging:

- ❖ Biomedical imaging refers to the visualization of biological tissues, organs, and processes using various imaging modalities.
- ❖ It allows non-invasive or minimally invasive examination of living organisms, aiding in the understanding of normal physiology and pathological conditions.
- ❖ Biomedical imaging techniques are used across various fields, including medicine, biology, neuroscience, and veterinary science.

2. Medical Imaging:

- ❖ Medical imaging is a subset of biomedical imaging focused on the visualization of the human body for diagnostic and therapeutic purposes. *It enables healthcare professionals to visualize and analyze internal structures, aiding in the detection, characterization, and monitoring of diseases and injuries.
- ❖ Medical imaging plays a key role in clinical decision-making and treatment planning.

3. Types of Medical Imaging Techniques:

There are several types of medical imaging techniques, each utilizing different physical principles and technologies. Some common modalities include:

- ❖ X-ray Imaging
- ❖ Computed Tomography (CT)
- ❖ Magnetic Resonance Imaging (MRI)
- ❖ Ultrasound Imaging
- ❖ Nuclear Medicine (e.g., PET and SPECT)
- ❖ Optical Imaging (e.g., Endoscopy)

4. Applications of Biomedical and Medical Imaging:

Biomedical and medical imaging find extensive applications in various medical specialties:

- ❖ Radiology: Diagnosis and treatment planning in conditions like fractures, tumors, and infections.
- ❖ Cardiology: Assessing heart function and identifying vascular issues.
- ❖ Oncology: Detecting and staging cancerous tumors.
- ❖ Neurology: Visualizing the brain and nervous system for neurological disorders.
- ❖ Obstetrics and Gynecology: Monitoring fetal development and reproductive health.

5. Advancements and Challenges:

Advancements in medical imaging technology have led to higher resolution, faster imaging, and functional imaging capabilities. The integration of artificial intelligence and machine learning has shown promise in enhancing image analysis and automation. However, challenges remain, such as radiation exposure in certain modalities, cost, and accessibility, especially in resource-constrained settings.

6. Types of Medical Imaging Techniques:

Medical imaging offers a diverse array of techniques, each employing unique principles and technologies. X-ray imaging is a widely used modality for bone imaging and chest examinations. CT scans use X-rays to create detailed cross-sectional images of the body. MRI employs powerful magnets and radio waves to produce high-resolution images of soft tissues and organs. Ultrasound imaging utilizes sound waves to visualize internal structures and is commonly used in obstetrics and cardiology. Nuclear medicine techniques like PET and SPECT provide functional information, particularly valuable in oncology and neurology.

7. Clinical Applications:

Medical imaging finds applications across various medical specialties. In radiology, it aids in diagnosing fractures, detecting tumors, and assessing internal injuries. In cardiology, imaging techniques help evaluate heart function, identify vascular issues, and plan treatments for cardiovascular diseases. Neurology benefits from MRI and CT scans to visualize the brain and diagnose neurological disorders. In oncology, medical imaging is instrumental in cancer detection, staging, and treatment monitoring.

8. Advancements and Future Directions:

Recent advancements in medical imaging technology have led to faster and higher-resolution imaging, enabling improved diagnosis and treatment planning. Functional imaging techniques, such as fMRI, are uncovering insights into brain activity and neural networks. Additionally, the integration of artificial intelligence and machine learning is enhancing image analysis, aiding in automated detection and diagnosis of diseases.

9. Medical Imaging:

Medical imaging is a branch of biomedical science that involves the use of various imaging technologies to visualize and create detailed images of the internal structures and functions of the human body. These imaging techniques play a crucial role in medical diagnosis, treatment planning, and monitoring of various diseases and conditions.

Types of Medical Imaging:

There are several types of medical imaging modalities, each utilizing different principles and technologies to capture specific aspects of the human body. Some common medical imaging techniques include:

1. **X-ray Imaging:** X-rays use electromagnetic radiation to penetrate the body and create images of bones and dense tissues. They are commonly used to diagnose fractures, lung conditions, and dental issues.
2. **Computed Tomography (CT):** CT scans combine X-rays and computer technology to produce detailed cross-sectional images of the body. CT scans are useful for visualizing internal organs and soft tissues.
3. **Magnetic Resonance Imaging (MRI):** MRI uses powerful magnets and radio waves to create high-resolution images of soft tissues, such as the brain, muscles, and joints. It is valuable for diagnosing neurological disorders, joint injuries, and tumors.
4. **Ultrasound Imaging:** Ultrasound uses high-frequency sound waves to create real-time images of internal organs and structures. It is commonly used in obstetrics to monitor fetal development and in various medical specialties for diagnostic purposes.
5. **Nuclear Medicine:** Nuclear medicine imaging involves the use of small amounts of radioactive materials (radiotracers) to diagnose and treat diseases. Techniques like Positron Emission Tomography (PET) and Single Photon Emission Computed Tomography (SPECT) provide functional information about organs and tissues.
6. **Fluoroscopy:** Fluoroscopy is a real-time imaging technique that uses X-rays to visualize moving body parts, such as the digestive system and blood vessels.
7. **Mammography:** Mammography is a specific type of X-ray used to screen and diagnose breast conditions, including breast cancer.

Medical imaging is essential in various medical specialties, including radiology, cardiology, oncology, neurology, and orthopaedics. It enables healthcare professionals to make accurate diagnoses, plan appropriate treatments, and monitor the effectiveness of therapies. Advancements in medical imaging technologies continue to enhance diagnostic capabilities and contribute to improved patient outcomes in modern healthcare.

Advantages of Biomedical and Medical Imaging:

1. Early Disease Detection: Biomedical and medical imaging techniques allow early detection of diseases, leading to timely intervention and improved treatment outcomes.
2. Non-Invasive: Many imaging techniques are non-invasive, meaning they do not require surgical procedures, reducing patient discomfort and recovery time.
3. Visualization of Internal Structures: Medical imaging provides detailed images of internal structures, allowing healthcare professionals to observe organs and tissues without the need for exploratory surgery.
4. Treatment Planning: Imaging plays a crucial role in treatment planning, helping doctors identify the precise location and extent of abnormalities or diseases, guiding surgeries, and determining appropriate therapies.
5. Research and Education: Medical imaging aids in medical research and education by providing a better understanding of human anatomy and disease processes.
6. Real-Time Monitoring: Some imaging methods, such as ultrasound and fluoroscopy, allow real-time monitoring during medical procedures, improving accuracy and safety.
7. Minimal Radiation Exposure: Advancements in imaging technology have reduced radiation exposure in many imaging modalities, making them safer for patients.

Disadvantages of Biomedical and Medical Imaging:

1. Cost: Some medical imaging procedures can be expensive, making them less accessible to certain individuals or regions with limited healthcare resources.
2. Radiation Exposure: Certain imaging methods, such as X-rays and CT scans, involve ionizing radiation, which can pose risks in high doses over repeated exposures.
3. False Positives/Negatives: Imaging results may occasionally yield false positives (indicating disease when none is present) or false negatives (failing to detect a disease that is present), leading to potential misdiagnoses.
4. Overutilization: Over-reliance on imaging can sometimes lead to unnecessary tests, which may drive up healthcare costs and expose patients to avoidable risks.
5. Limited Sensitivity in Some Cases: Some imaging modalities may not be sensitive enough to detect certain conditions or early-stage diseases.
6. Interpretation Complexity: Interpreting medical images requires specialized training and expertise, and errors in interpretation can lead to incorrect diagnoses or treatment plans.
7. Patient Cooperation: Certain imaging procedures may require patients to remain still or hold their breath, which can be challenging for some individuals, particularly young children or patients with certain medical conditions.

It's important to note that while medical imaging has many advantages, its use should always be judicious and based on medical necessity to minimize potential risks and costs. Healthcare professionals carefully weigh the benefits and drawbacks of imaging to make informed decisions for their patients' well-being.

CONCLUSION: -

In conclusion, biomedical and medical imaging technologies have revolutionized the field of medicine, offering invaluable tools for early disease detection, non-invasive visualization of internal structures, treatment planning, and real-time monitoring during medical procedures. These techniques have significantly improved patient outcomes, advanced medical research, and enhanced medical education.

The advantages of biomedical and medical imaging, such as early detection, non-invasiveness, and accurate visualization, have contributed to better patient care and treatment strategies. Medical imaging has become a cornerstone in modern healthcare, enabling healthcare professionals to diagnose and manage diseases more effectively.

However, it is crucial to acknowledge the disadvantages associated with medical imaging, such as cost, radiation exposure, and potential false results. Striking a balance between appropriate utilization and minimizing risks remains essential to ensure the best possible outcomes for patients.

Continuous advancements in technology, safety protocols, and research efforts aim to further enhance the benefits of biomedical and medical imaging while mitigating potential drawbacks. With responsible and well-informed use, these imaging techniques will continue to play a vital role in improving healthcare and advancing medical knowledge, ultimately contributing to a healthier and more informed society.

REFERENCE:

1. M.A. McWade, et al., "Label-free Intraoperative Parathyroid Localization With Near-Infrared Autofluorescence Imaging," The Journal of Clinical Endocrinology & Metabolism, 99, Issue 12, December 2014, Pages 4574–4580.
2. O.A. Okoh, *et al.,* "Promising near-infrared non-targeted probes: benzothiazole heptamethine cyanine dyes," Journal of Sulfur Chemistry, 35:1, 42 (2014).
3. Kim, E., Ahn, T., & Kumazaki, S. "Changes in Antenna Sizes of Photosystems during State Transitions in Granal and Stroma-Exposed Thylakoid Membrane of Intact Chloroplasts in Arabidopsis Mesophyll Protoplasts," Plant and Cell Physiology, 56(4), 759-768 (2015).
4. J.Q.Nguyen, et al., "Near-infrared auto fluorescence spectroscopy of in vivo soft tissue sarcomas," Opt. Lett. 40, 5498 (2015).
5. J. Schleusener, *et al.*, "Depth-dependent auto fluorescence photo bleaching using 325, 473, 633, and 785 nm of porcine ear skin ex vivo," Journal of Biomedical Optics, 22(9), 091503 (5 January 2017).

ID: 45

MACHINE LEARNING AND DEEP LEARNING APPLICATIONS

Dr.J.SANTHOSH
Assistant Professor,
Department of Computer Applications,
Sri Krishna Adithya College of Arts & Science,
Coimbatore, Tamil Nadu, India

Mr.S.AHAMED JOHNSHA ALI
Assistant Professor,
Department of Computer Applications,
Sri Krishna Adithya College of Arts & Science,
Coimbatore, Tamil Nadu, India

ABSTRACT

In this fast growing era, financial sector is the most vulnerable sector of the economy which is susceptible to data loss and occurrence of fraud Hence Financial sector needs a continuous improvement in field of data protection as majority of the transactions are migrating towards virtual platforms in the name of technology upgradation and ease in carrying the transaction at one's fingertips. Though technology upgradation is this area is happening in a higher pace, the techniques to breach data is also steadily upgrading.

Machine learning and Deep Learning are widely used in finance for fraud detection, credit risk assessment, algorithmic trading, and customer churn prediction. These applications help financial institutions make knowledgeable decisions and avert potential risks.

KEYWORDS: *Machine Learning, Deep Learning, Credit Risk Assessment, Algorithmic Trading, Customer Churn Prediction.*

I. INTRODUCTION

Fraud detection using Machine Learning

Fraud detection using machine learning techniques has become a necessary application in the financial industry and after. The goal is to automatically identify and prevent defrauding activities, which can save businesses and individuals from notable financial losses. Here's a review of how machine learning is used in fraud detection:

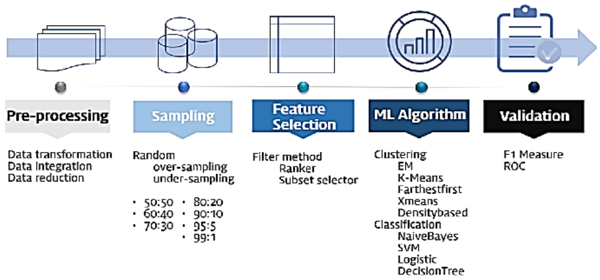


Fig1.Implementation of ML in Fraud Detection

1. **Data Collection and Pre-processing**: The first step in any machine learning application is data collection. In fraud detection, this involves assembling historical transaction data, including legal and false transactions. The data is then pretreated to handle missing values, normalize features, and remove noise.
2. **Feature Engineering:** Pertinent features are extracted from the data to constitute transactions effectively. These features may include transaction amount, time, location, device information, user behaviour patterns, and more. Proper feature engineering is decisive for the model's efficacy.
3. **Model Selection:** Various machine learning algorithms can be employed for fraud detection, including logistic regression, decision trees, random forests, support vector machines (SVM), and neural networks. Ensemble methods like Gradient Boosting and XGBoost are also commonly used to upgrade performance.
4. **Imbalanced Data Handling**: Fraudulent transactions are usually infrequent compared to legitimate ones, leading to imbalanced datasets. Special care is taken to handle this class imbalance to stop the model from being biased

regarding the majority class. Techniques such as oversampling, under sampling, and Synthetic Minority Over-Sampling Technique (SMOTE) are used to address this issue.

5. **Model Training**: The prepared dataset is divided into training and testing sets. The model is instructed on the labelled data (with both legitimate and fraudulent transactions) to learn the patterns and attributes of each class.
6. **Model Evaluation**: The trained model is evaluated using the testing set to assess its performance. Metrics such as precision, recall, F1-score, and area under the Receiver Operating Characteristic (ROC) curve are used to judge the model's ability to detect fraud precisely.
7. **Real-time Monitoring**: In real-world scenarios, fraud detection systems need to work in real-time. Therefore, the trained model is used in the production environment to monitor incoming transactions and quickly flag unsure ones.
8. **Continuous Learning**: Fraudsters continuously modify their techniques, making it necessary for fraud detection models to learn from new data continually. Periodic model retraining ensures the system stays up-to-date and effective.
9. **Anomaly Detection**: Fraudulent transactions often exhibit abnormal behaviour compared to legitimate ones. Anomaly detection algorithms, such as Isolation Forest or One-Class SVM, can be used to identify outliers in the data that may indicate fraudulent activities.
10. **Collaborative Filtering:** Sometimes, fraudsters collaborate and work together to commit fraud. Collaborative filtering techniques can be used to uncover hidden relationships between organizations involved in false activities.
11. **Feature Scaling:** Scale or normalize the features to make sure that they are on a similar scale. This step is especially important for algorithms that are sensitive to feature scales, such as distance-based algorithms.
12. **Model Selection:** Choose appropriate machine learning algorithms for fraud detection. Commonly used algorithms include:

    a. **Logistic Regression:** A simple and accountable algorithm that can provide the probability of a transaction being fraudulent.
    b. **Decision Trees:** Acceptable for handling non-linear relationships and can provide explicable rules for fraud detection.
    c. **Random Forest:** A method that combines multiple decision trees for improved performance.
    d. **Gradient Boosting:** Another method that boosts multiple weak learners to create a strong classifier.
    e. **Support Vector Machines (SVM)**: Effective for splitting data into different classes based on a hyperplane.

13. **Hyperparameter Tuning**: Maximize the hyperparameters of the chosen algorithm to improve the model's performance. Techniques like grid search or random search can be used for hyperparameter tuning.
14. **Ensemble Methods**: Consider using ensemble methods like voting classifiers or stacking multiple models to combine their predictions and improve overall fraud detection accuracy.

Financial fraud detection using deep learning algorithms:

Financial fraud detection using deep learning algorithms has shown significant promise due to their ability to automatically learn complex patterns and representations from data. Deep learning models, especially neural networks, can capture complex features that may be challenging for traditional machine learning algorithms to extract.

Deep learning models can productively capture complex patterns in financial transaction data, enabling accurate and efficient fraud detection. However, it's critical to have a sufficient amount of high-quality data and computational resources to train and deploy deep learning models successfully. Additionally, interpretability can be a challenge with deep learning models, so considering techniques to interpret the model's decisions may be necessary for certain applications.

Financial fraud detection and predictive algorithms are essential tools used by financial institutions and businesses to identify fraudulent activities before they cause significant financial losses. These algorithms aim to analyze transaction data and detect suspicious patterns or anomalies that may indicate potential fraud.

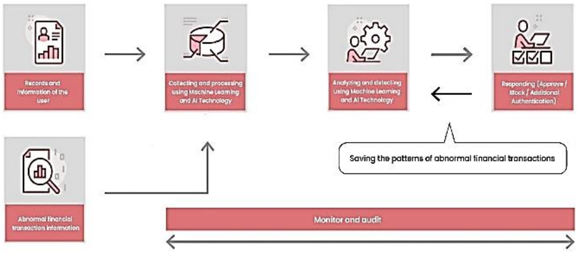


Fig 2. The Fraud Detection System Process

Financial fraud detection and predictive algorithms play a crucial role in shielding financial systems and protecting both businesses and customers from potential fraud. Continuous improvement, model updating, and the integration of multiple techniques are essential for staying ahead of ever-evolving fraud tactics. Additionally, the integration of human skill and oversight remains important for verifying and explore flagged transactions.

Efficient fraud detection and prevention using machine learning and deep learning algorithms:

Efficient fraud detection and prevention using machine learning and deep learning algorithms involve employing effective strategies to detect fraudulent activities accurately and quickly, while minimizing false positives and reducing the impact on legitimate transactions.

Here are some key approaches to achieve efficient fraud detection and prevention:

1. **Data Quality and Feature Engineering:** High-quality data is crucial for building effective fraud detection models. Make sure that the data is clean, consistent, and contains pertinent features that can help distinguish between legal and false transactions. Feature engineering is the process of bringing out meaningful information from the data, and it plays a crucial role in improving model performance.
2. **Imbalanced Data Handling**: Financial fraud datasets often suffer from class disparity, where the number of false transactions is much lower than legitimate ones. Apply appropriate techniques like oversampling, under sampling, or using synthetic data generation methods (e.g., SMOTE) to address this imbalance and prevent partiality toward the majority class.
3. **Model Selection and Ensemble Methods**: Choose appropriate machine learning or deep learning algorithms for the fraud detection task. Ensemble methods, which combine the forecast of multiple models, can help improve accuracy and wholesomeness. Combining the strengths of different algorithms can lead to more effective fraud detection systems.
4. **Real-time Monitoring:** Implement real-time monitoring systems that continuously analyze incoming transactions in real-time. This allows for quick detection of suspicious activities and immediate action to prevent fraudulent transactions.
5. **Feature Importance Analysis:** Conduct feature importance analysis to identify the most pertinent features that contribute significantly to fraud detection. This information can be used to focus on evaluative aspects and clarify the model.
6. **Continuous Learning and Model Updates:** Fraud patterns develop over time, so it's necessary to continuously update and hold on to the models with new data. Implement automated processes to include new information and adapt to emerging fraud tactics.
7. **Interpretability**: While deep learning models can provide excellent accuracy, they are often considered "black boxes" due to their twists. Consider using interpretable machine learning models or techniques to gain perception into how the model makes decisions. This can be essential for regulatory adherence and building trust in the system.
8. **Rule-based Systems:** Incorporate rule-based systems in conjunction with machine learning and deep learning models. Rule-based systems can capture known fraud patterns and help quickly identify straightforward fraud cases.
9. **Fraud Network Analysis:** Explore network analysis techniques to detect fraud rings or collaboration among fraudsters. Detecting fraud based on relationships between entities involved can be a powerful approach.
10. **Human Oversight and Expertise:** While machine learning and deep learning algorithms are powerful tools, human skill is crucial in investigating flagged

ISBN: 978-81-965283-7-9

transactions and making final decisions. Employ a hybrid approach that combines the strengths of both automated systems and human analysts.

Efficient fraud detection and prevention require a combination of enlightened algorithms, continuous learning, high-quality data, and collaboration between technology and human expertise. By hiring these strategies, financial institutions and businesses can build strong and effective fraud detection systems to protect themselves and their customers from fraudulent activities.

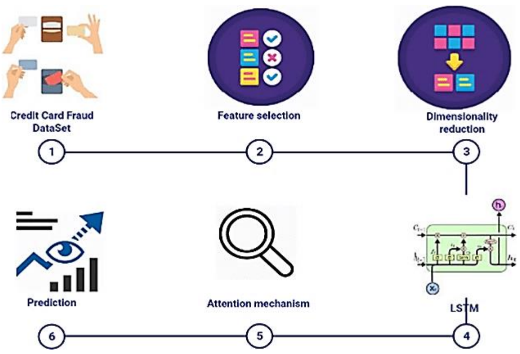


Fig.3. Prediction of Financial Fraud Detection

Optimizing machine learning and deep learning algorithms for financial fraud detection and prevention:

Optimizing machine learning and deep learning algorithms for financial fraud detection and prevention is necessary to achieve high accuracy, efficiency, and reliability in detecting false activities. Financial institutions and businesses can build highly successful and efficient fraud detection and prevention systems that can adapt to evolving fraud patterns and protect their benefit and customers from potential financial losses. The choice of machine learning and deep learning algorithms for financial fraud detection and prevention depends on various factors, including the nature of the data, the complexity of fraud patterns, computational resources, and the level of interpretability required.

Machine Learning Algorithms:

1. **Logistic Regression**: A simple and accountable algorithm that can provide the probability of a transaction being fraudulent. It's useful for getting a baseline performance and understanding the importance of features.
2. **Random Forest:** An ensemble method that combines multiple decision trees, offering high accuracy and wholesomeness. It can handle non-linear relationships in the data and deal with class imbalances effectively.
3. **Gradient Boosting:** Another ensemble method that builds multiple weak learners to create a strong classifier. It often exceeds random forests in terms of accuracy but might require more computational resources.
4. **Support Vector Machines (SVM):** Effective for separating data into different classes based on a hyperplane. SVM can handle complex decision boundaries and is useful for binary classification tasks like fraud detection.
5. **XGBoost:** An improved implementation of gradient boosting that offers high performance and scalability. It's widely used in various machine learning competitions and has proven effective for fraud detection.

Deep Learning Algorithms:

1. **Convolutional Neural Networks (CNNs):** CNNs are primarily used for image recognition tasks, but they can also be applied to analyze transaction data constituted as time series. They are useful when transaction data has a spatial or sequential aspect.
2. **Recurrent Neural Networks (RNNs):** RNNs are suitable for refining sequential data, making them suitable for fraud detection in transaction sequences. However, they might suffer from vanishing gradient problems.
3. **Long Short-Term Memory (LSTM) Networks:** A type of RNN that can handle long-term dependencies and address the vanishing gradient problem, making them effective for processing sequential data.
4. **Gated Recurrent Units (GRUs):** Similar to LSTMs, GRUs are a type of RNN that can handle sequential data effectively but with fewer parameters.

6. **Transformer-based Models:** Models like the Transformer architecture, particularly the Bidirectional Encoder Representations from Transformers (BERT) model, have been successfully applied to NLP tasks like sentiment analysis and can be used for fraud detection in linguistic data.

CONCLUSION:

It's important to note that fraud detection is an underway process, and machine learning models need to be regularly updated and refined to keep up with emerging fraud patterns and tactics. Additionally, integrating domain expertise and human analysts' intuition can enhance the overall fraud detection system's performance.

Moreover, the effectiveness of these algorithms depends on data quality, feature engineering, hyperparameter tuning, and the specific fraud detection problem at hand. In practice, a combination of machine learning and deep learning algorithms, along with ensemble methods, is often used to achieve higher accuracy and robustness. Regular updates and continuous learning are also crucial for modifying to emerging fraud patterns and ensuring the model's success over time.

REFERENCES:

1. **V. Almendra,** Finding the needle: a risk-based ranking of product listings at online auction sites for non-delivery fraud prediction, Expert Syst. Appl. (2013),
2. **Ll. Bermúdez et al,** A Bayesian dichotomous model with asymmetric link for fraud in insurance Insur. Math. Econ. (2008)
3. **Adadi, A., & Berrada, M.** (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, 6, 52138–52160.
4. **F. Pardo Tonic**: A deep Reinforcement Learning Library for Fast Prototyping and Benchmarking (2020).
5. **K. He, X. Zhang, S. Ren, J. Sun,** Deep residual learning for image recognition, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA (2016), pp. 770-778

ID: 46

HUMAN – COMPUTER INTERACTION (HCI) WITH GRAPHICS

Mr.PRAJAY PRASANTH
II BCA,
Sri Krishna Adithya College of Arts and Science,
Coimbatore, Tamil Nadu, India.

Mr.MAHENDRA PRASATH
II BCA,
Sri Krishna Adithya College of Arts and Science,
Coimbatore, Tamil Nadu, India.

I. INTRODUCTION:

Human-Computer Interaction (HCI) and Graphics have become inseparable components in the modern digital landscape. The convergence of these two fields has revolutionized the way humans interact with computers, resulting in intuitive interfaces, enhanced user experiences, and improved usability. This essay explores the dynamic relationship between HCI and Graphics, delving into their intertwined evolution, impact on user-centered design, challenges, and future prospects.

Evolution and Interplay:

The evolution of HCI and Graphics has been intertwined since the inception of graphical user interfaces (GUIs) in the 1970s. Early HCI research paved the way for graphical displays and user-friendly interfaces, marking a shift from command-line interactions to visually-driven experiences. Icons, buttons, and visual metaphors introduced by Graphics made digital interactions more intuitive and accessible, laying the foundation for modern user interfaces.

Impact on User-Centered Design:

HCI and Graphics share the common goal of placing the user at the center of design. Graphics contribute by providing visually appealing elements that enhance engagement and guide user actions. HCI principles guide the placement and behavior of graphical elements, ensuring usability and efficiency. The collaboration between these fields has led to the development of user-centered design methodologies, such as usability testing and user personas, which empower designers to create interfaces that cater to users' needs and preferences.

Usability and User Experience:

Graphics play a pivotal role in shaping user perceptions and experiences. Well-designed visual elements aid in cognitive processing, reducing the cognitive load on users and facilitating efficient decision-making. Effective use of colors, typography, and layout can convey hierarchy, establish brand identity, and evoke emotions, enhancing the overall user experience. HCI research contributes to the understanding of user behavior and cognitive processes, guiding the design of graphics for optimal usability.

Challenges and Considerations:

Despite the advancements, challenges persist in the integration of HCI and Graphics. Ensuring inclusivity and accessibility remains a concern, as visually impaired users may struggle to interact with graphical interfaces. Striking a balance between aesthetics and functionality can be challenging, and designers must consider cultural, age-related, and contextual factors that influence user interactions. Moreover, rapid technological advancements require HCI and Graphics practitioners to stay updated with emergingtrends and tools.

Future Prospects:

The future of HCI and Graphics holds exciting possibilities. Augmented Reality (AR) and Virtual Reality (VR) are expanding the scope of graphical interactions, creating immersive environments that redefine user experiences. Natural Language Processing (NLP) and AI-driven interfaces are further blurring the lines between HCI and Graphics, enabling more intuitive and conversational interactions. The evolution of HCI and Graphics will likely continue to be shaped by user needs, technological innovations, and interdisciplinary collaborations.

More about human-computer interaction and Graphics:

1. Usability: Ensuring that a system is easy to use and efficient, allowing users to accomplish their tasks with minimal effort.
2. User-Centred Design: Placing the user at the centre of the design process and involving them in the design and evaluation stages to create interfaces that meet their needs.
3. User Experience (UX) Design: Creating interfaces that provide positive and engaging experiences for users, taking into account factors like aesthetics, emotions, and overall satisfaction.
4. Interaction Design: Designing the way users interact with a system, considering factors like input methods (e.g., touch, voice), feedback, and responsiveness.
5. Accessibility: Designing interfaces that are usable by people with disabilities, ensuring equal access to information and functionality.
6. Cognitive Psychology: Understanding how users perceive, think, and make decisions when interacting with technology.

Graphics, in the context of HCI, refers to the visual aspects of user interfaces. Graphics play a crucial role in enhancing the user experience and communication. Some key points about graphics in HCI:

1. Visual Elements: Graphics include icons, buttons, images, typography, and layout, which aid in conveying information and guiding user actions.
2. Information Visualisation: Displaying complex data in a visual format makes it easier for users to comprehend and make informed decisions.
3. Visual Consistency: Maintaining a consistent visual design across an interface to provide a cohesive and recognisable user experience.
4. Visual Hierarchy: Organising content in a way that guides users' attention and helps them prioritise information.
5. Responsive Design: Adapting graphics and layout to different screen sizes and devices to ensure a seamless experience.
6. Animation: Using subtle animations for transitions and feedback can enhance user understanding and engagement.

In essence, HCI and Graphics collaborate to create interfaces that are intuitive, efficient, and visually appealing, resulting in satisfying user experiences with digital technologies.

What is Human-Computer Interaction (HCI)?

Human-computer interaction (HCI) is a multidisciplinary field of study focusing on the design of computer technology and, in particular, the interaction between humans (the users) and computers.

While initially concerned with computers, HCI has since expanded to cover almost all forms of information technology design.

ISBN: 978-81-965283-7-9

HCI Benefits:

- Gaining Market Share:

  People will intend to buy or use products with higher usability

  Eg: Google engine

- Improving productivity:

  By using HCI employees in a company perform their goes in a faster manner.

  Eg: Internet can increase employee's efficiency

- Lowering Support Cost:

  If a product or thing is not more usable for everything, we can call the customer support for the help.

  Eg: By reading the instructions or manually they can easily have done without customer request.

- Reducing Development Cost:

  Reducing or avoiding the unwanted features that the users don't want and if they are including or indulging the features it would create more annoying and inefficient.

  Eg: No other party apps are need when the one app is having full efficiency.

Principles of HCI:

General principles of HCI design:

- Discoverability
- Feedback
- Constraints
- Mapping
- Consistency
- Affordance
- Structure
- Simplicity
- Tolerance
- Equality
- Flexibility
- Perceptibility
- Ease
- Comfort
- Documentation

Principles of HCI Design:

Making our system / software easy to use and have the knowledge to learn about the usability applies to all the aspects conditions of the system.

Here the Principles of HCI design:

- Compatibility.
- Easy of learning.
- Memorability
- Simplicity
- Flexibility
- Responsiveness
- Protection
- Invisible technology
- Control
- WYSIINYG

What is Graphic Design?

Graphic design is a craft where professionals create visual content to communicate messages. By applying visual hierarchy and page layout techniques, designers use typography and pictures to meet users' specific needs and focus on the logic of displaying elements in interactive designs, to optimize the user experience.

Graphic design and the Human – Computer Interaction:

- Graphic design and the Human – computer iteration is an essential part of UX/UI designing. The both Graphic Design and the Human – computer is Inbounded in UI designing is a bloom for all of us.
- These graphics methods are developed for the clients of substantial e - commerce platforms with unlimited design in graphics.
- These graphics design method consists of the *Brand, HTML, WordPress, illustration* and building for the development of website and it can be also stored in "**drafts**".

The Graphics which populated:

In the earlier btext – based screen has two methods, they are Flat, one – dimensional appearance and the other stage they could have three dimensional appearances also and the control of those above screen is moved on and around. The information may have been materialized or vanished.

They are used in many fields; they are:

Radio button, Check button, List boxes and palettes.

ISBN: 978-81-965283-7-9

CONCLUSION:

Human-Computer Interaction and Graphics are symbiotic disciplines that have transformed the way humans interact with computers. Their intertwined evolution has led to user-centered design approaches, enhanced usability, and engaging user experiences. As technology advances, the collaboration between HCI and Graphics will continue to shape the future of digital Interactions, ushering in new paradigms and possibilities that redefine the boundaries of human-computer relationships.

ID: 47

A REVIEW PAPER ON PRIVACY AND CRYPTOLOGY

Mr.B.A.BABAPRADHANJAN
II B.sc AI & ML,
Sri Krishna Adithya College of Arts and Science,
Coimbatore, Tamil Nadu, India.

Dr.M.LALITHAMBIGAI
Assistant Professor,
Department of BCA
Sri Krishna Adithya College of Arts and Science,
Coimbatore, Tamil Nadu, India.

ABSTRACT:
In the digital age, privacy and cryptology are essential components in ensuring the security and confidentiality of personal data and communication. Privacy refers to the right of individuals to keep their sensitive information and activities confidential, protecting them from unauthorized access or disclosure. On the other hand, cryptology, particularly cryptography, offers the means to achieve this protection through the use of advanced mathematical algorithms and keys to encrypt data into an unreadable form, making it accessible only to authorized parties with the proper decryption keys. This abstract highlights the importance of safeguarding data privacy and the role of cryptology in providing secure communication and data protection in today's interconnected world.

KEYWORDS: *Privacy, Cryptology, Data Protection, Encryption, Digital Security.*

I. INTRODUCTION

Privacy and cryptology are two essential concepts that play a crucial role in our digital age, where information is constantly being shared, stored, and transmitted across various networks. The rapid advancements in technology have brought unprecedented Conveniences and opportunities, but they have also raised significant concerns about the protection of sensitive data and personal information.

Privacy is the right of individuals to keep their personal information and activities confidential and secure from unauthorized access or disclosure. In the digital realm, privacy has become a paramount issue as our online activities generate vast amounts of data that can be easily collected and analyzed by various entities, such as governments, corporations, and cybercriminals. Ensuring privacy is vital for maintaining autonomy, freedom of expression, and safeguarding against potential misuse of personal data.

Cryptology, on the other hand, is the science and practice of secure communication through the use of encryption and decryption techniques. It involves encoding information in such a way that only authorized parties can understand it while preventing unauthorized individuals from deciphering the content. Cryptography, a fundamental branch of cryptology, employs mathematical algorithms and keys to transform plaintext data into cipher text, making it incomprehensible to those without the necessary decryption keys.

The relationship between privacy and cryptology is intertwined. Cryptography provides the tools and methods to protect sensitive information and communication, ensuring that only intended recipients can access the content. It acts as the cornerstone of modern digital security, protecting personal data, financial transactions, confidential communications, and sensitive government information from prying eyes.

As technology continues to advance, so do the threats to privacy and the need for robust cryptology. Privacy breaches and cyber attacks are becoming more sophisticated, and governments and organizations are grappling with balancing

`The demands of security and convenience while preserving individual privacy rights. In this context, the development of cryptographic techniques is vital for staying ahead of malicious actors. Cryptologists and cyber security experts constantly strive to design stronger encryption algorithms and systems to thwart potential attacks and safeguard sensitive information.

In conclusion, privacy and cryptology are interlinked components of our digital age. Privacy is the right of individuals to protect their personal data, while cryptology provides the means to achieve this protection. As technology evolves, so too must our methods of ensuring privacy and security, making the study and implementation of cryptology crucial in safeguarding our digital world.

II. LITERATURE REVIEW

1. "Privacy in the Age of Cryptography" by S. Yu, X. Huang, and D. Evans (2019):

This paper explores the intersection of privacy and cryptography, highlighting how cryptographic techniques can be employed to protect user privacy in various applications and scenarios. The authors discuss advancements in privacy-preserving protocols and cryptographic primitives, offering insights into the challenges and potential solutions for achieving strong privacy guarantees in the digital age.

2. "Privacy-Preserving Cryptography: A Comprehensive Survey" by J. Camenisch and A. Lysyanskaya (2017):

This survey provides an in-depth overview of privacy-preserving cryptographic techniques and protocols. It covers various topics, including anonymous credentials, secure multiparty computation, private information retrieval, and privacy-preserving data mining. The paper examines the theoretical foundations and practical implementations of these cryptographic schemes to protect sensitive information while ensuring data utility.

3. "Privacy and Cryptography in Mobile Health Systems: A Survey" by Z. Zhang, Y. Xue, and J. Zhang (2020):

Focusing on the domain of mobile health systems, this survey delves into the challenges of preserving privacy in the context of healthcare applications. The authors discuss how cryptographic methods can be applied to ensure the confidentiality of medical data while allowing for efficient data sharing and analysis, emphasizing the importance of balancing privacy and usability in such systems.

4. "The Right to Privacy" by S. Warren and L. Brandeis (1890):

Although not a recent work, this landmark article remains highly relevant to the discourse on privacy and has significantly influenced the development of privacy law and discussions around it. The authors argue for the recognition of a right to privacy in response to emerging technologies, emphasizing the need for legal protections against intrusions into personal affairs and data.

5. "Post-Quantum Cryptography" by D. J. Bernstein, et al. (2015):

With the rise of quantum computing, this paper explores the impact on traditional cryptographic systems and the need for post-quantum cryptographic solutions. It discusses the vulnerabilities of current encryption algorithms to quantum attacks and examines potential post-quantum cryptographic schemes that can withstand quantum computing's computational power while maintaining data privacy and security. These literature pieces represent a fraction of the vast research available on privacy and cryptology. They highlight the ongoing efforts in academia and industry to address the evolving challenges of preserving privacy and developing robust cryptographic techniques to secure data and communication in the digital landscape. Researchers continue to explore new avenues and develop innovative solutions to ensure privacy and data protection in a rapidly changing technological environment.

III. OVERVIWE OF PRIVACY

1. **Online Privacy:** As more individuals share personal information online, it is essential to secure their data and protect them from cyber attacks.
2. **Offline Privacy:** Individuals who take steps to protect their privacy offline can better prevent identity theft and other physical crimes.
3. **Privacy Laws:** Privacy laws and other regulations play a significant role in protecting information and enforcing penalties for violations.

IV. CRYPTOLOGY CONCEPT

Cryptology is an essential field that revolves around securing information and communication in the digital world. It comprises two key aspects: cryptography and cryptanalysis. Cryptography involves devising algorithms and methods to transform data into unreadable formats, ensuring unauthorized parties cannot decipher it without proper decryption keys. It encompasses encryption techniques, digital signatures, and secure communication protocols, providing the foundation for secure data transmission and storage. On the other hand, cryptanalysis involves attempting to break encrypted messages without having access to the decryption keys. Cryptologists focus on creating robust encryption methods, while cryptanalysts assess the security of these systems by attempting to crack them. In today's digital age, cryptology plays a vital role in protecting sensitive data, maintaining privacy, and facilitating secure communication across diverse networks and platforms. Its continuous evolution is crucial in combating ever-evolving cyber threats and ensuring the confidentiality and integrity of digital information.

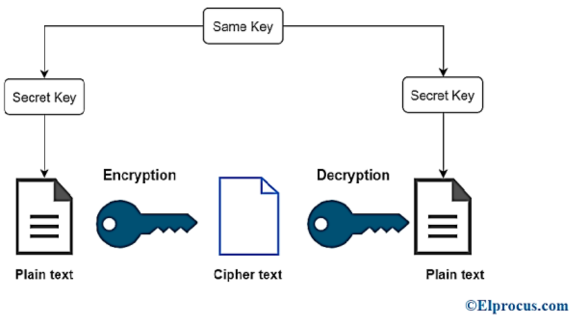


Fig.1: Cryptology Concept

V. HISTORY OF CRYPTOLOGY

Pre-Modern era: The history of cryptology can be traced back ancient civilizations, including the Egyptians, Babylonians, and Hebrews. Early examples include simple substitution ciphers and transposition techniques, which involved rearranging the order of letters in a message to obfuscate its meaning.

For example, The Caesar Cipher (1st century BCE):

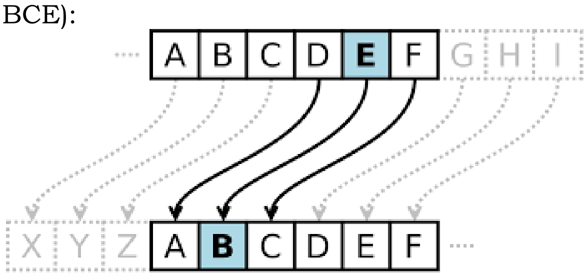


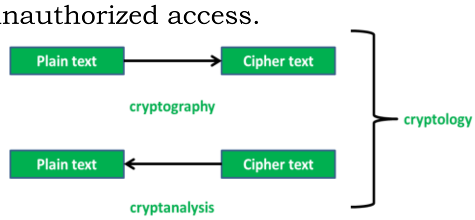
Fig.2: Encrypting with Caesar cipher

The Caesar cipher involves shifting each letter of the plaintext by a fixed number of positions down the alphabet. For example, if the fixed number is 3, the letter 'A' in the plaintext would be replaced by 'D'.

- **Plaintext:** HELLO
- **Ciphertext:** KHOOR

Fig.3: Caesar Cipher encryption wheel

Modern era: Today, cryptology is used to protect data over the internet, secure financial transactions, and safeguard intellectual property from theft. Modern cryptology is the study and practice of secure communication and data protection in the digital era. It encompasses both cryptography (encryption) and cryptanalysis (decryption), with a focus on developing advanced algorithms and protocols to safeguard sensitive information from unauthorized access.



Key Features of Modern Cryptology:

- Strong Encryption: Modern cryptology employs robust encryption algorithms, such as AES (Advanced Encryption Standard) and RSA (Rivest-Shamir-Adleman), ensuring secure transmission and storage of data.
- Public-Key Cryptography: The introduction of public-key cryptography allows for secure communication between parties who have never met before, using a pair of keys (public and private) for encryption and decryption.
- Digital Signatures: Cryptology enables the creation of digital signatures, verifying the authenticity and integrity of digital documents, messages, and transactions.
- Secure Protocols: Modern cryptology facilitates the development of secure communication protocols like

SSL/TLS (Secure Socket Layer/Transport Layer Security) to protect online data exchanges.
- Quantum-Resistant Cryptography: With the advent of quantum computing, cryptographers are exploring quantum-resistant algorithms to maintain data security against quantum attacks.
- Homomorphic Encryption: Emerging techniques like homomorphic encryption allow computations on encrypted data, preserving privacy during data analysis and processing.
- Cryptanalysis Advancements: Modern cryptanalysis employs advanced techniques and computational power to break weak encryption, enhancing security evaluation and threat mitigation.
- Application Diversity: Cryptology finds applications in various fields, including cybersecurity, finance, healthcare, government, and cloud computing.

As technology advances, modern cryptology continues to evolve, addressing emerging challenges and staying at the forefront of protecting data and communication in an increasingly interconnected world.

Ciphers and Encryption:

Ciphers and encryption are essential methods of protecting sensitive information from being intercepted and read by unauthorized persons.

Ciphers

1. Substitution Ciphers
2. Transposition Ciphers
3. One-Time Pads

Encryption

- Symmetric Key Encryption
- Asymmetric Key Encryption
- Hash Functions

TYPES OF CRYPTOGRAPHY:

Cryptography can be broken down into different categories depending on its purposes and mechanisms, including:

Public Key Cryptography	- Uses two separate keys to encrypt and decrypt messages
Symmetric Key Cryptography	- Uses the same key for both encryption and decryption
Blockchain Cryptography	- Uses complex algorithms to secure and verify blockchain transactions
Quantum Cryptography	- Uses quantum mechanics to create secure communication channels

Few more types are,
1. Hash function
2. Digital signatures
3. Homomorphic Encryption

VI. IMPORTANCE OF PRIVACY AND CRYPTOLOGY

"**Privacy is not something that I'm merely entitled to; it's an absolute prerequisite**."

-Marlon Brando

The importance of privacy and cryptology cannot be overstated in today's digital age. They play vital roles in ensuring data security, safeguarding individual rights, and maintaining trust in digital interactions. Here are some key reasons why privacy and cryptology are crucial:

1. Protection of Sensitive Information: Privacy is essential for safeguarding sensitive personal and financial information. Cryptology provides the tools to encrypt and secure data, making it inaccessible to unauthorized individuals and protecting against data breaches and identity theft.
2. Preserving Individual Rights: Privacy is a fundamental human right, enshrined in various international conventions and laws. It ensures individuals have control over their personal information, preventing unauthorized surveillance, and maintaining autonomy and freedom.
3. Confidentiality in Communication: Cryptology enables secure communication by encrypting messages and data during transmission. This confidentiality ensures that sensitive information, such as personal conversations, financial transactions, and medical records, remains private and is only accessible to authorized parties.
4. Trust in Digital Transactions: In an increasingly interconnected world, trust is vital for the smooth functioning of digital transactions and interactions. Cryptography provides the foundation for secure online transactions, e-

commerce, and digital signatures, instilling confidence in users and businesses.

5. National Security and Defense: Cryptology is instrumental in protecting sensitive government and military information from foreign adversaries. It helps secure classified data and communication channels, preventing espionage and ensuring national security.
6. Data Integrity: Cryptology, including hash functions and digital signatures, ensures the integrity of data and documents. It allows for the verification of data authenticity, preventing tampering and unauthorized changes.
7. Protection against Cyber Threats: In the face of ever-evolving cyber threats and hacking attempts, robust cryptology becomes crucial in defending against unauthorized access and attacks on sensitive information.
8. Advancements in Technology: As technology advances, the need for strong cryptology becomes more critical. Quantum computing and other emerging technologies pose new challenges and require innovative cryptographic solutions to maintain data security.
9. Compliance with Regulations: Privacy regulations, such as the European Union's General Data Protection Regulation (GDPR) and various data protection laws worldwide, mandate the protection of personal data. Proper cryptology implementation helps organizations comply with these regulations and avoid potential penalties.

In summary, privacy and cryptology are inseparable pillars of data security and personal rights in the digital era.

VII. FUTURE OF PRIVACY AND CRYPTOLOGY

The future of privacy and cryptology is uncertain, as technology continues to evolve at an unprecedented pace. New threats will arise, and new methods of protection will need to be developed.

- Advancements in artificial intelligence and quantum computing could revolutionize the field of cryptology and transform how we protect our priva
- Data privacy will become increasingly essential as more people rely on digital technologies to store and share personal information.

- Integration with IoT and AI: As the Internet of Things (IoT) and Artificial Intelligence (AI) technologies expand, cryptography will be essential to secure data transmission and protect sensitive information collected by interconnected devices and AI-driven applications.
- Focus on User-Centric Security: With increased public awareness of privacy concerns, there will be a greater emphasis on user-centric security approaches. Cryptography will be designed with usability and transparency in mind, empowering individuals to have more control over their data.

VIII. CONCLUSION

In conclusion, privacy and cryptology are inseparable components that play crucial roles in the modern digital age. Privacy is a fundamental human right that empowers individuals to control their personal information and protects them from unauthorized surveillance and data exploitation. Cryptology, on the other hand, provides the necessary tools and techniques to secure sensitive data, ensure confidential communication, and maintain data integrity.

The dynamic and interconnected nature of today's world demands robust cryptographic solutions to safeguard against cyber threats and protect sensitive information across various platforms and networks. As technology continues to evolve, the challenges to privacy and data security grow in parallel. Therefore, the continuous development and implementation of advanced cryptographic methods, including quantum-resistant algorithms, will be vital to stay ahead of potential attackers and preserve the privacy of individuals and organizations.

ISBN: 978-81-965283-7-9

ID: 48

SIGNIFICANT ASPECTS IN CLOUD DATABASE

Ms.V.VAISHNAVI
II - B.Sc Data Science,
Department of Data Science,
Sri Krishna Adithya College of Arts and Science,
Coimbatore, Tamil Nadu, India.

Ms.A.SURETHA
II -B.Sc Data Science,
Department of Data Science,
Sri Krishna Adithya College of Arts and Science,
Coimbatore, Tamil Nadu, India.

ABSTRACT:
Cloud databases are an uprising path to store and maintain data in cloud computing environment with a rapid growth of data and the need for the scalability and flexible solution, cloud database provides a minimum costs and better way to store access and analyze large volume of data. Cloud database provides a flexible, scalable and secure environment for storing and maintaining the data in the cloud. By using the power of cloud computing, IT sector can unlock the future of their data, driving prequel and development in digital era.

Keywords: *Storing, Maintaining, Analyzing, Accessing, Scalability, Flexible, Management, Cost-Effective, Reliable, Arrange, Control, Transform.*

I. INTRODUCTION

Cloud database technology has transformed the way of business to store, maintain and process their data. Traditional on-premises usually come with restrictions such as scalability upkeep cost and limited accessibility. On the other contra, cloud database provides a reliable, scalable minimum-cost solution by using the ability of cloud computing. These databases are introduced and maintained by cloud service providers, allowing businesses to concentrate on their main operations

While favor from the pros of cloud-based depository and management. The main project aims to survey the abilities and the benefits of cloud databases, analyze their use instance and grow a complete understanding of their execution and administration.

II.WHAT IS A CLOUD DATABASE?

A Cloud database is a database that is arranged, dispatched and approached [3] in the cloud. A cloud database is a systematize and it contains structured, unstructured and semi-structured data. They are similar to cloud computing, which includes speed, measurement, dexterity and reduce overheads. A cloud database can be performed as a managed database as a service or extended on the cloud based Virtual machine (VM).

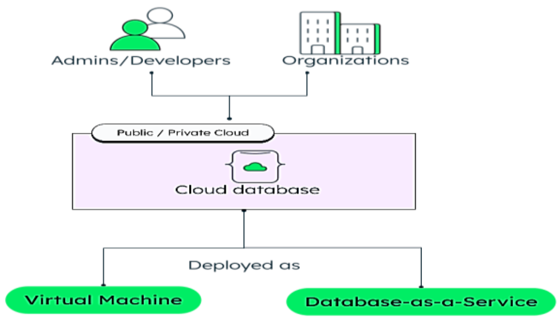


Figure 1: Cloud Database [5]

III. TYPES OF CLOUD DATABASE

Cloud database has two types namely, relational database and non-relational database.

- **Relational cloud database**

Relational cloud databases contains of one or more tuple and attribute and it allows you to systematized data[3] in well-defined relationship to acknowledge how the data is coherent. These databases are usually used to static data schema and you can utilize structured query language (SQL) to query and to control the data. They are extremely stable, dependable, and convenient to handle with enormous amounts of structured data.

An Instance of relational database involves: SQL Server, Oracle, MySQL, PostgreSQL , Spanner, and cloud SQL.

- **Non- relational cloud database:**

Non- relational cloud database store and control unstructured data such as email, text messages , archive and sensor data .They don't pursue a well-defined schema like a relational database and permit you to save and arrange information anyway. Non-relational database is otherwise known as "NOSQL", which stands for Not Only SQL [3]. An non-relational database stores data in non-tabular form and favor to be highly adaptable then the traditional, SQL-based, Relational Database Structure.

Instances of Non -relational database involves: MongoDB, Redis, Cassandra, Hbase, and Cloud bigtable**.**

IV. WHY DO WE USE A CLOUD DATABASE?

The capacity of data that's consumed in our today's world is very large. It is only different but also wildly disparate. Cloud database is reliable, assured, cost - effective, extensible, and so on. It gives a solid foundation for developing modern business applications and technology. In specific, they can quickly accommodate changing workload and claim without enlarging the workload of earlier overburdened group [2].

V. BENEFITS OF CLOUD DATABASE:

- **Scalability and flexibility:**

The cloud database can be measured horizontally or vertically to manage an increase in data traffic[1]. Horizontal scaling includes concatenate more servers to the database to manage high appeal, in other hand vertical scaling includes concatenating high assets to an existing server to increase its volume.

- **Optimized costs:**

Cloud database is used to reduce cost by rejecting the necessity for the costly hardware and infrastructure in the business. Because cloud users be in charge of the infrastructure and control the perpetuation of the database and it reduces need of employees.

- **High availability and reliability:**

Cloud user use load balancing and clone. Load balancing allocates appeal across multiple servers while clone assures that data is copied across multiple servers for superfluous. Cloud secure that the database is always accessible and can manage a high volume of requests.

- **Improved security:**

Cloud user a use data encryption and have the ability to control and secure the data which is stored in database. Data encryption assures that the data is secured when data is transmitted over the internet by the users.

VI. CLOUD DATABASE MANAGEMENT

Cloud database tag along with two deployment models: traditional self-managed and managed database as a server. Traditional self-managed cloud databases: a cloud database is upgraded and accommodate on virtual machine in a traditional deployment model. Using this model you can manage more control over your database.

Managing server database: The cloud database is entered as a service. It speeds on the cloud service user's physical infrastructure,[4] where user is in control for the most of the operational, handling and managing the database management tasks. It involves in automated provisioning, measuring, reliability and improve that need to keep your database statistically available and efficient.

Cloud data management is used to handle data across cloud platform, either with or alternative of on premises storage. The cloud is a helpful as data storage tier for backup. Using cloud data management resources can be easily purchased.

VII. HOW DO CLOUD DATABASE WORK?

In industries the databases are used to collect organize, store and send the data to executives and employees for operational and analytics applications. Typically cloud database gives the same data processing, administration and access ability as on premises ones. Existing on the premises databases usually can be transferred to the cloud along with the applications[4]. Alternatively of traditional software licenses, costing is based on use of system resource which can be provided on request as needed to meet processing work overburdened. Alternatively user can reserve database samples for an year to get discount cost on regular workloads with static capacity requirements.

VIII. ADVANTAGES OF WORKING WITH CLOUD DATABASES

In this busy and developing environment many industries are changing to cloud databases which has many advantages. Using the help of cloud database the information technology sector (IT) can become more agile and innovative. Because cloud database provide IT sector to set up confirm and examine new projects comfortably and fast without incurring a lot of costs. Cloud databases decrease the overall cost of managing and storing enlarge amount of data[2]. The capability to work with cloud databases enables IT sector to focus on other features of business rather than just working on maintaining the data in the ordered form and to assure it stays safe. Cloud databases provides many security aspects and have in place procedures that decrease the chances of error as well as stolen of data. This decreases the risk element associated with using and retrieve the data.

IX. CLOUD DATABASE TOOL

- **Amazon Web Services(AWS)**

Amazon provides a broad array of cloud database services which involves NOSQL as well as relational database[1]. Amazon RDS (Relational database services) that works on either Oracle, SQL or MySQL Server instance varies Amazon SimpleDB is firstly a schema less database that is meant to manage smaller workloads. Amazon provides additional data management services such as Redshift- a data warehouse and data pipeline- a data integrating service for easier data management.

- **System, Application and Products(SAP)**

SAP is big in offering industrial software, now offers a cloud database platform called High-Performance Analytic Appliance(HANA) for complementing the on premises database related tools of an enterprise. One of the main database platform called SAP HANA, which includes Sybase, and this tool is available in AWS Cloud.

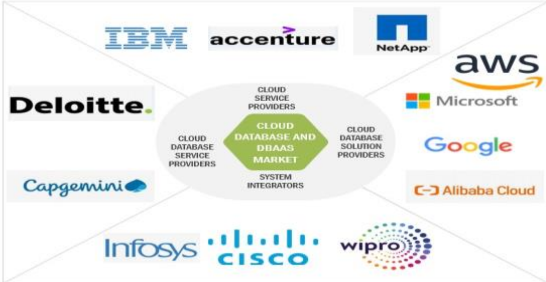


Figure 2: Tools in cloud database [6]

X. WHAT ARE THE BENEFITS OF MIGRATING TO A CLOUD DATABASE?

Cloud database have many advantages over traditional on site database. Using a cloud database, organization can provide raising data management needs without increasing infrastructure [2]. They can also control the data faster-efficiently dividing, delivering, and moving their data near to their users.

- **Traffic Speed:** Massive networks of mobile users and remote devices create giant volumes of engagement and application data. This can be measuring and accessibility management difficult and it consumes time for traditional database, as traditional databases need upgrades to work through a central "Master" database.
- **Easy axis:** Like any other asset on the cloud an user can access any data from anywhere using any device[4]. This simple connectivity sense data query more efficiently than traditional Local Area Network (LAN) access, which has only limited number of option.
- **Elasticity:** When the volume of data increases faster, a cloud database can extend capacity on a needed basis. This kind of scalability is not possible in a traditional database which needed observing and adjustment to be made by employee members.
- **Security:** With the help of cloud database you can gain a protective environment for a data because the

cloud servers are highly secured warehouse off site, separated from both internal and external users. In addition, the data itself is encrypted and making it hard for criminals to hack.

XI. BEST CLOUD DATABASE OF 2023:

Cloud computing has stimulated various technology innovation. It is easy to develop an app and host it with the help of cloud computing providers such as Microsoft Azure and Amazon Web Service (AWS). Before the services bound up, developers need to buy and maintain their own servers. All the cloud based apps are fueled by database software. These database helps the apps to collect data when needed and display it to the user[4]. There are n number of database developers can use for a cloud based app with each has its advantages and disadvantages.

A cloud database is created and accessed through a cloud platform. It provides the similar aspects as traditional database software but with more convenience and flexibility of cloud computing. It enables developers to host database software without buying costly hardware. The software could be relational database are as a NoSQL database and can be controlled through web interface or API.

XII. CONCLUSION

In the end, cloud database have arise as a transmutation technology that offers multiple benefits to the business. The capability to expedient on demand, easy accessibility and minimum cost of way models make cloud database appealing option for warehousing, maintaining and processing the data. The main project has provided an understanding into basic concepts of cloud database, their architecture and the numerous types available in the market. As business continue to enfold to digital transformation and search for scalable and flexible solution, cloud database place a vital role in designing the future of data warehousing and management.

REFERENCE

1. https://en.m.wikipedia.org/wiki/Cloud_database
2. https://www.simplilearn.com/cloud-databases-across-the-globe-article
3. https://cloud.google.com/learn/what-is-a-cloud-database#:~:text=A%20cloud%20database%20is%20a%20database%20built%20to%20run%20in,an%20in%2Dhouse%20IT%20team.
4. https://www.thoughtspot.com/data-trends/data-storage/cloud-databases
5. https://images.app.goo.gl/ark4ApYfRuCttL4H8
6. https://images.app.goo.gl/QQd7sLXtChLfBT2h8

ID: 49

A RECENT STUDY ON DATA SECURITY AND IT'S ELEMENTS

Ms.K.NIVETHA
II B.Sc Data Science,
Department of Data Science,
Sri Krishna Adithya College of Arts and Science,
Coimbatore, Tamil Nadu, India.

Ms.R.NITHYASRI
II B.Sc Data Science,
Department of Data Science,
Sri Krishna Adithya College of Arts and Science,
Coimbatore, Tamil Nadu, India.

ABSTRACT:
Data security is very important in our day today life, data security secure our personal information from unauthorized access. In data security availability, confidentiality, integrity is some of the important element. Nowadays data are widely misused in order to prevent that data security is used, data security methods kept on updating because of the online frauds being increased. Modern problem requires modern solution so; in future we may face even more data security issues to overcome all that we need updated data security tools.

KEYWORDS: *Protection, Data breach, encryption, Data privacy, Availability, Confidentiality, Integrity, Firewall and Deter.*

INTRODUCTION:

Data security is constructed to safeguard private data (personal information). It secures our data from unauthorized users like cyber-attack, data breach. Data breach is unauthorized cracking of data for copying and reading the information. The taken data may store several sensitive information such as unique identifiers like passport, adhaar number, biometric data and Personally Identifiable Financial Information (PIFI).data security is used in countless places mainly it used in our mobile phones to build a strong security.

Fig.1

DATA SECURITY

HISTROY OF DATA SECURITY:

Data security became an important element in the 1980's, that the time where computer clubs were started, as did malware.

Hackers and Viruses were generating a confusion disturbance in 1990's and in the 1990's modern version of data security came into the computer fields. [2] It was created to stop unauthorized entry into our computer network and also caution and memos were announced to computer staff to detect viruses. Encryption and passwords have become famous in terms of blocking hackers.

DATA SECURITY TRENDS:

- ***Generate data protection regulation:***

It is a kind of data protection that is used among EU citizens. It is not widely implemented yet. If things work out this protection system can also be used in other nations. If this trend has proved its importance, it will be used widely around nations.

- ***Artificial intelligence and machine learning:***

They can detect hacking and fewer security websites way before we could find. we cannot able to find those frauds and system errors as quickly as artificial intelligence do. They can detect any security less area and it let us know about the less security website, at the second we enter into the website but they cannot be right all the times, still this protection has something to work on.

- ***Ransomware:***

Ransomware is a type of malware that encrypts personal information and documents and they demand some amount to give access to their own documents. This ransomware may occur because of clicking some unknown links or an unknown email send to them or accepting a condition while entering into a website, that's how the victims' files are hacked. even after paying their demanded amount there is no grantee that the hackers will release our files. how we can prevent our files from ransomware? We can prevent our files from ransomware by using a physical backup on a hard drive, keeping the system updated to latest security measures, having anti-virus software etc.

- ***Patch management:***

Patching is an operation system updates for security purposes. Due to the increase in hacking, it is important to use patching as our documents. Most software companies often releasing patches. For example: Microsoft, releasing patches have three main

a) For providing security.
b) To fix bugs, to improve the system performance.
c) To introduce new features.

ELEMENTS OF DATA SECURITY:

- **Encryption:**

Commonly an encryption task is to transform a plain text into cipher text. encryption gives an extra layer of protection for the data. it converts the normal readable text into an unreadable cipher text format and it can't be readed without a decryption key

- **AES** – Advanced encryption standard is a kind of encryption that is utilized to protect the data transfers in online.

- ***Data privacy:***

The major focus of data privacy is to secure our sensitive information that we have in our devices as same as the company data even the companies have more sensitive data that has to be secured and maintained privately and even if companies fetch our information for some purpose, our information has to be maintained privately by the company and it's their work to keep them safe and secure.

- ***Availability:***

When the data can be readily used for any purpose safely is known as availability. for example: Adhaar card information widely used in most of business, private and government activities by just entering our adhaar number all our adhaar details will be retrieved and can be used for these purpose.

- ***Confidentiality:***

Confidentiality measures that only authorized persons can access a particular data. for example: The password we use and sometimes is recognition and finger print, we often use these to keep our data safe.

- ***Integrity:***

Maintaining the original form of the data being shared or submitted it ensures that the data being shared is not modified or altered. It remains the same before and after submission and it also ensures that there is no 3rd party intervention in this. It is secured and it cannot be changed by any 3rd party it remains as it is.

TYPES OF DATA SECURITY TECHNOLOGY:

Data security technologies come in variety of form, its main aim is to protect data.

- ***Firewall:***

More informative data is secured by firewall because firewall is basic layer in data security. Commonly firewall secure the data from fraud detector and ethical hackers. Each firewall is the layer in the system, it was make the system secured.

- ***Data masking:***

It is process of the wrapping the data in the sender to receiver. Receiver only can understand the data is wrapping by other fake data. it can help to secured the data from third party.

- ***Authentication:***

It is used in more application, it helps to show the show the some informative data in only sender and receive. Because it is process of the two factor authentication. Two factor authentication is only who know the pin or password that can saw the data from the sender.

- ***Hardware based security:***

It is based on security that involves the physical protection in the system because most of the IT layer and company's protection in build the silicon.

- ***Data erasure:***

Data erasure is an effective data security management technique that removes the data in that system. The deleted data will not be recoverable.

5 D's OF DATA SECURITY:

- ***Deter:***

It stops any potential intruders not to do something, and it stops the process before they even attempt it for example: if we are entering into a website, notification warns as to not enter into the website for some security reasons it is known as deter process.

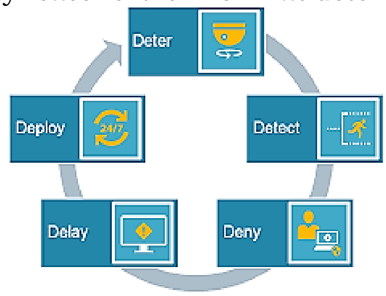


Fig.2

5 D's OF DATA SECURITY

- ***Detect:***

This process is used to scan that is there any malware or security issues are raised in data. It is used to find if any functions were performing excess other than its own process. for example: installation of CCTV cameras in house for security reasons is a form of detect, same as if we have installed any security app in our device it will help us to detect malware.

- ***Deny:***

It is the process of denying access to your data or website. It makes the data more secure and prevent our data from unauthorized users for example: using password. If someone else is entering into your email id by using your username and password, you will get notification on regard of that if you deny it's not you they cannot enter into and access your email.

- ***Delay:***

This is one of the layer, which protects the data by delaying the access to unauthorized users. It delays the time of the intruder and gives us more time to deal with security issue which is caused by the intruder.

- ***Defend:***

This is the final layer of the security.it happens when the intruder breaks all the layer the security and only this layer can help us to protect our data. This is the process of reaching to a technical team who can help us with this or to file a proper QQcomplaint and get help from the police cyber security team.

ADVANTAGES OF DATA SECURITY:

1. Protect all our personal information (data) for example: some information is very sensitive and those data have to be protected at any cost, [4] military information is one among them as military database hold all the information about military activities if that information got leaked it may affect the country's security so this sensitive information have to be protected.
2. Crucial for your fame for example: online shopping like flipkart, amazon etc. uses our location information and our personal information to order a product and for their delivery we give those information because of the trust we have with the brand and we believe that our personal information will be protected.
3. Show outfoxing in marketing: maintaining confidential information from a unlawful source by doing this, [4] the company will be on the top priority list of the customer because the company protect the personal information of the customer and make them to feel that their information are secured.

CONCLUSION:

In conclusion, Data security is an important element to protect our data and there are professionals to help as, if we were in trouble with our data security. Sensitive information that we share should protected by the company or the government entertprise, we share our details with they are done by data security.

REFERENCE:

1. https://www.loginradius.com/blog/identity/data-security-best-practices/
2. www.dataversity.net
3. https://images.app.goo.gl/uNTpx1BXwU1nbv2d8
4. https://www.digitalguardian.com/blog/what-data-security#:~:text=BENEFITS%20OF%20DATA%20SECURITY&text=Safeguards%20all%20valuable%20information%3A%20Sensitive,meant%20for%20every%20prying%20eye

ISBN: 978-81-965283-7-9

ID: 50

DATA MINING TECHNIQUES AND ITS APPLICATIONS

Mrs.R.GOMATHI
Assistant Professor,
Department of Computer Science,
KGiSL Institute of Information Management,
Saravanampatti, Coimbatore, Tamil Nadu, India.

Dr.R.KARTHIK
Assistant Professor,
Department of Computer Applications,
Sri Krishna Adithya College of Arts and Science,
Kovaipudur, Coimbatore, Tamil Nadu, India.

ABSTRACT:
Data Mining is the process used to store and retrieve the refined data from large datasets for solving business problems. It is used to understand the uncovering pattern from large dataset and to make predictions accurately. Data mining has various techniques like Machine Learning, Statistics, Information Retrieval, Data Warehouse, Pattern Recognition, algorithms and high performance computing. Data mining techniques are used to produce better result for business process logically and by decision making. Clustering and Classification are used to group and extract the data from preprocessor. Knowledge Discovery is the process to perform knowledge mining and knowledge extraction from dataset.

KEYWORDS: *Data Mining, Dataset, Patter Recognition, Knowledge Discovery, clustering and Classification.*

I. INTRODUCTION

Digital mining has emerged as a new research hotspot in the wake of the Internet, particularly high-dimensional, large-scale, distributed digital mining, which has limitless potential for economic value. Among them, classification prediction technology will support future smart business activities and provide crucial reference decisions. Scholars domestically and abroad have offered several references about data mining technologies and economic intelligence systems. Li and Long investigated data mining-based quantitative detection analysis for gastrointestinal disorders and image identification [1].

A computer data mining module was created by Zuo after extensive investigation and analysis of network virus characteristics. In order to uncover secret data and establish whether a virus exists, he integrated the technology for data mining with that for dynamic behavior interception. To identify Trojan viruses on networks, he used this technique [2].

Huge databases and data have been produced in many different fields as a result of information technology advancement. A method to preserve and handle this valuable data for future decision-making has been developed thanks to research in databases and information technology. The practice of extracting usable information and patterns from vast amounts of data is called data mining. It is also known as knowledge extraction, knowledge mining from data, knowledge discovery process, or data/pattern analysis [3].

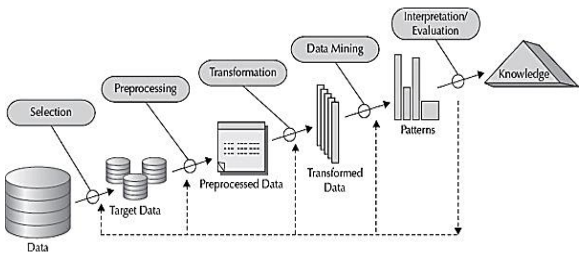


Fig-1 Knowledge Discovery Process in Data Mining.

In order to uncover meaningful data from a big volume of data, a method known as data mining is utilized. This method seeks to identify previously undiscovered patterns. Once these patterns have been identified, they may then be utilized to guide specific business growth decisions.

The process comprises three steps:

1. Exploration
2. Identification of patterns
3. Deployment.

Exploration: Data exploration begins with data cleaning and transformation into a different form, followed by the identification of key factors and the nature of the data in relation to the issue.

Pattern Identification: The next stage is to identify patterns in the data once it has been examined, clarified, and specified for the particular variables. Find and choose the patterns that are most predictive.

Deployment: Patterns are used to get the desired result.

DATA MINING TECHNIQUES

1. Classification:

Data records are classified using classification algorithms into one of a number of predetermined classes. They operate by building a model using a training dataset made up of sample records with predetermined class labels. Classification is a technique for supervised learning [4]. The categorization of data involves two steps. A model is created in the first stage by examining data tuples from training data that have a certain set of properties. The class label attribute value for each tuple in the training data is known. The model may be used to categorize the unknown tuples if its accuracy is deemed satisfactory [5]. Applications for credit risk and fraud detection are especially well suited to this kind of investigation. This method usually uses classification algorithms based on neural networks or decision trees. Learning and classification are both involved in the data classification process. The training data are examined by a classification algorithm in learning. Data from classification tests are used to gauge how accurate the rules are. The rules can be applied to the new data tuples if the accuracy is deemed acceptable. This would comprise comprehensive records of both valid and fraudulent actions identified record-by-record for a fraud detection application. These pre-classified examples are used by the classifier-training method to establish the parameters necessary for accurate discrimination. The classifier model that the algorithm creates after encoding these parameters.

Models for categorization fall under the following categories:

1. Classification based on decision tree induction
2. Classification based on associations
3. Neural networks
4. Support vector machines (SVM)
5. Bayesian classification.

2. Clustering

Identification of comparable classes of objects is referred to as clustering. We can further define dense and sparse regions in object space and learn about the general distribution pattern and relationships between data attributes by employing clustering techniques. The classification strategy can also be used to distinguish between groups or classes of objects, but it is more expensive, therefore clustering can be used as a preprocessing method before choosing an attribute subset and classifying it. To categorize clients based on their purchase habits, for instance, or to classify genes with comparable functions.

There are different Clustering Techniques:

1. Partitioning Strategies
2. Hierarchical aggregative (divisive) techniques
3. Methods based on density
4. Grid-based procedures
5. Model-based approaches

3. Prediction

Prediction can be accomplished using the regression technique. The relationship between one or more independent variables and dependent variables can be modeled using regression analysis. Independent variables in data mining are characteristics that are previously known, whereas response variables are what we wish to forecast. Unfortunately, not all real-world issues can be predicted. Sales volumes, stock prices, and product failure rates, for instance, are all exceedingly challenging to forecast because they may be influenced by intricate interactions among numerous predictor factors. Therefore, it may be essential to estimate future values using more sophisticated techniques (such as logistic regression, decision trees, or neural networks). Frequently, the same model types can be applied to classification and regression. For instance, classification trees (used to categorize categorical response variables) and regression trees (used to forecast continuous response variables) can both be created using the CART (Classification and Regression Trees) decision tree technique [6]. Similarly to neural networks, classification and regression models can be produced by them.

There are various Regression Techniques:

1. Inverse Linear Regression
2. Nonlinear Regression
3. Multivariate Linear Regression
4. Regression using Multivariate Nonlinearity

4. Association rule

In large data sets, frequent item set findings are typically found using association and correlation. This kind of information aids firms in making choices regarding catalog design, cross-marketing, and customer buying behavior analysis, among other things. The ability to produce rules with confidence levels under one is a requirement for association rule algorithms. However, the number of Association Rules that can be applied to a given dataset is typically very enormous, and the majority of these rules are typically of little (if any) significance.

Types of Associations Govern:

1. Multilevel association rule
2. Multidimensional association rule
3. Qualitative association rule

5. Neural Networks

Each connection in a neural network has a weight associated with it. It is made up of connected input/output units. The network adjusts its weights as it learns in order to anticipate the right class labels for the input tuples. Neural networks have the extraordinary capacity to extract patterns and identify trends from complex or ambiguous data that are too subtle to be seen by humans or other computer techniques. These are ideal for inputs and outputs with continuous values. For instance, rearranging handwritten characters, teaching a computer to comprehend English text, and many other real-world business issues have all been effectively implemented in numerous industries. For prediction or forecasting needs, neural networks are excellent at spotting patterns or trends in data.

Neural Network Type:

- Back Propagation

Data Mining Applications

Technologies for data mining can be used to make a wide range of decisions in different corporate environments [7]. Because data mining technologies provide quick access to vast amounts of data as well as important information from that data, many different sectors have adopted them. The following is a list of a few of the primary applications:

A. Data mining in Science and Engineering

In fields of research and engineering like bioinformatics, genetics, medicine, education, and electrical power engineering, data mining is frequently employed. Data mining is considered an interdisciplinary practice because of this. Understanding the relationship between inter-individual variation in human DNA sequences and variance in disease susceptibility is a key objective in the field of research on human genetics. It is particularly beneficial in the detection, prevention, and treatment of diseases.

B. Data mining in Banking and Finance

In the banking and financial sectors, data mining has been extensively employed. Data mining is used in the banking industry to forecast credit card fraud, gauge risk, examine trends, and determines profitability. To aid in the identification of credit card fraud, a number of data mining approaches, including distributed data mining, have been studied, modeled, and developed. Data mining allows banks to establish stock trading principles from past market data and uncover hidden relationships between various financial indicators.

C. Data mining in Sales and Marketing

Data mining is frequently used in the marketing industry to analyze consumer behavior based on their purchasing habits, such as identifying products that are bought at the same time. Additionally, data mining enables companies to choose marketing tactics like advertising, warehouse location, etc. Finding customer and product segmentation is the ultimate purpose of market analysis, which enables companies to promote their most lucrative items and increase profit. By positioning these goods next to one another and increasing their visibility and accessibility to clients while they are shopping, the retailers may make use of this information [8].

D. Data mining in earthquake prediction

The earthquake is predicted via data mining using satellite images. An earthquake is a quick shift of the Earth's crust brought on by the abrupt release of stress that has built up along an inner geologic fault. Forecasts (made months to years in advance) and short-term predictions (made hours or days in advance) are the two main kinds of earthquake predictions [9].

E. Data mining in Telecommunication

Because the telecoms sector has a vast amount of data, a sizable client base, a dynamic business climate, and fierce competition, it uses data mining technology. In the telecommunications sector, data mining aids in pattern recognition, fraud detection, efficient resource management, and service quality enhancement.

F. Data mining in Agriculture

Data mining is being used in the field of agriculture to analyze crop yields in relation to four factors: year, rainfall, production, and sowing area. Based on the information now available, yield prediction is a significant agricultural challenge that has to be solved. Data mining approaches like K Means, K Nearest Neighbor (KNN), Artificial Neural Networks, and support vector machines (SVM) can be used to tackle the yield prediction problem.

G. Data mining in Cloud Computing

Data Mining techniques are used in cloud computing. The implementation of data mining techniques through Cloud computing will allow the users to retrieve meaningful information from virtually integrated data warehouse that reduces the costs of infrastructure and storage .Cloud computing uses the Internet services that rely on clouds of servers to handle tasks The data mining technique in Cloud Computing to perform efficient, reliable and secure services for their users.

H. Data mining in Retail Industry

Data mining in retail industry helps in identifying customer buying patterns and trends that lead to improved quality of customer service and good customer retention and satisfaction.

I. Data mining in Bio Informatics

Data Mining ideally suited for Bioinformatics, since it is data-rich. Mining biological data helps to extract useful knowledge from massive datasets gathered in biology, and in other related life sciences areas such as medicine and neuroscience. Applications of data mining to bioinformatics include gene finding, protein function inference, disease diagnosis, disease prognosis, disease treatment optimization, protein and gene interaction network reconstruction, data cleansing, and protein sub-cellular location prediction.

J. Data mining in Corporate Surveillance

Corporate surveillance is the act of a corporation keeping an eye on someone or some group of people. The information gathered is frequently sold to other businesses or used for marketing purposes, but it is also frequently shared with government organizations. The company might use it to make their items more appealing to their target market. The information may be utilized for direct marketing initiatives, such as the tailored adverts on Google and Yahoo that are shown to search engine users based on their email and search history [10].

CONCLUSION:

In-depth descriptions of data mining methods and their uses in various industries are provided in this study. Classification, clustering, and other data mining techniques aid in identifying patterns that can be used to predict future business trends. Different Data Mining methods can be employed for various objectives. Every method has advantages and disadvantages. Future research will focus on different clustering algorithms and their significance.

REFERENCES:

1. T. Li and L. Long, “Imaging examination and quantitative detection and analysis of gastrointestinal diseases based on data mining technology,” *Journal of Medical Systems*, vol. 44, no. 1, pp. 1–15, 2020. View at: Publisher Site | Google Scholar
2. C. Zuo, “Defense of computer network viruses based on data mining technology,” *International Journal on Network Security*, vol. 20, no. 4, pp. 805–810, 2018. View at: Publisher Site | Google Scholar
3. A. Buczak and E. Guven, “A survey of data mining and machine learning methods for cyber security intrusion

detection," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2017. View at: Publisher Site | Google Scholar

4. Jiawei Han and Micheline Kamber (2006), "Data Mining Concepts and Techniques", published by Morgan Kauffman, 2nd ed
5. Dr. Gary Parker, vol 7, 2004, "Data Mining: Modules in emerging fields", CD-ROM.
6. W.-K. Chen, Linear Networks and Systems (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.
7. Ramageri," Data Mining Techniques and Applications", Indian Journal of Computer Science and Engineering Vol. 1 No. 4 301-305
8. Yudho Giri Sucahyo, Ph. D, CISA: "Introduction to Data Mining and Business Intellegence".
9. Kabra. R, Bichkar. R, "Performance Prediction of Engineering Students using Decision Tree", International Journal of computer Applications, December, 2011
10. Customer Successes in your industry from http://www.spss.com/success/?source=homepage&hpzone=nav_bar.

ISBN: 978-81-965283-7-9

ID: 51

THE STUDY OF DATA PRIVACY (CRYPTOLOGY) AND IT'S TECHNIQUES

Ms.S.SWETHA
II - B. Sc Data Science,
Department of Data Science,
Sri Krishna Adithya College of Arts and Science,
Coimbatore, Tamil Nadu, India.

Ms.D.RINEETHA GRACE
II - B. Sc Data Science,
Department of Data Science,
Sri Krishna Adithya College of Arts and Science,
Coimbatore, Tamil Nadu, India.

ABSTRACT:
The whole concept is about how to keep our datum secured from the intrusion of outsiders. The stored personal data can be kept safe with the study of cryptology. Cryptology is nothing but the science that involves encrypting and decrypting which prevents the data from unauthorized access. Cryptology involves the coverage of some texts in the background. Encryption takes place on the sender's side and has its own benefits and decryption takes place on the receiver's side.

KEYWORDS: *Privacy-definition, Origin word "Privatus"-Latin, History of privacy, Issues of privacy-THT, Study of cryptology-encryption and decryption, Five benefits of encryption and its drawbacks, Pros and cons of decryption.*

I. INTRODUCTION:

In today's sophisticated world, privacy is something which is considered to be very important by an individual. To maintain a private data, there are some few techniques which will be of some help to us. Among some of these techniques one of them is going to be discussed here.

Firstly, what is privacy?

If an individual is willing to do something secretly or without the knowledge of others and trying to maintain it for themselves or doesn't want anyone to involve with it is known as privacy.

Fig.1 Data Privacy

The domain of privacy partly imbricates with security, which can involve the notions of right use and safety of intelligence. The privacy may possess the form of inviolability of a person.

Etymology:

The word privacy has been derived from the Latin word "Privatus", [1] meaning to be away from what is civic, subjective and belonging to an individual.

History:

- ❖ In the old English, privacy is known as secrecy, mystery. Privacy for today's people is an advanced technology. But it has emerged in 1890's. Firstly people were only concentrating in protecting their own land and lives. Lately, as days passed, they started to impinge on physical protection and considered the safety of information and its privacy.[1] In 1819, the American lawyers Samuel Warren and Louis Branders gave a brief definition about the right to privacy in a famous article. It is right to be let alone.
- ❖ In the before 90's, this privacy was used to maintain the secrecy of the information in armies. They were used

to cover some politic or national information. Later as the technology began to grow, even the privacy term became very advanced. They are used to keep safe a country's database from the intrusive thoughts of others. Rather than a country's, today it is subjected to every individual and they are indubitably in need of it.

❖ With the invasion of technology, the argument considering privacy terms has elongated from a corporal sense to include a cybernated sense.

Do we all need privacy?

Yes, in today's situation privacy is a salient feature for every person. Because, in the era each and every individual starting from a 5year old to a 75year old are subtended to have a mobile in their hands. Using these mobiles, they do a lot of all activities like clicking pictures, bank transaction via online and so on. Since a lot of activities is being done through mobiles, our personal datum would also survive in these mobiles. Hence, only if we keep our data private or secure our data from the invaders, we can have a safe electronic life. Unless we secure our data, our personal information's would be exposed without getting the permission of the subject holders and this is known as hacking. But today, it should he termed s ethical hacking which is certified by the government of India.

PRIVACY ISSUES:

The main three issues in privacy terms are Tracking, Hacking and Trading which can be represented as "THT".[2]

Tracking:

It is like finding the location of the information where it is stored.

Hacking:

It is tacking the information from the stored location in simple terms.

Trading:

It is selling the information with some payment for it.

These cybercrimes, especially hacking can never be stopped fully, instead we can only take some protection actions which is also known as cyber security.

Techniques used to protect data:

A let of techniques are involved in the protection of data. Among them let us have a look about encryption under the study of cryptography. Cryptography is a technique used to secure information. Where, cryptology is the study of secure knowledge.

What is cryptology?

Cryptology is the study of how to secure information's. It is also known as the study of encryption and decryption.[3] It is the process of transformation of normal text to cipher text and the opposite.

Major aspects of Cryptology:

❖ This cryptology is effective on both the sender and the recipient side.
❖ Here, both sender and receiver can share the relevant information's with each other.
❖ It secures the information or data via theoretical and mathematical aspects.
❖ Cryptology also involves the domains of codes, ciphers and cryptanalysis.
❖ In cryptology it is important to know about those domains because it will be helpful in encrypting and decrypting the message sent between the sender and the receiver.

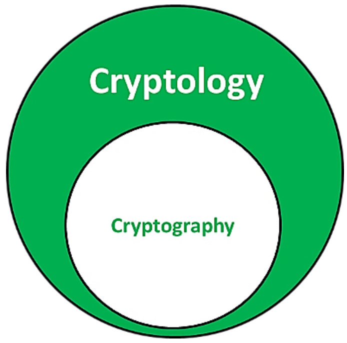


Fig.2 Cryptology

What is encryption and decryption?

Encryption:

It is the method of converting an important information into a coded meaningless information.[4] It takes place at the sender's side. In other words, it trans forms a plain message to cipher message. Any information can be encrypted with key called a secret or public key. It is used to protect the secrecy of the information. The encrypted message is unknowledgeable to anyone who doesn't have the decryption key.

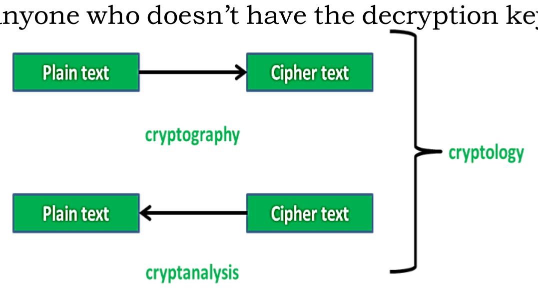


Fig.3 Overview of Encryption and Decryption

Assets of using encryption:

There are some few reasons of encryption for some corporates.[5]

1) Encryption is inexpensive to execute:

In today's world, each and every gadget is assisted with some kind encryption technology. It acts as an automated privacy set.

For example: Microsoft Windows is encrypted with a program called Bit locker which encrypts the whole volumes of hard disk.

2) Encryption can provide total data protection:

- When it comes to business datum, encryption provides a high security to the data stored.
- In business there might be a lot of competitors, who may try to intrude our data. There encryption is very supportive.

3) Encryption can support remote workers:

- A lot of employees are not static in their position. So, when they are moving to different places, their kind of data may require a lot of safety. In order to secure the data this encryption technology is very much helpful for remote workers.

4) Encryption enhances the integrity of data:

- Encryption data does not fully guarantee the truthfulness of data at rest, as data is dynamic, it can be used to verify the integrity of our backups.

5) Encryption can expand consumer trust:

- As said earlier, encryption is not a mandatory requirement, but some companies uses encryption to gain their customers belief.
- On studying a recent survey, we see that about "53% of the customer replies were around online security now than the previous years".

Drawbacks of encryption:

- Encryption provides a full security for the data which sometime makes hard for the employer to access the data.
- Managing the keys is a big task in encryption. The keys have to be given to only the decoder.

Decryption:

- It is the process of transforming the meaningless information or the encrypted information to the original message. It takes place on the receiver's side. Any information can be decrypted with the help of a key named secret or private key[4]. Decryption can only be performed by the person who have the decryption key. Decryption is the reverse process of encryption (i.e.,) it converts cipher message to a plain message.

Advantages of decryption:

- It's one of the main advantages is that it provides a high rate of security.
- It helps in handling the business issues like a piece of cake.
- This is what is called as cyber security.

Disadvantages of decryption:

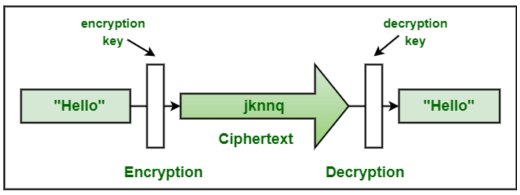
- It's main dis-advantage is that the decryption key is only given to the recipient of the data. This may become hard sometimes for remembering the key.

Threats:

- In between the encryption and decryption process a lot of hackers try to hack the data to get the needed message. This type of hacking here, is known as "man in middle attack".

Examples:

- These encryption and decryption of message are most commonly done in exchanging the army related information's.
- Another example can be taken as, an encrypted message is given to the manager by the higher official and the manager is decrypting the message via codes.



Cryptography

Fig.4 Working Flow of Encryption and Decryption

CONCLUSION:

Therefore, this encryption and decryption together forms cryptology which is used to study about and how to provide security for the data and some ways of how to prevent the data from hackers. So, finally the conclusion is the of protection of data is privacy and the studies ways to protect the data is cryptology which involves encryption and decryption.

REFERENCES:

1. https://en.m.wikipedia.org/wiki/privacy -Privacy.
2. https://www.globalbrandsmagazine.com/internet-privacy-issues-tracking-hacking-and-trading/#:~:text=but%20the%20%E2%80%9ctop%203%E2%80%9D%20privacy,how%20it%20impacts%20youra23%20privacy.-Internet privacy issues: THT.
3. https://www.techtarget.com/searchsecurity/definition/cryptology - Cryptology means.
4. www.geeksforgeeks.org Encryption and decryption.
5. www.lepide.com, 5 benefits of using encryption technology for data protection by Philip Robinson.

ID: 52

ADVANCE PREDICTION USINGHUPM ALGORITHM

Dr.K.HEMA SHANKARI
Associate Professor,
Department of Computer Applications,
Women's Christian College,
Chennai, Tamil Nadu, India.

Dr.KONDURI SUCHARITHA
Associate Professor,
Department of Computer Applications,
Women's Christian College,
Chennai, Tamil Nadu, India.

ABSTRACT:
HUPM has been researched successfully in the field of processing of data. The Traditional Pattern Mining cannot think fully of the features of the databases that have been used in the real world. In contrast, in many applications industries and interaction with details on internet services, the database sizes are increasingly closer to retail sales data by little bit. This is a general approach that is not adequate for static databases to prepare complex datasets and retrieve useful data. Most incremental utility pattern mining strategies have been developed. It has been proposed that previous methodologies, independent of the use of any framework, require more scanning for incremental utility sequence mining. Guy. In either case, there are proposed for different scans. It's just not enough for stream conditions. We have suggested and Effective algorithm that uses mine High utility trends based list data structure to involving just one search of the database and interested user restrictions limit the space search. This is the Transactional Retail Algorithm uses the Database, raises we add restrictions to the efficiency of our method, such as Duration, object date that helps to make forecasts more precise.

KEYWORDS: *Data mining; high utility pattern mining (HUPM); frequent pattern mining.*

I. INTRODUCTION

The procedure for identifying the right data the patterns and understanding of massive data sets is Data mining is a. The key goal Data mining method is to mine existing information Dataset and change it to a humanly understandable form. Advanced usage [7]. In the long term, a good part of it was shot for a long time Representations viewed as a constant structure Mining calculation can not be insightful to terminate Clients. Irrespective, paying no regard to the norm separating the device, the end-customersshould have had Representations, whichtake place at a low level of Datasets, guy [9].

The sense of utility is linked to the interest, usefulness or profitability of patterns in high utility model mining and the purpose is to remove patterns. Utility greater than or equal to Uh, threshold, minimal utility [10]. In the Big Data age, thus, data mining techniques and Applications are becoming increasingly relevant in differentareas Industry. We will have a better processing power Look at all the details. Corporations are calling for more Performance and more sophisticated approaches for data mining Disclose secret knowledge so that they can make it fast and simple. Adequate business choices [13].

Frequent artefacts and mining contributes to the exploration of similarities and similarities between Objects in large transaction and relational data sets. And large volumes of data on an on-going basis Collected and processed, many of the sectors are being involved in removing those trends from their Database [8]. The data does not have any significance or value unless it is sent a variety of processes for delivery [13].

Many algorithms have been developed to find frequent patterns like FP-Growth, Apriority. Utilities High Data mining is a Template mining is the continuation mining in pattern frequent activities. Strong trends are utility patterns that make money that uses information on costs and quantities of transactions. High utility decision making applications in the realworld Patterns are more useful than common patterns. Mine of trends The knowledge was found in terms of trends from the giant Database. The basic principle of the regular The issue ofpattern mining is to find out the pattern whose level of existence in the database is greater than a particular threshold. [6]

Utility pattern mining has recently been primarily the subject of research.Numerous algorithms have been programmed for static databases in the past, and few algorithms for dynamic databases have been created. The size of the database increases every day in the realworld. We're in High Utility Mining [HUM] Find a collection of products that produce a high profit. Database when they're sold together. User must have a threshold value named 'The minute.' The HIM algorithm produces all of the Strong the utility item set that is the item set that produces at least a minute profit [15].

A large number of candidate patterns that use a large amount of memory are generated by prior algorithms. Since researchers are designing any algorithm that does not create patterns of candidates. To stop building a nominee pattern, the list-based data layout is beneficial. The full data is re-searched as new data is added in complex datasets. But more than one scan takes more processing time, memory.

The list-based data structure algorithm is created to be preserved gradually improve the efficiency of data and its creation and Techniques in restoration. Then there's a mining algorithm Proposed on the basis of limitations such as minutiae threshold, duration, Attribute etc. for high utility mining trends more specifically.

II. LITERATURE SURVEY

HYOJU NAM et al. [1] proposed Traditional model mining is intended to manage a binary database that contains all objects. Database is of the same value, there is a restriction on the identification ofreliable knowledge from the real world. The databases are using the conventional form. To solve this dilemma, the high utility model of mining is approaching. A number of scholars have suggested and actively researched non-binary databases. Recently, there are new results gradually generated with the passage of time in anumber of areas, Biometric information of the patient diagnosed the scale of the database is gradually increasing in the medical devices and log data of Internet users. A database with these characteristics is called a link database. Recently, in these conditions, a broad scale of utility mining strategies suitable for rigorous study of databases has been explored. A new list-based algorithm is introduced in their article, which takes account of the time of arrival of utility models. Any transactionin an incremental database environment. That is to suggest that our algorithm basically performs a pattern cutting using adamp window model which understands the value of the data already achieved. It isless than the latest data which recognises solid service patterns. Experimental findings reveal that their proposed solution is more effective than memory, scalability,state of the art runtime approaches.

Shuning Xing et al. [2] introduced One of the top phases mentioned is UP-Growth. Data structure-based utility pattern mining algorithm The UP-Tree. There is a need, though for building trees structure for many search of the database and wasting a lot of time programming. In order to solve these problems, the introduction of a Fast Utility Tree (FU-Tree) is demonstrated by better houses. In this step, we add LinkQue to reduce the number and prefix utility of scans in the original database to reduce the overestimated utility. In time consumption the FU-Tree outperforms UP-Tree build-up trees and the improvement of high mining productivity utilities theoretical studies and experimental findings show that.

Swapnali Londhe wt al. [3] illustrated TSW scanners Fill in and break in a Two Screens Sliding Pattern Mining Algorithm (TSW) into two parts. Scans both sections to expand the search space simultaneously. This is the algorithm suggested (TSW), which uses two slides to search the text. Windows, allowing multiple alignments when searching for a form. There are three primary approaches in the mining pattern. Area, sliding window, damp window, landmark window. In the sliding window system, stream data is split into two. Multiple chunks are called batches. There are no recent batches Used for the mining pattern approach. New in the damped browser Incoming data viewed as more significant than the previous data One of them. Because of the value of data decays as time goes by. The only way to use data in the landmark window is to use the data in Between the relevant time period and the time period between the present time point and the particular time point. Uh,method. Introduction of a list layout and a new algorithm Use the Sliding control method of the Pane for generating a broad data high utility pattern. Stop this strategy Pattern production of the claimant. The algorithm did not use a large amount of memory space due to the reduction in the candidate pattern. As well as computer software for authentication trends for the nominee, guy. It's an efficient way to understand all these parameters Approach to it.

Guodong Fang et al. [4] proposed a method it is important to keep the network secure Track network traffic in a timelyand efficient way. This is the Orthodox techniques for the identification of network irregularities mainly focused ontechniques such as sampling, counting, andaggregating, but they can't fix the issueOh, get reliable and successful data. In thatpaper, we are introducing a new approach based on the fundamentals Properties ofrecurrent mining trend problems and using vertical mining approaches to mine frequent variations of network traffic.Based on that, Algorithm, we are designing a prototype framework to test our results. Big Net flow Data algorithm for the campus network. This is the experimental result shows that this algorithm is capable of detecting Network abnormalities are timely and efficient and can support Network managers are generating more productivity Network monitoring.

G.S.Vishalini et al. [5] proposed a fascinating proposal Arrangement of laws that show substantial inordinate usefulnessStyles in the unmarried stage without the formation of candidates. The oddities lie inthe architecture of a high utility dependent programme Approach to growth, a method to look ahead, and Straight form records. Solidly, raise our example Method is to go to look at the inverse set count tree and Prune's quest for space by method of utilisation jumping zenith programming. In comparison, we look until you end up distinctly mindful of cases of high usefulness without numbering from the useof the conclusion Tools and a singleton resource. Our straight line Actualities frame our licence to record a respectable positive to be able to prune and to easily pick high Programming software types in an unpractised and unpractised manner. A flexible way that envisions the purpose of thinking with previous calculations. Immense studies of ineffective and thickness, Established and genuine global measurement advocate that our setup ofdirections is as much as it is not less. One or three extra centrality demands it is more adaptable than the measurement of the front line and Finding Acceptable minimum threshold for effectiveness by trial and Error is an expensive process for consumers. When the min util is set Too low, too many HUIs will be produced, It could trigger the mining process to be really serious It's inefficient. In the other hand, if the min util is set, High, no HUIs are likely to be identified.[8].

For the high utility data mining, we give some definitions Itemset mining utility LetJ= {j1, j2, i3,... jm} be a set of The objectsand E = {T01, T02, T03 ,... T0n} are a transaction database. Here, Each Td transaction is distinguished by a specific number known as d, named T0id. The sumand unit benefit of the object jp (1 <= p <= m) They are represented as q(jp) and pr(jp), respectively. These are the In TablesI and II, shown. The X entry set is defined as a set of elements. Separate objects, which are a subset of J. [12].

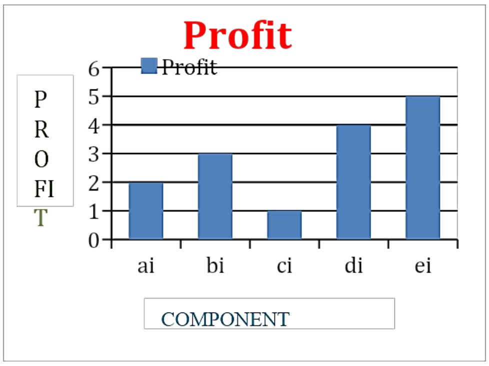
Table1. Table (profit)

Component	ai	bi	ci	di	ei
Profit	2	3	1	4	5

III. METHODOLOGY

Mining with high utility patter method finds all object sets in a Transaction database with a utility value greater than or equal to The minimum usefulness threshold is defined by the user. It also learns that Semantic value among the elements in the mining process [11].

High Utility Sets (HUIs) mining refers to Discovering all the products that have a utility meeting User-specific minimumutility threshold Min util, guy. However, setting min util properly This is a complicated issue for consumers. Tracking



Graph1. Profit graph for the components.

The table1 and graph1 shows the profit forthe components ai, bi, ci, di, ei.

Table2. Database Example

T0id	Transaction
T01	(ai, 1) (bi, 3) (di, 1) (ei, 1)
T02	(ai, 2) (ci, 4) (di, 2)
T03	(bi, 2) (ci, 1) (di, 1) (ei, 2)
T04	(bi, 1) (ci, 3) (di, 2) (ei, 1)
T05	(ai, 2) (ci, 5) (ei, 1)
T06	(ai, 1) (bi, 1) (ci, 3) (di, 1) (ei, 1)
T07	(ai, 3) (ci, 2) (e,i 2)

$u(i, T) = q(i, T) * pr(i)$ (1)
incorporates the downward method Unable to extend the closure property in HUPM. The association rule for the mining algorithmApriori uses theGenerate-and-test process in order to mine useful laws. In other words, that's itrepeatedly scans and lists the candidate trends to identify the Frequent collections of products. Next it counts the assistance for each object to be sought Those with greater support than the minimum support value. Then by increasing, it seeks the next frequent itemsets their length is one in the 'level-wise' way, and repeats this Process before frequent objects are detected [12].

Algorithm1. HUPM based List

$$u(A, T) = \sum_{i \in X \leq T} U(i, T)$$

Input: Minutil, Object, Length, Date, Transaction Database.

Output: Trends of high utility

1. Choose a database
2. Pick Restriction Item/Date
3. Build the layout of list data
4. In the dataset for each transaction, do
5. Creating a list of utilities for each object
6. Sorting utility lists in order as descending
7. Repeat phase 2 to step 6 for incremental details
8. Taking the consumer's barrier, i.e.
9. All changes are minimised
10. Choose the length
11. Return of useful models

Table3. DATABASE in NON-BINARY TRANSACTIONAL

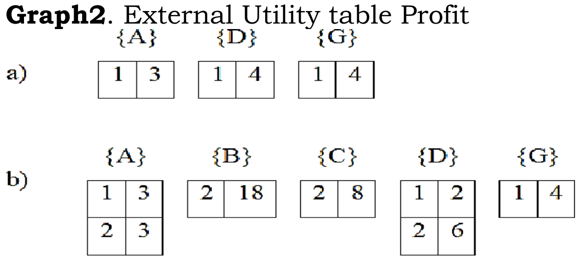
ToID	The T-Transaction	utility sum
To1	(Ai,1) (Di,2) (Gi,1)	12
To2	(Ai,1) (Bi,3) (Ci,2) (Di,3)	36
To3	(Bi,2) (Ci,2) (Di,1) (Ei,3)	38
To4	(Ai,3) (Di,2) (Fi,2)	24

Table4. External utility profit

Component	ai	bi	ci	di	ei	fi	gi
Prof	3.1	6.1	4.1	2.1	3.1	5.1	4.1

The recommended solution utilises Utility data transactional database as input along with Minutil (threshold) minimum utility information [14] and restrictions such as object, period or date the customer. Construction of the structure of data list: List data structure Data list structure the data list structure contains a collection called the utility list. List data structure Information The nominee's pattern is clustered and stored on the utility list. Some utilities list with the duration one it is rendered for a nominee itemset.

Graph2. External Utility table Profit



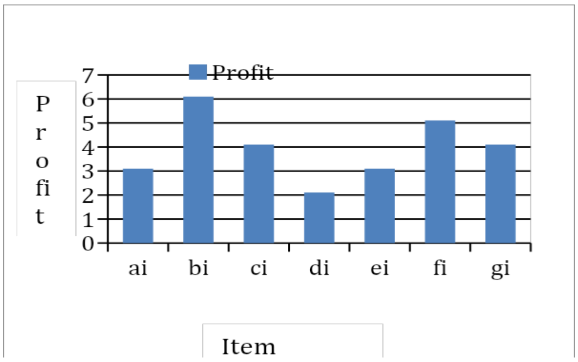


Figure1. List based constructed data structure

T1'	A	D	G	11

A	D	G
11	11	11

T2'	A	B	C	D	35

A	B	C	D	G
46	35	35	46	11

T3'	B	C	D	E	37

A	B	C	D	E	G
46	72	72	83	37	11

T4'	A	D	F	23

A	B	C	D	E	F	G
69	72	72	106	37	23	11

Figure2. Sorting process TWU Table

The items utility of lists G, A, D after T11 is seen in Fig.1 a). The Figu.1 b) indicate the number of products in the utility After the T2 transaction, A,B,C,D,G. List of utilities for each object Includes account idand price (i.e. additional xx utility) Amount) of the product. Utility improves as transaction progresses the list will be revised.

The sorting of utility lists is based on the values of the twu (transaction) [15] Weighted utility) transaction. Sorting algorithm the first transaction is T1'= {A, D, G} according to the initial order and consequence and establishes candidate patterns for A}, {D}, and A} {G}. The next transaction is sorted from the utility set for {B and C} as T02'= {A, B, C, D, G} and the Construct pattern candidate for {C and B}. {D}, {G} and {A}, are compiled in the first trans. Outcomes of the transactions of T01, T02, T03 and T04 are seen in Fig.2 Two. Therefore the $ D<C<B<A<E<F<G order is their twu, the ascending order.

HUPM: After the data structure has been developed when the customer asks and joins the minutility and sorting, Value and limits, for example, length, date or object. Utilization of Restrictions decrease the search space and tend to increase the Algorithm Performance. An algorithm performs a sequence of mining operations Recursively operation.

Let us assume that the values entered by the consumer are minutil=30 and chosen restraint duration=2. Second, our algorithm considers the G, UL (G) List of Resources. The utility value for G is 4, so the pattern is not picked. Algorithm, algorithm checking for all longitudinal service lists in order of trial. The design is picked as a good utility template if the volume of utility is greater than or equal to minutil. Period One, Algorithm Called Mine Now it is two years until theproducts are done. This technique lasts to the end of the limit. Nice, good. Good Performance patterns include: B{B}=30, BC}=46 and BD}=38.1. Performance patterns are: Adding restriction lowers the scanning space on our algorithm so that high utility models are identified more easily.

IV. CONCLUSION

HUPM based-List algorithm discovers restriction dependent on strong utility trends. HUPM algorithms are reminiscent of HUI-Miner, HUI-list INS, which discover patterns of high utility but involve a HUI-Miner. Additional infoscan, which takes a lot of time to sort. Via The HUPM List Based algorithm is usedto discover high usefulness patterns andcan be used for a lot of fine prediction and analysis, by using the threshold and other user-interested limitations as length, variable, date or other. The proposedimprovement to the method for handling complex databases is very powerful. HUPM List-Based Algorithm This is the implementation of various restrictions that reduce time and criteria for manymagnitude recall orders.

V. REFERENCES:

1. Nam, H., Yun, U., Vo, B., Truong, T., Deng, Z. H., & Yoon, E. (2020).Efficient approach for damped window-based high utility pattern mining with list structure. IEEE Access, 8, 50958-50968.
2. Xing, S., Liu, F., Wang, J., Pang, L., & Xu, Z. (2015, December). Utility Pattern Mining Algorithm Based on Improved Utility Pattern Tree. In 2015 8th International Symposium on Computational Intelligence and Design (ISCID) (Vol. 2, pp. 258-261). IEEE.
3. Londhe, S., & Mahajan, R. (2018, August). Two Sliding Window Control Based High Utility Pattern Mining. In 2018 Fourth International Conference on Computing Communication

Control and Automation (ICCUBEA) (pp. 1-4). IEEE.

4. Fang, G., Deng, Z., & Ma, H. (2009, August). Network traffic monitoring based on mining frequent patterns. In 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery (Vol. 7, pp. 571-575). IEEE.
5. Canessane, R. A., Vaishnavi, S., & Vishalini, G. S. (2017, March). Mining high utilities patterns in a medical dataset. In 2017 Third International Conference on Science Technology Engineering & Management (ICONSTEM) (pp. 152-157). IEEE.
6. Lin, K. W., & Luo, Y. C. (2010, October). Efficient strategies for many-task frequent pattern mining in cloud computing environments. In 2010 IEEE International Conference on Systems, Man and Cybernetics (pp. 620-623). IEEE.
7. Patil, D. L., & Tirmare, H. A. (2018, August). Efficient Method for Mining Frequent Itemsets using Temporal Data. In 2018 International Conference on Information, Communication, Engineering and Technology (ICICET) (pp. 1-4). IEEE.
8. Khode, S., & Mohod, S. (2017, April). Mining high utility itemsets using TKO and TKU to find top-k high utility web access patterns. In 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA) (Vol. 1, pp. 504-509). IEEE.
9. Kumar, R., & Canessane, R. A. (2017, March). Improving the efficiency for extracting high utility sequential pattern. In 2017 Third International Conference on Science Technology Engineering & Management (ICONSTEM) (pp. 394-400). IEEE.
10. Alkan, O. K., & Karagoz, P. (2015). CRoM and HuspExt: Improving efficiency of high utility sequential pattern extraction. IEEE Transactions on Knowledge and Data Engineering, 27(10), 2645-2657.
11. Ramaraju, C., & Savarimuthu, N. (2011, June). A conditional tree based novel algorithm for high utility itemset mining. In 2011 International Conference on Recent Trends in Information Technology (ICRTIT) (pp. 701-706). IEEE.
12. Lee, S., & Park, J. S. (2016, January). Top-k high utility itemset mining based on utility-list structures. In 2016 International conference on big data and smart computing (BigComp) (pp. 101-108). IEEE.
13. Hu, H. W., Chang, H. C., & Lin,
14. W. S. (2016, December). An optimized frequent pattern mining algorithm with multiple minimum supports. In 2016 IEEE International Conference on Big Data (Big Data) (pp. 1734-1741). IEEE.
15. Patil, S. K., & Bojewar, S. (2017, May). Application of high utility mining for pattern prediction. In 2017 International Conference on Trends in Electronics and Informatics (ICEI) (pp. 1154-1158). IEEE.
16. Rajendra, I. M., Vyas, C., Moghe, S. S., Deshmukh, D., Sakhare, S., & Gonge, S. (2018, August). Implementing a Hybrid of Efficient Algorithms for Mining Top-K High Utility Itemsets. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1-4). IEEE.
17. 16. Xie, Y., & Philip, S. Y. (2010, December). Max-clique: A top-down graph-based approach to frequent pattern mining. In 2010 IEEE International Conference on Data Mining (pp. 1139-1144). IEEE.

ID: 53

FEATURE SELECTION TECHNIQUES FOR A MACHINE LEARNING MODEL TO DETECT OPTIC NERVE DAMAGE AND DISEASE RECOMMENDATION SYSTEMS

Mr.S.P.SANTHOSHKUMAR
Assitant Professor & Head,
Department of Information Technology,
Rathinam Technical Campus,
Coimbatore, Tamil Nadu, India.

Ms.K.SUSITHRA
Assitant Professor,
Department of Information Technology,
Rathinam Technical Campus,
Coimbatore, Tamil Nadu, India.

ABSTRACT:
Early detection and accurate diagnosis of optic nerve damage and diseases are crucial for effective treatment and preservation of vision. Machine learning models have shown promising results in medical image analysis and healthcare recommendation systems. In this research, we explore feature selection techniques tailored specifically for detecting optic nerve damage and diseases using machine learning models. We curate a comprehensive dataset comprising a variety of medical data, including fundus photographs, optical coherence tomography (OCT) scans, clinical measurements, and patient demographics. We investigate a range of feature selection methods, including correlation analysis, univariate feature selection, recursive feature elimination, LASSO regularization, and tree-based feature importance, to identify the most relevant and informative features for optic nerve disease detection. Through rigorous experimentation and evaluation, we aim to build an optimized machine-learning model with the highest predictive accuracy and generalizability. Furthermore, we develop recommendation systems for personalized treatment guidance based on the identified features and patient profiles. Collaborative filtering and hybrid filtering approaches are integrated to leverage both content-based and collaborative information, tailoring treatment recommendations to individual patients' specific optic nerve conditions. The proposed feature selection techniques and recommendation systems are thoroughly validated using real-world clinical data from ophthalmology practices. Feedback from eye care professionals is collected to assess the system's usability and clinical impact. Throughout the research, we address ethical considerations, including patient data privacy and the role of the systems as decision aids rather than substitutes for medical expertise. The results of this study have the potential to significantly enhance early diagnosis and treatment planning for optic nerve damage and diseases, leading to improved patient outcomes and better management of visual health. By harnessing the power of machine learning and recommendation systems, this research aims to advance the field of ophthalmology and contribute to better eye care for patients affected by optic nerve-related conditions.

KEYWORDS: *Optical Coherence Tomography, LASSO regularization, Glaucoma, Collaborative Filtering, Medical-Diagnosis Method.*

I. INTRODUCTION

The early and accurate detection of optic nerve damage and diseases is critical for preserving vision and providing timely medical intervention. With the increasing availability of medical data and advancements in machine learning, there is a growing interest in developing automated systems to assist ophthalmologists and eye care professionals in diagnosing and managing optic nerve-related conditions. However, the success of such systems heavily depends on the selection of relevant and informative features from the data to build effective machine learning models.

Feature selection is a vital step in the development of machine learning models for optic nerve disease detection and recommendation systems. The procedure entails determining the key elements that directly affect the goal variable, which in this case could be the presence or degree of optic nerve injury or the best course of action for a certain patient. The performance of the model, the computational complexity, and the interpretability of the model can all be improved by choosing the most pertinent features.

This research focuses on exploring various feature selection techniques tailored specifically for optic nerve disease detection and recommendation systems. We aim to develop an efficient and accurate machine learning model that can assist in diagnosing optic nerve damage at early stages and provide personalized treatment recommendations for patients with optic nerve-related conditions.

To achieve these objectives, we begin by curating a comprehensive dataset comprising diverse medical information related to optic nerve health. The dataset includes a range of data sources, such as fundus photographs, optical coherence tomography (OCT) scans, clinical measurements, and patient demographics. The dataset serves as the foundation for training and evaluating our machine learning models and recommendation systems.

We investigate a variety of feature selection techniques, including correlation analysis, univariate feature selection, recursive feature elimination, LASSO regularization, and tree-based feature importance. Each method is applied to identify the most relevant and informative features for optic nerve disease detection from the curated dataset. The comparative analysis of these techniques will help us understand their effectiveness in optimizing model performance.

In addition to optic nerve disease detection, we explore the development of recommendation systems for personalized treatment guidance. These systems aim to assist medical professionals in selecting the most suitable treatment options for individual patients based on their specific optic nerve disease profiles. We integrate collaborative filtering and hybrid filtering approaches to leverage both content-based and collaborative information, thereby ensuring tailored and well-informed treatment recommendations [3].

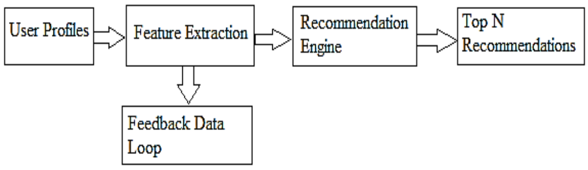


Fig.1 The recommendation systems diagram - high-level representation

Ethical considerations, including patient data privacy and the role of the recommendation systems as decision aids, are thoroughly addressed throughout the research. We emphasize that the machine learning models and recommendation systems are designed to complement the expertise of medical professionals rather than replacing it.

In summary, this study aims to advance the field of ophthalmology by utilising feature selection methods and recommendation systems to identify optic nerve injury and offer specialised treatment advice. The proposed methods have the potential to enhance early diagnosis, improve patient outcomes, and optimize the management of optic nerve-related conditions, ultimately leading to better eye care for individuals affected by these conditions.

II. LITERATURE SURVEY

To find the most pertinent features for detecting optic nerve damage from fundus pictures, Li, F., Yan, L., Wang, et al. study a variety of feature selection strategies, including correlation analysis and recursive feature reduction. The authors suggest a machine learning approach that is highly accurate in detecting early signs of damage to the optic nerve. [1].

In order to diagnose glaucoma using OCT scans of the optic nerve, Juneja, Singh, Agarwal, and others study feature selection techniques such univariate feature selection and LASSO regularisation. The performance of various machine learning models is compared by the authors, who emphasise the significance of feature selection for effective glaucoma identification. [2].

In order to diagnose glaucoma using OCT scans of the optic nerve, Juneja, Singh, Agarwal, and others study feature selection techniques such univariate feature selection and LASSO regularisation. The performance of various machine learning models is compared by the authors, who emphasise the significance of feature selection for effective glaucoma identification. [3].

The use of genetic algorithms to perform feature selection for optic nerve illness categorization from retinal pictures is investigated by R.A. Welikala, M.M. Fraz, J. Dehmeshki, A. Hoppe, V. Tah, S. Mann, T.H. Williamson, S.A. Barman, et al. The study demonstrates the effectiveness of genetic algorithms in selecting informative features for accurate disease detection [4], [14]

Al-Bander B, Williams BM, Al-Nuaimy W et al., this research proposes a hybrid approach that combines feature selection with decision tree classification for optic nerve disease diagnosis. The study examines the effectiveness of several feature selection techniques, demonstrating the hybrid approach's potential for enhancing accuracy. [5], [13].

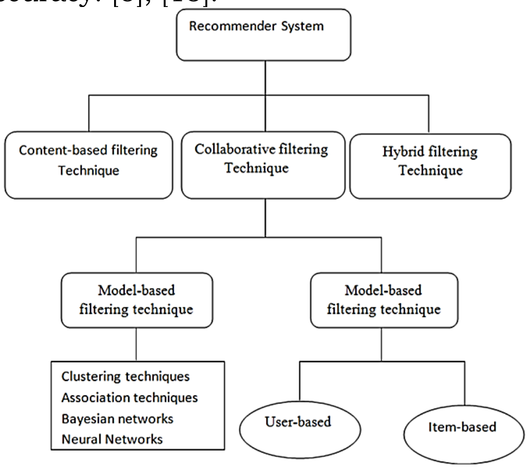


Fig.2 Hierarchy of Recommender System based on filtering

Vij, R., Arora, S. et. al., this systematic review presents an overview of machine learning techniques applied to optic nerve disease detection. The review emphasizes the significance of feature selection methods in improving the interpretability and performance of machine learning models [6].

This study, by Jose E. Valdez-Rodrguez, Edgardo M. Felipe-Riveron, Hiram Calvo, et al., explores the use of feature selection techniques for categorising optic nerve pictures for glaucoma diagnosis. The authors talk about how feature selection affects how well a categorization model performs. [7].

Anuradha Chandra, Arun Kumar Bandyopadhyay and Gautam Bhaduri, et. al., this comparative study evaluates various feature selection techniques for optic nerve disease classification using OCT images. The authors analyze the effectiveness of different methods in achieving accurate disease detection [8].

The review of the literature reveals an increasing interest in feature selection strategies and how they might improve the precision and understandability of machine learning models for diagnosing optic nerve damage and proposing specific course of action. Researchers are examining several strategies to enhance illness detection and treatment planning, resulting in improvements in the area of ophthalmology and better care for patients with disorders affecting the optic nerve.

III. FEATURE SELECTION TECHNIQUES

You can use a variety of feature selection strategies to find the most pertinent and instructive features to develop a machine learning model for identifying disease and injury to the optic nerve. Here are some commonly used feature selection techniques:

A) Univariate Feature Selection:

By calculating each feature's correlation with the target variable, this technique assesses each one separately. To rank features according to their relevance, common statistical tests like chi-square, ANOVA, or mutual information can be utilised.

B) Recursive Feature Elimination (RFE)

RFE is an iterative process that starts with all features and gradually eliminates the ones that aren't crucial based on the performance of the model. To assess the effect of each feature on the model's performance, cross-validation is used.

C) L1 Regularization (Lasso)

The model's cost function receives a penalty term from L1 regularisation, which forces some coefficients to exactly equal zero. As a result, features that are irrelevant will have zero coefficients, resulting in automatic feature selection.

D) Tree-Based Feature Importance

You may gauge the significance of features in decision tree-based models (like Random Forest and Gradient Boosting) by how much they aid in lowering impurity or increasing information acquisition.

E) Feature Importance from Model Coefficients

For linear models (e.g., logistic regression), the absolute values of the coefficients can indicate the importance of each feature.

F) Principal Component Analysis (PCA)

The original features are changed by PCA into a new set of uncorrelated components. The reduced feature set can be chosen from the top main components, which account for the majority of data variance.

IV. FEATURE SELECTION TECHNIQUES

A) Non-Personalized Recommenders

These are the simplest recommender systems that provide recommendations without considering individual user preferences. They offer generic and non-personalized recommendations, such as recommending popular items to all users.

B) Filtering Based on Content

This method suggests products based on the characteristics or aspects of their content. It focuses on recommending products that are comparable to those that a consumer has already interacted with or found appealing.

C) Collaborative Filtering

Recommendations from collaborative filtering techniques are based on user-item interactions. Depending on whether they concentrate on locating comparable persons or comparable products, they can be further separated into user-based collaborative filtering and item-based collaborative filtering

D) .Hybrid Recommenders

Multiple filtering methods are combined by hybrid recommender systems to increase the accuracy and breadth of recommendations. They might incorporate collaborative filtering, content-based filtering, or other techniques.

E) Context-Aware Recommenders

To deliver more individualised and pertinent recommendations, context-aware recommender systems take into consideration additional contextual data, such as time, location, or user context.

F) Demographic Filtering

According to demographic data like age, gender, location, or occupation, this method suggests products. It recommends products that are well-liked by individuals who share similar demographics

.

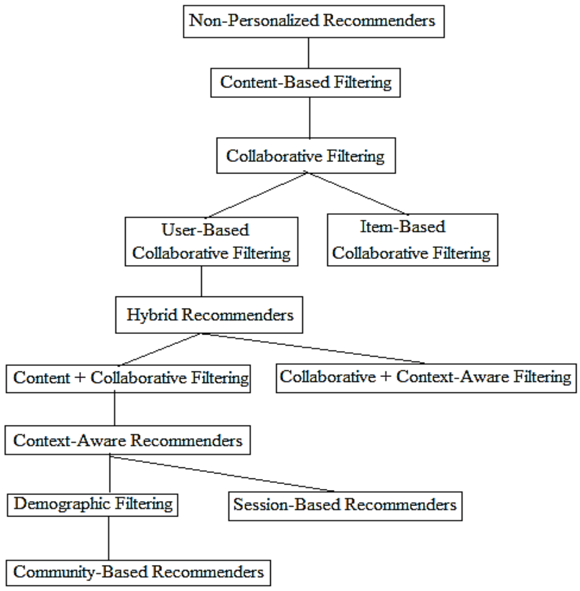


Fig.3 Feature Selection Techniques for a Machine Learning Model to Detect Optic Nerve damage and disease Recommendation Systems Diagram

G) Session-Based Recommenders

User preferences throughout a single session or brief period of time are the main focus of session-based recommender systems. They make suggestions depending on the interactions from the current session.

H) Community-Based Recommenders

Social network data and community behaviour are used by community-based recommender systems to generate recommendations. They suggest things based on the tastes of consumers' online groups or social networks.

This hierarchy shows how recommender systems progress from straightforward, non-personalized methods to more sophisticated, tailored strategies, taking into account elements like content, collaborative filtering, context, demographics, and social connections. The exact use case and data at hand determine which recommender system should be used.

V. COLOR PLANE EXTRACTION

In image processing and computer vision, the colour channels of an image are separated using a technique called colour plane extraction. Red, Green, and Blue (RGB) are the three fundamental colour channels that are used to represent the colour information in the majority of digital colour images. These three colour channels are combined to form the full-color image, which is represented by each pixel in the image. [9].

By offering distinct grayscale representations of each colour component, colour plane extraction enables you to isolate specific colour channels. This can be helpful for a variety of things, including object detection, feature extraction from images, and colour alteration. [10].

To extract a specific color plane (e.g., red, green, or blue) from an RGB image, you can follow these steps:

1. Load the RGB image: Read the RGB image using an appropriate image processing library, such as OpenCV or Pillow in Python.
2. Separate the color channels: Split the RGB image into its red, green, and blue channels. Most image processing libraries provide functions to perform this task [11]. For example, in Python with OpenCV [12].

```
import cv2

# Load the RGB image
image = cv2.imread('path_to_image.jpg')

# Split the image into color channels
blue_channel, green_channel,
red_channel = cv2.split(image)
```

Create grayscale images by combining the colour channels you received in the previous step, each of which holds the image's intensity values for that specific colour. You can convert each color channel into a grayscale image by discarding the other color information:

```
import cv2

# Load the RGB image
image = cv2.imread('path_to_image.jpg')

# Split the image into color channels
blue_channel, green_channel,
red_channel = cv2.split(image)

# Convert individual color channels to
grayscale images
blue_grayscale =
cv2.cvtColor(blue_channel,
cv2.COLOR_BGR2GRAY)
green_grayscale =
cv2.cvtColor(green_channel,
cv2.COLOR_BGR2GRAY)
red_grayscale = cv2.cvtColor(red_channel,
cv2.COLOR_BGR2GRAY)
```

Now, you have three grayscale images representing the blue, green, and red color planes extracted from the original RGB image. These grayscale images can be further processed or analyzed independently to suit your specific application.

Keep in mind that there are other color representations besides RGB, such as HSV (Hue, Saturation, Value), Lab, etc. The process of color plane extraction can vary slightly depending on the color space used, but the general concept remains the same.

VI. CLASSIFICATION IMAGES USED COLOUR PLANES EXTRACTION

A helpful preprocessing step for picture classification applications is colour plane extraction. You can make distinct grayscale images that represent various colour components by taking particular colour planes (such as red, green, or blue) from RGB images. Then, during picture classification, a machine learning model can be utilised with these grayscale photos as input features. [7], [15]

Figure 4: Color planes extraction used for classification

VII. EVALUATION METHOD AND RESULT

To evaluate our methodology we used state of the art metrics depicted in the following equations:

Accuracy = (True Positive+True Negative) / (True Positive + False Positive + False Negative + True Negative) ---- Equ. (1)

Precision = True Positive / (True Positive+ False Positive) ---- Equ. (2)

Recall = True Positive / (True Positive + False Negative) ---- Equ. (3)

Healthy Case = 2 (Precision * Recall / (Precision+Recall)) ---- Equ. (4)

The studies' findings suggest that data augmentation was advantageous and enhanced classification performance generally. However, some experiments did not show significant improvements and can be discarded.

Here is a summary of the results:

RGB + Depth: The combination of RGB color channels with depth information did not lead to improved classification results. This indicates that depth information might not be helpful for classifying glaucoma in this specific task [17]

RGB + INV-Depth: Similarly, combining RGB color channels with inverted depth information did not yield better classification results. This reinforces the observation that depth information might not be relevant for this particular glaucoma classification task.

G, GR: Using only the green channel (G) or the combination of green and red (GR) channels did not significantly improve classification compared to using the original RGB image. These experiments can be discarded as they don't seem to provide any additional benefits [16]. The original RGB image and the union of the green and blue planes produced the best categorization outcomes. Both situations resulted in better categorization performance when combined with data augmentation. [4], [18]. For healthy instances, the categorization outcomes using the grayscale image were successful. In this situation, grayscale photographs may be useful for identifying glaucomatous from healthy eyes. [19]. Based on these results, it is advised that, for the glaucoma classification task, one concentrate on using the original RGB image and the combination of the green and blue planes, coupled with data augmentation. Additionally, grayscale images can be considered for identifying healthy cases [20}

Keep in mind that the dataset, the particular classification method employed, and other factors all affect how well the classification strategy works. In order to ensure the robustness and dependability of the classification model, it is crucial to carefully analyse the findings and perhaps carry out further experiments or analyses.

TABLE 1: Computable results (higher is better).

Experiment	Precision	Recall	Accuracy	Healthy Case
RGB	0.726	0.7946	0.8412	1.5892
RGB-DA	0.947	0.8899	0.9448	1.7798
RGB+Depth	0.763	0.7258	0.8367	1.4516
RGB+Depth-DA	0.9668	0.8432	0.943	1.6864
RGB+INV-Depth	0.78	0.7122	0.8445	1.4244
RGB+INV-Depth-DA	0.765	0.8334	0.942	1.6668
G	0.7942	0.6234	0.8541	1.2468
G-DA	0.89	0.8263	0.924	1.6526
GR	0.6834	0.8247	0.7912	1.6494
GR-DA	0.9771	0.8356	0.934	1.6712
GB	0.9462	0.5978	0.8571	1.1956
GB-DA	0.97	0.8674	0.945	1.7348
Grayscale	0.82	0.5662	0.8241	1.1324
Grayscale-DA	1	0.7122	0.934	1.4244

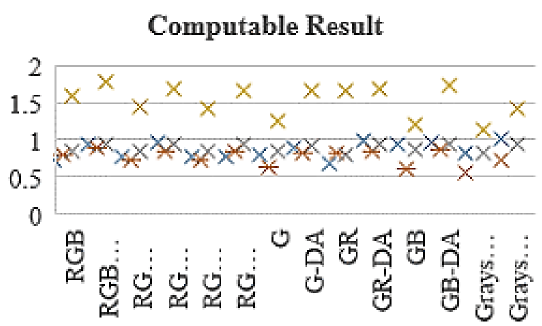


Fig.5 Computable Results

VIII. CONCLUSIONS AND FUTURE WORKS

The work presented in this study focuses on developing a simple Convolutional Neural Network (CNN) model to classify glaucoma in digital retinal fundus color images, specifically under low data conditions. The study introduces a series of pre-processing methods, including optic disc extraction and various combinations of color planes, to enhance the classification performance. The study shows that using the original RGB image yields the best accuracy. However, the pairing of the green and blue colour planes also demonstrated encouraging outcomes, probably as a result of the contrast offered by the optic discs in these images. Grayscale images produced a high level of precision (100%) but also a decline in recall. This shows that while grayscale images may be acceptable for recognising healthy instances, correctly identifying glaucomatous cases may provide some difficulties. The research found that adding depth information did not contribute significantly to the detection of glaucoma. This indicates that depth information might not be useful for this particular classification task. The novelty of the work lies in the comprehensive comparison of different combinations of color planes derived from the RGB image. Despite the complexity of the problem, the study shows that a simple CNN architecture is sufficient to achieve adequate glaucoma classification. The research identified specific types of images that can influence classifier performance. This highlights the importance of understanding the dataset characteristics and considering appropriate pre-processing techniques.

The study suggests further investigation of pre-processing techniques to improve contrast and assess their impact on classification as future work. The study also recommends enlarging the dataset to evaluate the classifier's performance on more photos. Overall, the work advances the field of glaucoma classification by demonstrating the utility of a CNN model in the face of sparse data. It highlights the value of pre-processing methods and the influence of various colour planes on classification outcomes. These results can be used as a foundation for future research to enhance the model's functionality and capacity for handling larger datasets.

IX. REFERENCES

1. Li, F., Yan, L., Wang, Y. et al. Deep learning-based automated detection of glaucomatous optic neuropathy on color fundus photographs. Graefes Arch Clin Exp Ophthalmol 258, 851–867 (2020). https://doi.org/10.1007/s00417-020-04609-8
2. Juneja, M., Singh, S., Agarwal, N. et al. Automated detection of Glaucoma using deep learning convolution network (G-net). Multimed Tools Appl 79, 15531–15553 (2020). https://doi.org/10.1007/s11042-019-7460-4
3. Abhaya Kumar Sahoo, Chittaranjan Pradhan, Rabindra Kumar Barik, and Harishchandra Dubey et al. DeepReco: Deep Learning Based Health Recommender System Using Collaborative Filtering, Computation 2019, 7(2), 25, 22 May 2019; https://doi.org/10.3390/computation7020025
4. R.A. Welikala,M.M. Fraz,J. Dehmeshki,A. Hoppe,V. Tah,S. Mann,T.H. Williamson,S.A. Barman, "Genetic algorithm based feature selection combined with dual classification for the automated detection of proliferative diabetic retinopathy", Computerized Medical Imaging and Graphics, Elsevier, July 2015.
5. Al-Bander B, Williams BM, Al-Nuaimy W et al (2018) Dense fully convolutional segmentation of the optic disc and cup in colour fundus for glaucoma diagnosis. Symmetry-Basel 10(4):87. https://doi.org/10.3390/sym10040087
6. Vij, R., Arora, S. A Systematic Review on Diabetic Retinopathy Detection Using Deep Learning Techniques. Arch Computat Methods Eng 30, 2211–2256 (2023). https://doi.org/10.1007/s11831-022-09862-0
7. Jose E. Valdez-Rodríguez,Edgardo M. Felipe-Riveron and Hiram Calvo, "Optic Disc Preprocessing for Reliable Glaucoma Detection in Small Datasets", Mathematics 2021, 9(18), 2237; https://doi.org/10.3390/math9182237
8. Anuradha Chandra, Arun Kumar Bandyopadhyay and Gautam Bhaduri, "A comparative study of two methods of optic disc evaluation in patients of glaucoma", Oman J Ophthalmol. 2013 May-Aug; 6(2): 103–107, Doi: 10.4103/0974-620X.116643
9. Barros, D.M.; Moura, J.C.; Freire, C.R.; Taleb, A.C.; Valentim, R.A.; Morais, P.S. Machine learning applied to retinal image processing for glaucoma detection: Review and perspective. BioMed. Eng. OnLine 2020, 19, 1–21.
10. Sultana, F.; Sufian, A.; Dutta, P. Advancements in image classification using convolutional neural network. In Proceedings of the 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN),

Kolkata, India, 22–23 November 2018; pp. 122–129.
11. Gonzalez, R.C.; Woods, R.E. Digital Image Processing, 4th ed.; Prentice-Hall, Inc.: Upper Saddle River, NJ, USA, 2018.
12. Wang, S.; Kang, B.; Ma, J.; Zeng, X.; Xiao, M.; Guo, J.; Cai, M.; Yang, J.; Li, Y.; Meng, X.; et al. A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). Eur. Radiol. 2021, 1–9.
13. Raghavendra, U.; Fujita, H.; Bhandary, S.V.; Gudigar, A.; Tan, J.H.; Acharya, U.R. Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images. Inf. Sci. 2018, 441, 41–49.
14. Liu, H.; Li, L.; Wormstone, I.M.; Qiao, C.; Zhang, C.; Liu, P.; Li, S.; Wang, H.; Mou, D.; Pang, R.; et al. Development and validation of a deep learning system to detect glaucomatous optic neuropathy using fundus photographs.
15. Gonzalez, R.C.; Woods, R.E. Digital Image Processing, 4th ed.; Prentice-Hall, Inc.: Upper Saddle River, NJ, USA, 2018.
16. Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. Deep Learning; MIT Press: Cambridge, MA, USA, 2016; Volume 1.
17. Sivaswamy, J.; Krishnadas, S.; Chakravarty, A.; Joshi, G.; Tabish, A.S. A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis. JSM Biomed. Imaging Data Pap. 2015, 2, 1004.
18. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv 2014, arXiv:1409.1556.
19. Kauppi, T.; Kalesnykiene, V.; Kamarainen, J.-K.; Lensu, L.; Sorri, I.; Raninen, A.; Voutilainen, R.; Uusitalo, H.; Kälviäinen, H.; Pietilä, J. DIARETDB1 diabetic retinopathy database and evaluation protocol. In Medical Image Understanding and Analysis; University of Wales: Aberystwyth, UK, 2007; Volume 2007, p. 61.
20. Otsu, N. A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. 1979, 9, 62–66.

ID: 54

POSTULATION OF MICROCALCIFICATION IN MAMMOGRAM IMAGES USING DEEP LEARNING ALGORITHM WITH WOA

Mrs.T.LEENA PREMA KUMARI
Research Scholar,
Department of Computer Application,
Madurai Kamaraj University,
Madurai, Tamil Nadu, India.

Dr.K.PERUMAL
Professor,
Department of Computer Application,
Madurai Kamaraj,
Madurai, Tamil Nadu, India.

ABSTRACT

Now a days cancer is a deadly disease and especially breast cancer is the dangerous disease which increase the mortality rate in women. Early detection or prediction of this can easily cure. When predicting the disease with accurate diagnosis and giving the proper treatment at the early stage can reduce the death rate. Machine Learning and Deep Learning can join hands with the diagnosis of the radiologist to give the better result. The mammography is the technique used to identify the micro calcification in the breast at the early stage. Micro calcification in the breast is a cluster of calcium deposit which indicates that cancer is present .Micro calcification are small, they often occur because of benign on cancer) changes but occasionally micro calcification can be an early sign of cancer.the most common form of cancer that see with calcification can be a ductal carcinoma in situ which is considered as stage 0 cancer.the micro deposit can be identified accurately by deep learning techniques. In this Paper we proposed various deep learning algorithms such as VGG -19, Resnet, Inception, VGG 19 with WOA. When Deep Dense neural network combined with WOA it gives better result as 98% of accuracy when compare with the other CAD algorithms.

KEYWORDS: *Calcification, Whale Optimization algorithm, ResNet, VGG, Inception-Net.*

I. INTRODUCTION

Breast Cancer is the leading cause of mortality rate in women. Breast cancer screening plays a major role to reduce the death rate. Diferent Screening methods are there such as MRI, Ultrasound, Nuclear Medicine Imaging, Tomosynthesis, Mammography. Among this mammography is the easiest and accurate way to predict breast cancer.It is a X-ray method used to examine the breast for the early detection of breast cancer. In this use the low dose x rays systems to see in side the breast. A mammography examined called mammogram. Three recent advances in mammography include digital mammograms, computer aided detection and breast tomosynthesis. It is used to identify the micro calcification in the breast examination. A micro calcification are the tiny deposit of calcium, it may indicate the presence of cancer.the most common type of cancer is ductal carcinoma in situ. Which considered as stage 0 cancer. Deep learning techniques such as VGG -19, Resnet, Inception, VGG 19 with WOA are used to predict the micro calcification. AI which is in the form of CAD system is used to assist radiologist and reduce the workload. Deep learning a subset of ML operates directly on the image data there by defining appropriate features without human interference. In CV outstanding achievements were obtained by DL and when it combines with optimization it gives best result and better prediction. In recent years many application and research scholars used deep learning in breast cancer imaging.

However, the role of AI & DL has exceed the early objective of CAD system to aid radiologist. DL is used in a plethora of different task such as reconstructing image and assessment. This paper gives a different breast cancer modalities and providers. Some future role of these DL techniques re in clinical practice.

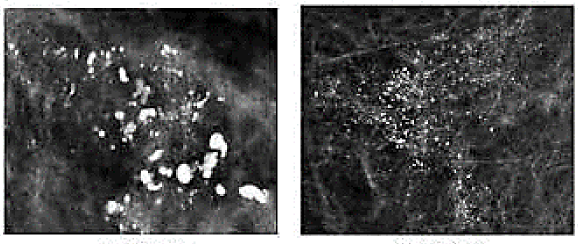


Fig 1: Micro Calcification Images

METHODOLOGY

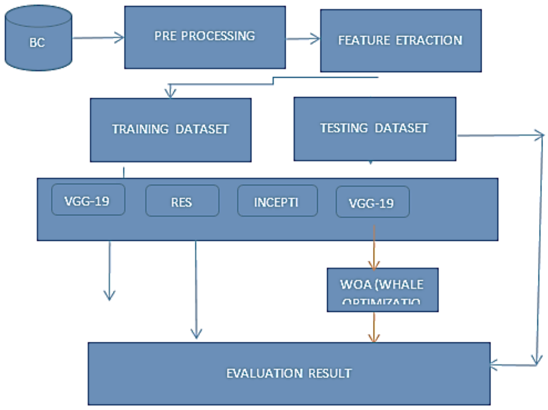


Fig 2: Proposed Method

The proposed system of predicting BC consists of two approaches:

1. Regular DL approach
2. DL combine with optimization algorithm

In the regular DL approach different types of CAD algorithms such as VGG -19, Resnet, Inception, VGG 19 with WOA were analysis. In the second type of approach. Deep learning is combined with whale optimization algorithm

The steps of the proposed system includes feature selection methods splitting datasets, training models and evaluation models.

FEATURE SELECTION METHODS BASED ON ABSOLUTE MAXIMUM SUPPORT FEATURE SELECTION (AMSFS)

It provides some key information about an image.It is one of the most important part to recognize the patterns. The accuracy of the result depends on selecting the right feature.The overlapping of the mass tissues make to identify the calcification and to classify the result.So the combination related to the shape,density contour , texture are preferred.

The weights are calculated using the bias-increasing key technique.Consider the two classes Q and R. These two re the feature vectors. A and B are the feature vectors of the samples belongingto X = $(x_1, x_2, \dots x_n)$ and Y = $(y_1, y_2, \dots y_n)$, difference between in the $d^{th}(d = 1,2, \dots n)$ features d_x and d_y of the two samples the belongs x and y. is more, then that feature plays an important section.

$$P(d_x, d_y)=\int_{\infty}^{+\infty} |d_x - d_y|\ f_x(d_x, d_y)pd_x * pd_y$$

Structure the model for maximizing the deviation are as follows a

$$\max P(\lambda_d) = \sum_{d=1}^{n} \lambda_d\ P(d_x, d_y)$$

This gives weight to each feature of the model, Classification is performed using a featureset of statistical, textual and clinical data. A predictive support vector is built and trained using weighted features from the selecting features. The test models are then checked against the absolute vector matrix of test features.

Steps:

Step 1: Input sample features set = $\{p_x, q_x\}x = 1$ to M // where p_x is features vector and q_x is the classes

Step 2: The vector matrix is constructed by the features

Step 3: The appropriate feature parameter is be select.

Step 4: Structure the Absolute the function $s(a) = sgn(\sum u = 1 \text{ to } l\ a_x P \times V(a_x, x) + v^*)$ //v^* – vector image pixels

The AMSFS algorithm based on weighted features Absolute Maximum Support Feature Selection (AMSFS) is similar to the absolute support feature vectors, except that the vector matrix absolute values is replaced by maximum feature weights.

DEEP LEARNING ALGORITHMS

INCEPTION:

In Deep learning inception network consists of repeating blocks which act as an input to the next upcoming block.It is a CNN neural network design for the image classification. It is known for inception models to be used for image classifications. Here they are used with the blocks of layers which are designed to learn a combination of local and global features from the input data. It is used for the more efficient computation and for deeper network with the reduction in the dimensional.These models are used for the problem solving with computational

ISBN: 978-81-965283-7-9

expense as well as over fitting.

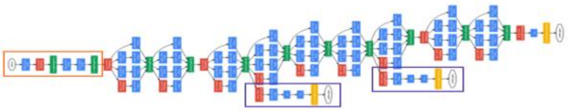


Fig 3: Inception Architecture

VGGNET-19:

It is a Visual geometry Group it is deep convolution neural network with multiple layers . The deep is used to mention number of layers in the network VGG-16 & VGG-19.It consists of (16 CV layers,3Fully connected layers,5 Max Pool layers and 1 Soft max layer).In fake image classification also it provides the better accuracy than the Dense Net models.The minimum image size required by the model is 32X32.The pre-trained model can classify images into 1000 object categories.

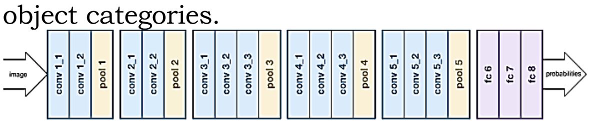


FIG 4: VGG-19 ARCHITECTURE

RESNET 50:

It is a constitutional network with deep 50 layers. We can per-trained the model with more than million of images. In this model it has 48 convolution layer with 1 MaxPool and 1 Average Pool layer. It overcomes the vanishing gradient problem it is to construct network with up to thousand of conventional layers which is used for out perform shallower network.It uses images o 224 pixels X 224 pixels.

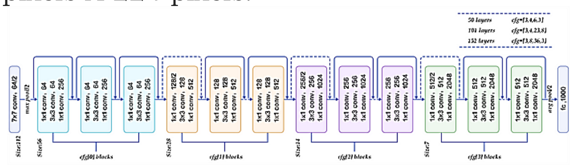


Fig 5: Res Net 50 Architecture

WHALE OPTIMIZATION ALGORITHM

The Procedure for obtain the optimum value of the function is called Optimization Algorithm. Meta heuristic optimization algorithm is more popular in soft computing applications. It is easy to implement and it do not require the gradient information,it can by pass local optima.It is used in the wide range of problems. This WOA is used for the search process and select the relevant features for classification of images. It changes the position of the search agent related to get the best fitness value.

A nature inspired meta-heuristic optimization algorithm which mimics the hunting behaviour of humpback whales. The algorithm is inspired by the bubble-net hunting strategy .Two maneuvers associated with bubble net feeding are 'upward-spirals' and 'double-loops'.In **'upward-spirals'** maneuver humpback whales dive around 12 m down and then start to create bubble in a spiral shape around the prey and swim up toward the surface.'**double-loops**' maneuver includes three different stages: coral loop, lobtail, and capture loop.

Encircling Prey

Current best candidate solution is assumed to be closes to target prey and other solutions update their position towards the best agent

$$\vec{D} = |\vec{C}.\vec{X}_{best}(t) - \vec{X}(t)| \quad (1)$$

$$\vec{X}(t+1) = \vec{X}_{best}(t) - \vec{A}.\vec{D} \quad (1)$$

Where t indicates the current iteration, $\vec{A}$ and $\vec{C}$ are coefficient vectors, $\vec{X}_{best}$ is the position vector of the best solution, and $\vec{X}$ indicates the position vector of a grey wolf

$$\vec{A} = 2\vec{a}\vec{r_1} - \vec{a} \quad (2)$$

$$\vec{C} = 2\vec{r_2} \quad (3)$$

Where $\vec{r_1}, \vec{r_2}$ are random vectors in [0, 1].

Bubble-net attacking method (exploitation phase)

In order to mathematically model the bubble-net behaviour of humpback whales, two approaches are designed as follows:

1. Shrinking encircling mechanism

This behaviour is achieved by decreasing the value of $\vec{a}$. a is decreased from 2 to 0 over the course of iterations.

2. Spiral updating position

$$\vec{D}' = |\vec{X}_{best}(t) - \vec{X}(t)| \quad (4)$$

$$\vec{X}(t+1) = \vec{D}'.e^{bl}.\cos(2\pi l) + \vec{X}_{best}(t) \quad (5)$$

l is a random number in [-1, 1]

Search for prey:

Humpback whales search randomly according to the position of each other

$$\vec{D} = |\vec{C}.\vec{X}_{rand}(t) - \vec{X}(t)| \quad (6)$$

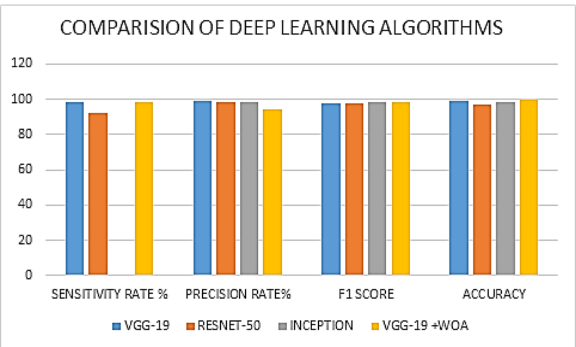
$$\vec{X}(t+1) = \vec{X}_{rand}(t) - \vec{A}.\vec{D} \quad (7)$$

ISBN: 978-81-965283-7-9

EVALUTION AND RESULT

We compare the proposed frame work with some are of art techniques which is used with same data set.It compares the proposed methods with the recent techniques. The VGG-19d gives the accuracy of 98.5% withVGG-19. When the CNN algorithm combines with the Whale Optimization Algorithm (WOA) the accuracy improves.The WOA is used to selection the need feature from the image. So the accuracy is improved by 99.1% when the proposd method VGG-19 combine with WOA.

CNN Algorithms	Sensitivity Rate %	Precision Rate%	F1 Score	Accuracy
VGG-19	98.2	99.1	98.04	98.5
RESNET-50	92.1	98.3	97.55	97.3
INCEPTION	0.5	98.6	98.2	98.3
VGG-19 +WOA	98.7	94.05	98.2	99.1



CONCLUSION

To predict the calcification from the mammogram images various deep learning algorithms were used to increase the accuracy. In the proposed method we combine the VGG-19 with the optimization algorithm (WOA) which gives the better accuracy than the other methods.The accuracy of the proposed method gives Sensitivity as 98.7%, Precision rate 94.05%, F1 Score 98.2% and accuracy 99.1%.

REFERENCES:

1. *Y. Ma, P. C. Tay, R. D. Adams et al., "A novel shape feature to classify* microcalcifications," in Proceedings of IEEE International Conference on Image Processing, pp. 2265–2268, Hong Kong, China, September 2010.View at: Google Scholar
2. H. D. Cheng, X. Cai, X. Chen, L. Hu, and X. Lou, "Computer-aided detection and classification of microcalcifications in mammograms: a survey," Pattern Recognition, vol. 36, no. 12, pp. 2967–2991, 2003.View at: Publisher Site | Google Scholar
3. Hamidinekoo A, Denton E, Rampun A, Honnor K, Zwiggelaar R. Deep learning in mammography and breast histology, an overview and future trends. Med Image Anal. 2018;**47**:45-67. [PubMed ID: 29679847]. https://doi.org/10.101 6/j.media.2018.03.006.
4. Shen, L.; Margolies, L.R.; Rothstein, J.H.; Fluder, E.; McBride, R.; Sieh, W. Deep Learning to Improve Breast Cancer Detection on Screening Mammography. Sci. Rep. **2019**, 9, 12495. [**Google Scholar**] [**CrossRef**]
5. Yunus, M., Ahmed, N. & Masroor, I. Mammographic criteria for determining the diagnostic value of microcalcifications in the detection of early breast cancer. J Pak Med Assoc **4**(1), 24–29 (2004).
6. Melekoodappattu, J.G.; Subbian, P.S. A Hybridized ELM for automatic micro calcification detection in mammogram images based on multi-scale features. J. Med. Syst. **2019**, 43, 183. [**Google Scholar**] [**CrossRef**]
7. Suresh, R.; Rao, A.N.; Reddy, B.E. Detection and classification of normal and abnormal patterns in mammograms using deep neural network. Concurr. Comput. Pract. Exp. **2019**, 31, e5293. [**Google Scholar**] [**CrossRef**]
8. Rulaningtyas, R.; Hyperastuty, A.S.; Rahaju, A.S. Histopathology Grading Identification of Breast Cancer Based on Texture Classification Using GLCM and Neural Network Method. J. Physics: Conf. Ser. **2018**, 1120, 012050. [**Google Scholar**] [**CrossRef**]
9. Abdel-Basset M, Hawash H, Chakrabortty RK, Ryan M. Energy-net: a deep learning approach for smart energy management in iot-based smart cities. IEEE Internet of Things J. 2021.
10. Michailovich, O.; Tannenbaum, A. Despeckling of medical ultrasound images. IEEE Trans. Ultrason. Ferroelectr. Freq. Control **2006**, 53, 64–78. [**Google Scholar**] [**CrossRef**] [**PubMed**]
11. Boudouh, S.S.; Bouakkaz, M. Breast Cancer: Breast Tumor Detection Using Deep Transfer Learning Techniques in

Mammogram Images. In Proceedings of the 2022 International Conference on Computer Science and Software Engineering (CSASE), Duhok, Iraq, 15–17 March 2022; IEEE: Duhok, Iraq, 2022; pp. 289–294. [**Google Scholar**] [**CrossRef**]

12. Mrs.T.Leena Prema Kumari, Dr.K.Perumal “Mammography Images Classification and Detection Based On Deep Dense Net Convolution Generative Neural Network, Journal of Data Acquisition and Processing, 2023, **38** (2): 3119-3136.

ID: 55

AN ENHANCED PANCREATIC CANCER CLASSIFICATION USING DEEP NEURAL NETWORK

Mrs.R.DHIVYA
Research Scholar (FT),
Department of Computer Science,
Dr.SNS Rajalakshmi College of Arts and Science,
Coimbatore, Tamil Nadu, India.

Dr.N.SHANMUGAPRIYA
Associate Professor & Head,
Department of Computer Applications (PG),
Dr. SNS Rajalakshmi College of Arts & Science,
Coimbatore, Tamil Nadu, India.

ABSTRACT:
The classification of pancreatic cancer (PC) in CT images is useful for PC diagnosis, increase rate prediction, tumour volume measures, and treatment planning. PC is classified via a biopsy, which is not normally performed prior to flawless surgery. Technological advancements and machine learning can aid radiologists in tumour diagnosis without the use of invasive procedures. The Convolutional Neural Network (CNN) is a deep-learning system that has produced significant achievements in segmentation and classification. In order to address the optimization issue during high dimensional data classification, this paper presented an Enhanced CNN with LSTM (Long Short Term Memory), which incorporates Pancreas Segmentation technique. The proposed Hybrid CNN-LSTM technique can be used to find tumors in high dimensional datasets in huge quantities. The experimental findings showed that the proposed hybrid CNN-LSTM outperformed alternative ensemble learning-support vector machine (EL-SVM) and existing machine learning techniques, especially in terms of AUC.

KEYWORDS: *Pancreatic cancer, diagnosis and staging, CNN,-LSTM, deep learning, CT.*

I. INTRODUCTION

Image processing is a method for conducting several processes on a single image in order to enhance it or extract significant information. It's a type of signal processing in which images serves as the input and the result is either a new image or the image's characteristics. In today's world, image processing is one of the most rapidly evolving technologies. It's also a hub of engineering and computer science research [1].

One of the most common cancers, pancreatic cancer, has a dismal prognosis and is usually fatal 1–3. Patients without a tumor just need to be examined further; however a pancreatic tumor diagnosis necessitates immediate treatment and a specific surgical strategy. If treatment is delayed, the risk of aggravation and death increases, making an accurate diagnosis of a pancreatic tumor critical to its successful surgical treatment. Figure 1 depicts a CT scan of a pancreatic cancer patient.

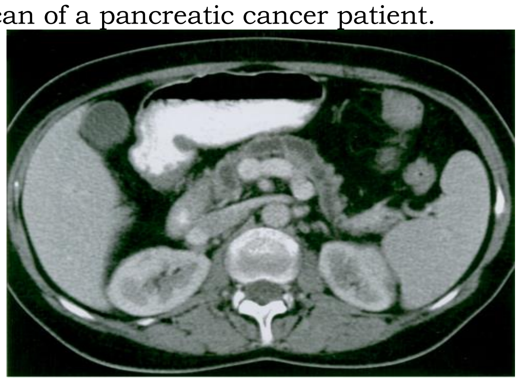


Fig.1: CT Pancreatic cancer image

Natural Language Processing (NLP), speech recognition, and computer vision have all benefited from deep learning [2]. Convolutional networks [3], often referred to as CNNs, are a subset of neural networks that are used to process data with a grid-like structure, such as image data, which may be regarded as a two-dimensional grid of pixels. CNNs have been used to produce state-of-the-art performance in image recognition [4], semantic segmentation [5], and object detection [6], stereo matching [7], and other computer vision applications. Convolutional neural networks are essentially neural networks with at least one layer that uses convolution instead of traditional matrix multiplication.

Convolutional neural networks are similar to traditional neural networks in that they consist of neurons with learnable weights and unpredictable behavior. Each neuron takes a few inputs, conducts a dot product, and then, if desired, does non-linearity activation function. The complete neural network represents a unique differentiable function, from the unprocessed input picture pixels on one end to class scores on the other. They still have a loss function on the last layer (e.g. cross entropy). All of the general neural network parameter learning techniques applies to CNNs as well. CNN architectures are based on inputs images, to add specific attributes to the architecture.

Each output unit in a typical neural network layer is linked to all input units, as shown in Figure 2. However, light interactions among input and result units, also known as light connectivity, are common in convolutional neural networks, which can be attained by making the kernel dimension of CNNs lesser the input dimension illustrated in Figure 3.

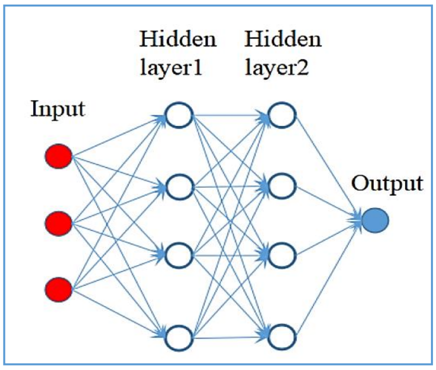


Fig.2 FC Layer

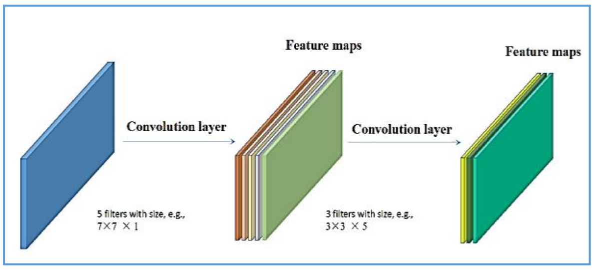


Figure.3 Illustrations of convolutional networks

The split of a computer image into distinct pieces is referred to as PC segmentation. Segmentation seeks to reduce and/or modify an image's representation to make it more understandable. The process of image segmentation is widely used to locate borders and objects in images.

The present research proposes a unique method for automatically classifying pancreatic cancer utilizing computed tomography (CT) images using hybrid CNN-LSTM model.

II. LITERATURE REVIEW

Texture analysis approaches, which included arithmetical methods and the wavelet transform, were utilized to mine texture features of non-enhanced CT images, according to Qiu et al. (2018) [8]. A mixing mode was devised, which combined the wavelet transform with an arithmetical technique. Wavelet transform (WT) combined with histogram (WTH), and wavelet transform combined with co-occurrence matrix (WTCM) were used to analyzed texture (WTCM). The proposed combination mode was employed by the WTH and WTCM. The approach based on mutual information was then used to pick features. The support vector machine (SVM) and the probabilistic neural network (PNN) are then used as classifiers (PNN).

Liu et al. (2019) [9] discussed precise imaging processing procedure system that allows it to correctly read CT images and speed up pancreatic cancer diagnosis. Two processes were used to create the system for pancreatic cancer analysis based on chronological contrast-enhanced CT images: training and verification.

According to **K. Holub and C. Conill (2020)** [10], pancreatic cancer (PC) is one of the largest part hostile tumors, with no valuable treatment if diagnosed at a superior stage. Systemic inflammation is a well-known hallmark of cancer growth, and they believed that a greater understanding of the function of inflammatory factors in PC treatment could lead to better results.

Huang et al. (2020) [11] proposed a LASSO technique for attribute collection, the typical optimization time was condensed by more than 25 seconds while maintaining classification accuracy, and serial synthesis was executed. After the EL-SVM classifier was used to recognize and categorize more than four hundred biopsy tissue images, the Receiver Operating Characteristic (ROC) curve and error curve were utilized to test the classifier's generalization performance.

Huo et al. (2020) [12] suggested the Sparse Group Lasso and Support Vector Machine as a new technique of tumor classification (SGL-SVM). The Kruskal–Wallis rank sum test is used to do the key choice of attribute genes on normalized tumor datasets. Second, a sparse group Lasso is used for additional selection, and ultimately, a support vector machine is used as a classifier. On microarray and NGS datasets, they tested the proposed technique.

The technique presented by S. E. Roshan and S. Asadi (2020) [13] utilized multi-objective evolutionary to generate a set of diversified, well-performing, and (almost) unbiased bags. As a result, the suggested method allows for the generation of diverse and high-performing classifiers as well as the determination of the number of classifiers in the Bagging algorithm. Furthermore, the proposed strategy employs two separate strategies to increase diversity.

According to Togaçar, et al. (2020) [14], lung tumor identification is accomplished utilizing the deep learning models. An assortment of computed tomography (CT) images that are readily accessible were used for the tests. The experiment employed feature extraction and classification using convolutional neural networks (CNNs).

III. PROPOSED METHODOLOGY

The present research provides a new classification method for pancreatic cancer utilizing CNN with LSTM (Long Short Term Memory) version in MATLAB simulation on a dataset of CT images. Figure 4's basic planned flow diagram uses a PC category technique from start to finish.

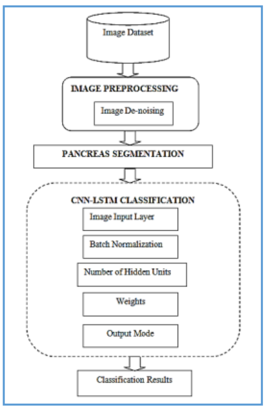


Fig.4 Proposed Flow diagram

A. IMAGE PREPROCESSING

The image preprocessing is prior to training or prediction, conduct a minimum quantity of preprocessing and several label-preserving changes on the photos, largely following the method employed, Image Resizing using the 'bicubic' approach, the entire images are scaled to a minimum dimension of 256 pixels.

- Image denosing: Using the Wiener filter, remove unnecessary noise in the original image.

Random crops: The approach requires images with specific dimensions; when an image is chosen for training, a random crop with a size of 256 by 256 pixels is chosen. During training, the size of inputs is fixed, allowing for significantly faster CPU processing.

The predictable image preprocessing methodology responds as it prepares image data that is undefined and ready for analysis, with no feedback and no communication for the data gathering procedure. The most difficult aspect of data preparation is the data variation between data sets. Clear out customs work in data preprocessing to clean up the data by filling in missing values, smoothing noisy data, detecting or removing outliers, and determining abnormalities. Image denoising is a technique for removing noise from a corrupted image while preserving the borders and other fine details. With a filter size of 3 × 3, a Wiener filter is employed to eliminate noise in this data preparation.

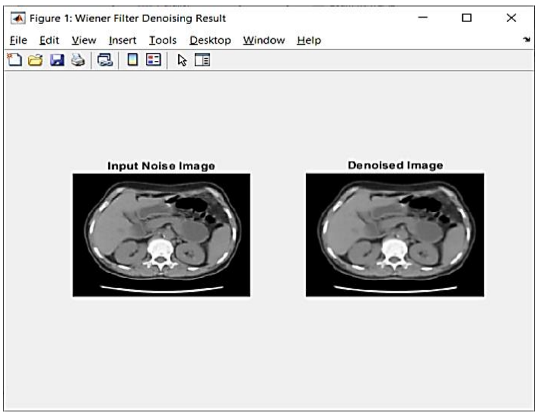


Fig.5 Wiener Filter Denoised Result

B. PANCREAS SEGMENTATION

The computer-assisted diagnosis of PC CT images has been a game-changer in the early detection of lung illnesses. Preprocessing the image in order to segment it is the best way for adopting computer aided diagnosis for medical image analysis.

The first step is to use a wiener filter to remove 2-D adaptive noise-removal filtering. Noise removal combined with the segmentation procedure is a simpler approach. The suggested OTSU approach, First order cumulative and Histogram based lung segmentation technique can help radiologist in early diagnosis of lung disorders, but the segmentation algorithm can also be utilized to early diagnose other benign or malignant condition.

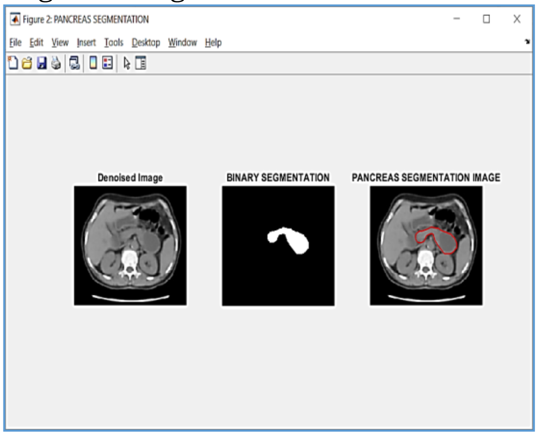


Fig.6 Pancreas Segmentation Result

C. *CNN-LSTM NETWORK CLASSIFICATION*

Using improved CNN-LSTM, a mixture technique was formed to automatically classify pancreas cancer. This architecture's structure was created by fusing CNN and LSTM networks, where CNN is utilized to extract intricate details from images and LSTM serves as a classifier. Layers make up the network architecture, and each one defines a specific computation. Deep Learning allows you to layer-by-layer construct a Neural Network. There are the following layers available:

- **imageInputLayer** - A layer for displaying images (The network's capacity for handling different types and volumes of data is determined by the input layer)
- **convolution2dLayer** - Convolutional Neural Networks' 2D convolution layer
- **ReLU**- (Rectified Linear Unit) layer (reluLayer)
- **maxPooling2dLayer** - Maximum pooling layer (Sample data is laid down as it moves through the network.)
- pooling layers
- **fullyConnectedLayer** - Fully connected layer
- **SoftmaxLayer and Classification Layer** - Softmax layer and class layer

Three gates—an input gate, a "forget" gate, and an output gate—are combined with an LSTM., where m_t refers to the present input; PS_t and PS_{t-1} denote the novel and earlier cell states, correspondingly; and o_t and o_{t-1} are the present and earlier results, correspondingly.

IV. EXPERIMENAL RESULTS

The proposed hybrid CNN-LSTM technique has been used to estimate the outcomes. The simulations are done using MATLAB R2018b and Python 3.7, and the findings are implemented using an Intel I7 processor, 16GB main memory, and Windows 10 as the operating system. This paper is implemented with by Pancreatic Cancer CT image dataset capture from The Cancer Imaging Archive (TCIA) Public Access database. The resulting parameters of PC Classification activity and overall CNN-LSTM training accuracy with loss is described in below figures,

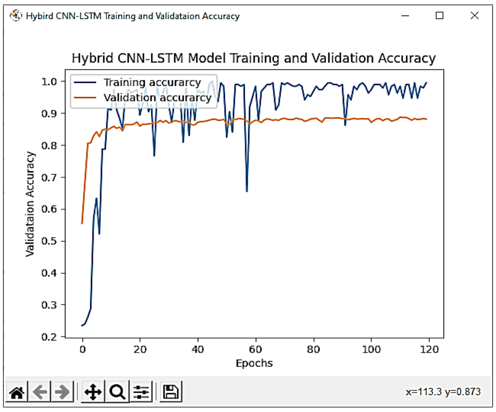


Fig.7: Hybrid CNN-LSTM model accuracy.

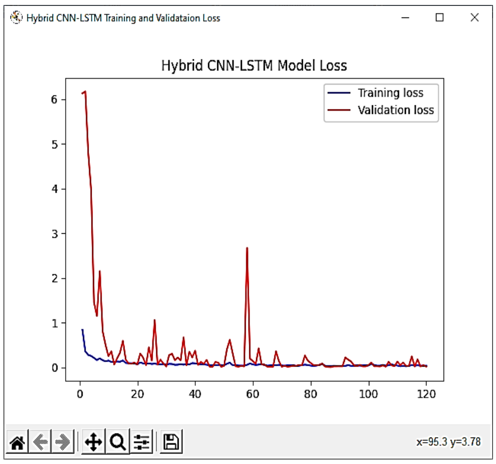


Fig.8: Hybrid CNN-LSTM model Loss.

The Area Under Curve (AUC) curves of SVM [15], BP, KNN, LASSO-SVM [16], EL-SVM [17], and proposed Hybrid CNN-LSTM as shown in Figure 9. It can be seen from Figure 9 that the AUC value of CNN-LSTM classification increased by 0.0762 compared to EL-SVM classification. The results show that CNN-LSTM has good feasibility.

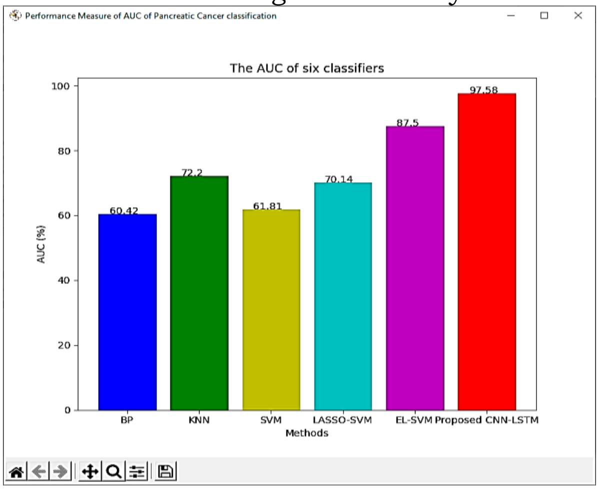


Fig.9 ACU of six Classifiers

V. CONCLUSION

The paper developed and implemented an improved method of CNN with LSTM classification algorithm for Pancreatic Cancer dataset that includes Pancreas Segmentation with CNN-LSTM network model is trained to exploit the probability of the cancer portions given the image. The suggested method uses overlapping crops of complete slices of PC data to train deep convolutional neural networks. Segmenting tumours from the backdrop is a highly unbalanced dense prediction job since cancer pixels make up such a small fraction of the overall slice image.

VI. REFERENCES

1. Stefan Bauer, Lutz-P., Nolte, and Mauricio Reyes., "Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization". In Proceedings of International Conference on Medical Image Computing and ComputerAssisted Intervention (MICCAI), pages 354–361, 2011.
2. Ian Goodfellow, Yoshua Bengio, and Aaron Courville., "Deep learning". MIT press, 2016.
3. Yan LeCu, "Generalization and network design strategies", Technical report, University of Toronto, 1989.
4. Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for largescale image recognition". In Proceedings of International Conference on Learning Representations (ICLR), San Diego, USA, 2015.
5. Jonathan Long, Evan Shelhamer, and Trevor Darrell., "Fully convolutional networks for semantic segmentation", In Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, USA, 2015.
6. Ross Girshick, "Fast r-cnn", In International Conference on Computer Vision (ICCV), 2015.
7. Jure Zbontar and Yann LeCun, "Stereo matching by training a convolutional neural network to compare image patches", Journal of Machine Learning Research, 17:1–32, 2016.
8. J.-J. Qiu, Y. Wu, B. Hui, Z.-X. Huang, and J. Chen, "Texture analysis of computed tomography images in the classification of pancreatic cancer and normal pancreas: A feasibility study," J. Med. Imag. Health Informat., vol. 8, no. 8, pp. 1539-1545, Oct. 2018.
9. S.-L. Liu, S. Li, Y.-T. Guo, Y.-P. Zhou, Z.-D. Zhang, S. Li, and Y. Lu, "Establishment and application of an articial intelligence diagnosis system for pancreatic cancer with a faster region-based convolutional neural network," Chin. Med. J., vol. 132, no. 23, pp. 2795–2803, Dec. 2019.
10. Holub and C. Conill, "Unveiling the mechanisms of immune evasion in pancreatic cancer: May it be a systemic inammation responsible for dismal survival?" Clin. Transl. Oncol., vol. 22, no. 1, pp. 8190, Jan. 2020.
11. P. Huang, S. Zhang, M. Li, J. Wang, C. Ma, B. Wang, and X. Lv, "Classification of cervical biopsy images based on LASSO and EL-SVM," IEEE Access, vol. 8, pp. 2421924228, 2020.
12. Y. Huo, L. Xin, C. Kang, M.Wang, Q. Ma, and B.Yu, "SGL-SVM:Anovel method for tumor classication via support vector machine with sparse group lasso," J. Theor. Biol., vol. 486, Feb. 2020, Art. no. 110098.
13. S. E. Roshan and S. Asadi, "Improvement of bagging performance for classication of imbalanced datasets

using evolutionary multi-objective optimization," Eng. Appl. Artif. Intell., vol. 87, Jan. 2020, Art. no. 103319.
14. Togaçar, B. Ergen, and Z. Cömert, "Detection of lung cancer on chest CT images using minimum redundancy maximum relevance feature selection method with convolutional neural networks," Biocybern. Biomed. Eng., vol. 40, no. 1, pp. 23-39, Jan. 2020.
15. Liu, M. Li, W. Lan, F.-X. Wu, Y. Pan, and J. Wang, "Classification of Alzheimer's disease using whole brain hierarchical network," IEEE/ACM Trans. Comput. Biol. Bioinf., vol. 15, no. 2, pp. 624–632, Mar. 2018.
16. Mahdianpari, B. Salehi, M. Rezaee, F. Mohammadimanesh, and Y. Zhang, "Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery," Remote Sens., vol. 10, no. 7, p. 1119, Jul. 2018.
17. Min Li, Xiaohan Nie, Yilidan Reheman, Pan Huang, Shuailei Zhang, Yushuai Yuan, Chen Chen, Ziwei Yan, Cheng Chen, Xiaoyi Lv, And Wei Han, "Computer-Aided Diagnosis and Staging of Pancreatic Cancer Based on CT Images", IEEE Access, August 13, 2020.

ID: 56

INTRUSION DETECTION AND DATA SECURITY IN CLOUD ENVIRONMENT

Ms.S.AROCKIA PANIMALAR
Assistant Professor,
Department of Information Technology,
Nehru Arts and Science College,
Coimbatore, Tamil Nadu, India.

Dr.A.KRISHNAKUMAR
Associate Professor,
Department of Computer Science,
Nehru Arts and Science College,
Coimbatore, Tamil Nadu, India.

ABSTRACT

Smart applications aim to enhance productivity to support humans in terms of efficacy and security. Cloud computing is a hub for data analytics and intelligence due to its capacity for storage and computation. However, it faces critical concerns like security and privacy. Despite this, cloud computing offers advantages like flexibility, cost-effectiveness, and integration with new services. Yet, it also deals with security issues such as malicious attacks and data loss. The combination of IoT and cloud computing has led to technological revolutions, but they are susceptible to hacking and malicious control. Current efforts to address security issues have treated them as separate problems, leading to time complexity and data confidentiality issues.

KEYWORDS: *Cloud, Security, Machine Learning, Intrusion, Attack, Network, Services.*

I. INTRODUCTION

Cloud computing [1] increases the technological capability to deliver different types of services over the internet without the need of deploying them on a local computer or server. The word “cloud” in cloud computing mainly implies that the information accessed is remotely found in the virtual place. The cloud offers a wide range of services, tools, and resources via the internet such as servers, databases, software, network, and applications. The on-demand availability of the data centers does not require any active management by its consumers.

Cloud computing allows users to store data in a remote database, reducing reliance on costly storage devices. It supports organizations' business processes [2] by offering internet-based services, serving as a metaphor for the term "Internet." Various organizations of different sizes and types use the cloud for diverse services. For example, medical institutions utilize the cloud for patient services, and financial institutions employ it for real-time fraud detection.

Cloud computing provides a virtual office environment, facilitating connectivity for individuals worldwide. Its major advantages include cost reduction, scalability, business continuity, and automatic updates. Various cloud computing models cater to user needs, and not all clouds are identical. The types of cloud computing models [3] are public, private, and hybrid. To deploy a cloud service, users must select the suitable cloud deployment model based on their needs: public, private, or hybrid [4].

The cloud poses risks of unauthorized data access, making data and cloud protection crucial for ensuring confidentiality and security [5]. Organizations using cloud computing trust the CSPs for security and confidentiality, but errors in CSP-stored data can't be legally challenged [6]. Cloud security is a significant concern as users are unaware of data storage locations [6]. Online cloud users face various data attacks, making cloud data security vulnerable to hackers [7].

II. CLOUD SECURITY

Cloud Security protects cloud-stored data from unauthorized access, modifications, theft, and leakage. It is a vital aspect of cyber security for both Cloud

Service Providers (CSP) and users. CSPs must meet regulatory requirements to safeguard users' confidential information, like health insurance and banking credentials [8]. Major threats to cloud security include data breaches, insecure Application Programming Interfaces (API), user account hijacking, and service traffic hijacking. The following techniques mainly offer cloud security[9].

a. Firewalls: Firewalls prevent unauthorized access to the cloud network from a private network.
b. Obfuscation: It is a technique shielded cloud resources from harmful malware by securing the information using an intricate procedure. This process is not similar to encryption, and the details about the algorithm used to reverse the process are not well-known.
c. Penetration Testing: It is an authorized simulated attack conducted on the CSP to analyze the system's strength and weakness. This helps to understand the different attacks to which the CSP is vulnerable.
d. Virtual Private Network (VPN): The VPN hosts a private network remotely over the public network using an encryption technology that helps to integrate the data isolation feature of the private cloud along with the services and scalability of the public cloud. In this way, the traffic is sent through the public internet infrastructure securely by hiding the details.
e. Tokenization: It randomly generates a token value for the plaintext and saves its corresponding value in the database. It converts confidential information such as bank credentials and healthcare information into tokens. When the information is converted into tokens, the information significance value is protected even during the data breach. The database used to store the tokens is called a token vault, and it is often secured using encryption techniques.
f. Avoidance of public internet connections: The Public Wi-Fi network is less secure than the private or personal network because the details about the users connecting this internet are unknown. Some public Wi-Fi network retrieves many details from the users such as email id and phone number when they try to connect them. So, whenever a user tries to connect up with a public Wi-Fi network, they should always consider the privacy of their data being shared with other groups.

III. CLOUD COMPUTING ATTACKS

The cloud, despite its wide range of services, is vulnerable to both insider and outsider attacks [10]. Attackers aim to access user data and disrupt their access to cloud services, causing harm and eroding trust in cloud computing. These attacks exploit cloud vulnerabilities, such as obtaining sensitive information, accessing user accounts using stolen passwords, and acting as malicious insiders[11]. The different types of attacks that take place in the cloud environment are presented below:

a. Denial of Service (DoS): DoS attacks involve flooding the target system with redundant messages to render the service unavailable. This dangerous cloud attack can impact multiple users if a CSP is affected, as it consumes high computing power and slows down operations, denying legitimate users access to services. Distributed Denial of Service (DDoS) attacks use multiple cloud servers to block users from accessing their information.
b. Phishing Attacks: This social engineering attack aims to steal login credentials, insurance details, and bank passwords. The attacker acts as a trusted entity and sends a malicious link via email or text message. Clicking the link can inadvertently transfer sensitive information from the user's device to the attacker.
c. Port Scanning: This attack sends client requests to various server port addresses of the target system to identify an active port and exploit its weakness. It identifies many active listeners in the network by probing and use the one which is very active in satisfying their needs.
d. Hypervisor attack: In this attack, the user exploits cloud application vulnerabilities using various hypervisors like create, clone, migrate, etc. Using a compromised hypervisor, the attacker can target every Virtual Machine (VM) in the host or a group of servers. The complexity grows with the number of compromised servers.
e. Man in the middle attack: In this attack, the attacker intercepts communication between the user and the cloud service to eavesdrop or compromise either party

and obtain personal details like passwords and bank credentials. They remain hidden, making the exchange seem normal. The SaaS layer of the cloud, banking apps, business, and e-commerce sites are their primary targets, where user logins are required. The retrieved information can be used for illegal fund transfers, identity fraud, and unauthorized password changes.

f. Malware Injection attack: This attack controls user information in the cloud by adding infected service implementation to the SaaS or PaaS layer. Once the cloud server is compromised, all user requests are directed to the hacker. Malicious code execution leads to theft and eavesdropping of user information. Commonly conducted attacks include cross-site scripting and SQL injection.

IV. INTRUSION DETECTION SYSTEM

Intrusion compromises information confidentiality, availability, and integrity. Vulnerability scanners are generally not allowed in the cloud due to complex traffic distinctions. The Intrusion Detection System (IDS) identifies cloud network intrusions early and provides protection. The cloud's distributed nature makes it susceptible to exploitation, leading to significant impacts. Securing data and systems in the cloud becomes more challenging [12].

The IDS aims to identify intrusions early and prevent severe consequences. It notifies security personnel promptly about attacks, helping prevent various threats in the cloud environment. However, a disadvantage is the high false alarm rate, which may lead to unnecessary termination of communications and downtime for the cloud server.

The IDS is valuable in enhancing cloud security and protecting users' data when they lack complete control. Cloud attacks come in two forms: insider and outsider attacks. External attacks originate from outside the organization, while insider attacks occur within it, aiming to gain unauthorized access to user devices. The IDS monitors the cloud network to identify malicious activity, unauthorized entry, and illegal file modifications, detecting incidents in distinct groups.

Most cloud incidents are malicious, but some are accidental, like unauthorized access due to mistyped addresses. The IDS system's four main functionalities are surveillance, identification, investigation, and acknowledgment of intrusion events. Malicious intent of attackers is detected by analyzing acquired data. Network-based, host-based, hypervisor-based, and application-based are the four types of IDS monitoring environments[13].

a. Network-based Intrusion Detection System: Network traffic is monitored and analyzed for malicious behavior in specific network components or devices. Both internal and external traffic towards the network is examined to identify various intrusions. Once an attack is detected, necessary steps are taken to protect the cloud from further attacks. The cloud provider controls and implements the network-based IDS in the cloud, capable of identifying attacks on virtual machines and hypervisors, but not those occurring inside the virtual network.

b. Host-based Intrusion Detection System: This IDS monitors the host's state and dynamic behavior, identifying the resources used by host applications. Some hybrid approaches combine network and host-based IDS to form a flexible hybrid IDS. It detects malicious activity by monitoring inbound and outbound packets, alerting the cloud provider. The obtained information includes attack patterns and targeted users. Host-based IDS can be implemented in both virtual machines and host machines, and users can control them. Cloud providers can monitor host-based IDS deployed in the hypervisor.

c. Hypervisor-based Intrusion Detection System: The hypervisor, also called a virtual machine monitor, creates and runs virtual machines. The host machine runs the hypervisor and one or more virtual machines, which are referred to as guest machines. Hypervisor-based IDS enhances cloud security by adding protection between the abstract and its underlying host. It analyzes system metrics from cloud components to identify negative patterns, keeping the kernel free from threats.

d. Application-based Intrusion Detection System: The incidents that happen in the various application is analysed by tracking the application log files. The application-based IDS mainly receive input from data sources of the applications. It is mainly limited to specific application that needs monitoring.

Real-time intrusion detection identifies attacks while monitoring system or network traffic. Users are alerted about specific intrusion events, and precautionary measures are taken. Real-time IDS can also run in offline mode, analyzing historical data to identify past intrusions. Non-real-time IDS processing audit data may result in delays. Audit data is collected from various locations or databases, either from a single source or in a centralized manner. Intrusions can be identified using three models: misuse, anomaly, and hybrid [14].

a. Misuse Detection: It uses signatures to identify the behavior of malicious users. Based on the hackers' signatures, potential attempts to intrude the device using the same signature are recognized and predicted at the initial stage.
b. Anomaly Detection: Its main intention is to identify abnormal behavior patterns. This type of IDS has a particular threshold value for normal behavior and when a behavior exceeds this threshold value, they are automatically classified as intrusions.
c. Hybrid Intrusion Detection: By integrating the misuse and anomaly-based intrusion detection techniques, the Hybrid intrusion detection system is formed. The main aim of integrating both techniques is to identify the known and unknown attacks by misuse and anomaly detection techniques.

V. ANALYSIS OF IDS AND SECURITY IN CLOUD

IDS software automates the intrusion detection process, preventing and detecting network intrusions from various threats, including internal attacks, cybercrimes, and web threats. An effective security measure proposed by IDS [15] is critical for protecting network infrastructure. Intrusion detection involves monitoring networks or computers to identify abnormal activities caused by unauthorized entries. An efficient IDS should be free from false errors, have low overhead, uninterrupted availability, difficult to bypass, easily configurable, fault-tolerant, self-monitored, and fast. Its goal is to evaluate information systems and achieve early detection of malicious activity, reducing security risk to an acceptable level.

Limitations of an intrusion detection system based on cloud computing are as follows:

a. According to the cloud environment, the traditional IDS is inappropriate to detect suspicious or malicious activities. The ability of traditional intrusion detection systems to block and handle large malicious attacks for distributed nature of cloud computing infrastructures.
b. All the virtual machines examine traditional methods to detect network intrusion. But, most of the tasks become time-consuming and tedious.
d. Individual virtual machines launch attacks that are vulnerable to host handling. For any virtual machine, individual intrusion detection system is diverted using a host that will create complex network attack detection.
e. The anomaly and signature-based intrusion detection do not deliver efficient and robust results. The anomaly-based intrusion detection provided a high false-positive rate when signature-based IDS cannot detect unknown attacks.
f. The cloud service provider successfully deploys the intrusion detection system because the cloud architecture is distributed, scalable and dynamic in nature.
g. For intrusion detection, most of the works take a higher time to execute the process.
h. Higher data transfer costs are another major problem in cloud computing. Researchers rarely concentrate on these areas to reduce the bandwidth of the network.

VI. CONCLUSION

This chapter evaluates the works based on intrusion detection systems on cloud environments dependent on different attacks. According to the current results, the signature-based IDS, anomaly-based IDS, host-based intrusion detection system, network-based intrusion detection system are explored. Furthermore, this chapter includes reviews of articles on hypervisor-based intrusion detection systems, distributed-based intrusion detection systems, and machine learning-based intrusion detection systems.

REFERENCE

1. Sadeeq, M. M., Abdulkareem, N. M., Zeebaree, S. R., Ahmed, D. M., Sami, A. S., & Zebari, R. R. (2021). IoT and Cloud

computing issues, challenges and opportunities: A review. *Qubahan Academic Journal*, *1*(2), 1-7.

2. Alouffi, B., Hasnain, M., Alharbi, A., Alosaimi, W., Alyami, H., & Ayaz, M. (2021). A systematic literature review on cloud computing security: threats and mitigation strategies. *IEEE Access*, *9*, 57792-57807.
3. Bello, S. A., Oyedele, L. O., Akinade, O. O., Bilal, M., Delgado, J. M. D., Akanbi, L. A., ... & Owolabi, H. A. (2021). Cloud computing in construction industry: Use cases, benefits and challenges. *Automation in Construction*, *122*, 103441.
4. Chen, Y., Chen, X., Liu, W., Zhou, Y., Zomaya, A. Y., Ranjan, R., & Hu, S. (2020). Stochastic scheduling for variation-aware virtual machine placement in a cloud computing CPS. *Future Generation Computer Systems*, *105*, 779-788.
5. Priyadarshini, R., Quadir Md, A., Rajendran, N., Neelanarayanan, V., & Sabireen, H. (2022). An enhanced encryption-based security framework in the cps cloud. *Journal of Cloud Computing*, *11*(1), 64.
6. Xu, Z., Zhang, Y., Li, H., Yang, W., & Qi, Q. (2020). Dynamic resource provisioning for cyber-physical systems in cloud-fog-edge computing. *Journal of Cloud Computing*, *9*(1), 1-16.
7. Alouffi, B., Hasnain, M., Alharbi, A., Alosaimi, W., Alyami, H., & Ayaz, M. (2021). A systematic literature review on cloud computing security: threats and mitigation strategies. *IEEE Access*, *9*, 57792-57807.
8. Vinoth, S., Vemula, H. L., Haralayya, B., Mamgain, P., Hasan, M. F., & Naved, M. (2022). Application of cloud computing in banking and e-commerce and related security threats. *Materials Today: Proceedings*, *51*, 2172-2175.
9. Alam, T. (2020). Cloud Computing and its role in the Information Technology. *IAIC Transactions on Sustainable Digital Innovation (ITSDI)*, *1*(2), 108-115.
10. Kushwah, G. S., & Ranga, V. (2021). Optimized extreme learning machine for detecting DDoS attacks in cloud computing. *Computers & Security*, *105*, 102260.
11. Devi, B. T., Shitharth, S., & Jabbar, M. A. (2020, March). An Appraisal over Intrusion Detection systems in cloud computing security attacks. In *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)* (pp. 722-727). IEEE.
12. Ahmad, Z., Shahid Khan, A., Wai Shiang, C., Abdullah, J., & Ahmad, F. (2021). Network intrusion detection system: A systematic study of machine learning and deep learning approaches. *Transactions on Emerging Telecommunications Technologies*, *32*(1), e4150.\
13. Kasongo, S. M., & Sun, Y. (2020). A deep learning method with wrapper based feature extraction for wireless intrusion detection system. *Computers & Security*, *92*, 101752.
14. Aldweesh, A., Derhab, A., & Emam, A. Z. (2020). Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues. *Knowledge-Based Systems*, *189*, 105124.
15. Hady, A. A., Ghubaish, A., Salman, T., Unal, D., & Jain, R. (2020). Intrusion detection system for healthcare systems using medical and network data: A comparison study. *IEEE Access*, *8*, 106576-106584.

ID: 57

UNLEASHING THE POWER OF ARTIFICIAL INTELLIGENCE: THE FUTURE OF DIGITAL MARKETING

Mrs.M.CHITHRA
Assistant Professor,
Department of Commerce,
Sri Ramakrishna College of Arts & Science for Women,
Coimbatore, Tamil Nadu, India.

Ms.K.DARSHINI
Assistant Professor,
Department of Commerce,
Sri Ramakrishna College of Arts & Science for Women,
Coimbatore, Tamil Nadu, India.

ABSTRACT

Because of the shifting marketing landscape, tracking the customer journey has grown more difficult. With digital markets providing consumers with myriad new buying possibilities, the market is expanding and getting better. Customers express their wants, attitudes, and views through a variety of channels as demand for great customer service rises across all digital platforms. Delivering tailored information while improving the digital experience is possible with artificial intelligence (AI). This reservoir of customer-curated data that seems to never end is growing. To extract the information and exploit it, many marketers use AI. AI enables companies to gather and act on comprehensive real-time customer information, and with the help of these insights, they can create tailored digital marketing experiences. Before adopting, businesses still have a long way to go. Although there are numerous AI-based applications, many also recognise the significant advantage of incorporating AI into digital marketing strategies to create extraordinary client experiences during the purchasing process. Using the consumer decision process, this literature review examines AI-based digital marketing tools that can improve the online customer experience. It closely examines the impact of AI-integrated digital marketing on customer purchasing decisions.

KEYWORDS: *Artificial Intelligence, Digital Marketing, Virtual assistant, Chatbots, Augmented reality.*

I. INTRODUCTION:

According to Ramsbotham et al. (2017), artificial intelligence (AI) is likely to change customer interactions in digital marketing. AI is not intuitive like humans because it is data-driven. Machines can use artificial intelligence (AI), which is frequently referred to as human intelligence processes, to translate data into strategies that direct sensible customer behaviour (Haenlein & Kaplan 2019). When companies use digital marketing to change consumer behaviour, customer happiness is more likely to increase (Ramsbotham et al. 2017). Reaching the appropriate clients at the right moment is now even simpler for firms because to AI-based digital marketing (Ramsbotham et al. 2017). Unexpectedly, fostering AI innovation for the sake of assisting digital marketing results in a long-term collaboration that boosts client pleasure. Businesses are gaining from AI through the use of big data, AI-controlled chatbots, and cognitive technology outputs. Stores who use AI-powered marketing outperform conventional stores by a factor of five. Digital marketing has had a significant impact on consumer behaviour, and contemporary consumers now demand a more dependable and customised experience.

Here, we investigate how digital marketing and AI may be combined to boost client capability. Based on the AI-centred shopping experience, service quality, and trust-confidence premise, we created a model (Figure 1) that makes it easier to learn about the attitudes and behaviours of online shoppers. In order to close the gap between convenience, quality, and AI support, our model incorporates trust and confidence.

This allows us to capitalise on technology advancement and support successful marketing initiatives. This study gives firms advice on how to create customised AI experiences and makes sure that certain audiences get from the insightful data gathered using AI approaches. Additionally, it advances our knowledge of how people might work together to use AI-powered services.

II. DIGITAL MARKETING

Digital marketing is becoming more and more significant in the marketing sector. It helps create a new way of connecting with customers and ensures that they are satisfied with the end result and receive prompt assistance. Internet marketing may be more accessible to consumers than traditional marketing.

Digital Marketing Platforms:

Websites:

Businesses build websites to provide information and marketing to consumers. Websites should be developed using cutting-edge technologies and premium graphics because they have an impact on how consumers perceive a company. Websites can be used by businesses to connect with their clients, store information, and provide support (Kolesar & Wayne Galbraith, 2000).

Email:

The practise of email marketing involves sending advertisements to huge email lists of recipients. Email may enhance a business's communication with its clients, encouraging greater repeat business and client loyalty. Additionally, it helps firms attract new customers. This enables two-way communication between you and your clients. (Adikesavan 2014's)

Social Media

According to Dahnil et al. (2014), social media marketing refers to the use of social media networks to advertise goods, services, knowledge, and concepts to consumers. These technologies may be used by businesses to create content and entice social media users to spread it (Dahnil et al., 2014).

E-Commerce:

E-commerce is the practise of using communication networks to automate business transactions and processes (Al-Lami & Alnoor, 2021). On the other hand, claims categorically that the practise of trading products and services through telecommunications is what is meant by the term "e-commerce." Pay-per-click, affiliate marketing, influence marketing, blogs, and SMS marketing are further digital marketing tools (Akeel & Gubhaju, 2020).

III. OBJECTIVES

To determine how using AI in digital marketing affects customer buying behaviour and how this promotes improved marketing decisions.

Artificial Intelligence in Customer Segmentation and Targeting

AI-driven customer segmentation and targeting involve analysing vast amounts of data to identify meaningful patterns and segments. By leveraging AI algorithms, marketers can identify specific customer groups based on their purchasing behaviour, demographics, interests, and preferences. This level of granularity allows for highly targeted marketing campaigns that speak directly to the needs and desires of each segment.

For example, a clothing retailer can use AI to segment their customers based on factors such as age, gender, style preferences, and purchase history. Armed with this information, they can create personalized email campaigns featuring products that are most likely to appeal to each segment. This level of personalization not only increases the chances of conversion but also enhances the overall customer experience.

Personalization and Customization with Artificial Intelligence

Personalization and customization are key components of successful marketing campaigns. AI enables marketers to deliver personalized content and experiences at scale. By analysing customer data, AI algorithms can understand individual preferences, behaviour, and interests. This allows marketers to tailor their messaging, offers, and recommendations to each customer, increasing the likelihood of engagement and conversion.

For instance, an e-commerce website can use AI to recommend products based on a customer's browsing and purchase history. By analysing patterns and trends, AI can accurately suggest items that align with the customer's preferences. This level of personalization not only improves the shopping experience but also increases the chances of cross-selling and up-selling.

Chatbots and Virtual Assistants in Digital Marketing

Chatbots powered by AI have become increasingly popular in digital marketing. These virtual assistants can provide immediate and accurate responses to customer queries, enhancing the overall customer service experience. Chatbots are available 24/7, providing round-the-clock support and assistance to customers.

Chatbots can be programmed to handle various tasks, such as answering frequently asked questions, assisting with product recommendations, and even processing transactions. By automating these processes, businesses can improve efficiency, reduce response times, and provide a seamless customer experience. Additionally, chatbots can collect valuable data and insights about customer interactions, enabling businesses to further refine their marketing strategies.

Predictive Analytics and Machine Learning in Digital Marketing

Predictive analytics and machine learning are powerful tools in digital marketing. By analysing historical data and patterns, AI algorithms can make accurate predictions about customer behaviour and preferences. This enables marketers to anticipate customer needs, optimize marketing campaigns, and deliver highly targeted messaging.

Predictive analytics can help marketers identify which customers are most likely to churn, allowing them to implement retention strategies. Machine learning algorithms can analyse past purchase behaviour to predict future buying patterns and recommend products or services to customers. By leveraging these insights, businesses can optimize their marketing efforts and increase customer loyalty.

IV. LITERATURE REVIEW

Merendino and colleagues (2018) found in their analysis that board decisions are also impacted by the digitalization of information.

According to Kim et al. (2019), even when AI is significantly deployed, progress can be slowed down by the technology's reliance on data quality and quantity, as well as a lack of AI skills. Combining AI capabilities with moderately substantial business and marketing transcribing talents could result in a large rise in production.

This advancement in AI digital marketing organisation, benefits from mechanising various aspects of marketing and data products that can be used to support AI (Bag et al. 2021)

In digital marketing, trust and accountability in a relationship play important roles in connecting customers and retailers (Jarrahi, 2018).

When it comes to preferences, perceived value and transparency play an important role in determining opinions and conduct (Hoff & Bashir 2015).

Chatbots have identified humanity, social intellect and the presence of society, trust, skills, and usability as regards to social demands (Chopra 2020).

The social attitudes or the psychological character of a client's perception are also characterised by social elements (Keiningham et al. 2017).

Different digital marketing technologies such as augmented-virtual realities, vision-driven imaging and predictive inventory management are widely associated with AI in online businesses (Singh et al. 2019).

The Future of Artificial Intelligence in Digital Marketing

The future of digital marketing is closely intertwined with the evolution of artificial intelligence. As AI continues to advance, marketers can expect even more sophisticated tools and techniques to enhance their strategies.

One area that holds immense potential is natural language processing (NLP). NLP enables AI systems to understand and interpret human language, allowing for more effective communication and engagement. This opens up possibilities for voice-activated search, chatbot conversations that mimic human interaction, and personalized content delivery based on real-time conversations.

Another promising area is the integration of AI with augmented reality (AR) and virtual reality (VR). By combining AI algorithms with AR/VR technology, marketers can create immersive and personalized experiences for their customers. For example, AI-powered AR applications can allow customers to virtually try on clothing or visualize how furniture would look in their homes, enhancing the overall shopping experience.

Implementing Artificial Intelligence in your Digital Marketing Strategy

To implement artificial intelligence in your digital marketing strategy, it's essential to start with a solid foundation of data. AI relies on data to generate insights and make accurate predictions, so ensure that you have a robust data collection and management system in place.

Next, identify areas in your marketing strategy where AI can add value. This could include customer segmentation and targeting, content creation, campaign optimization, or customer service. Research and evaluate AI tools and platforms that align with your specific needs, and consider partnering with experts or agencies with AI expertise.

Once you have selected the right AI tools, ensure that your team has the necessary skills and knowledge to leverage them effectively. Provide training and support to help your team understand how to interpret AI-generated insights and integrate them into their workflows. Regularly analyse and evaluate the results of your AI-powered initiatives, and make adjustments as needed.

Artificial intelligence tools for digital marketing

There is a wide range of AI-powered tools and platforms available for digital marketers. These tools can automate various aspects of marketing, provide valuable insights, and enhance the overall customer experience. Here are some popular AI tools for digital marketing:

- **Chat fuel:** A platform that allows you to easily create and deploy chatbots on popular messaging platforms like Facebook Messenger. Chat fuel's AI-powered chatbots queries, provide recommendations, and even process transactions.
- **Optimizely:** An experimentation platform that uses AI to optimize customer experiences. It enables marketers to run A/B tests, personalized content, and make data-driven decisions to improve conversion rates.
- **Ad text:** A platform that uses AI algorithms to optimize digital advertising campaigns. Ad text analyses campaign performance data in real-time and automatically adjusts ad budgets and targeting to maximizing ROI.
- **Convers Ica:** An AI-powered sales assistant that engages with leads via email and chat. Convers Ica virtual assistant can qualify leads, schedule meetings, and follow up with prospects, freeing up sales team to focus on closing details.
- **Phrasee:** An AI-powered copywriting tool that generates engaging and effective email subject lines, social media posts and ad copy. Its algorithms analyse language patterns to create compelling marketing messages.

Findings

AI is reshaping the future of digital marketing and contributes to consumer trust-building and individualised experiences.

V. CONCLUSION:

Digital marketing is still very young, but AI has the potential to have a greater impact on customer behaviour. With the help of a new marketing playbook called AI marketing, businesses may more successfully transition from marketing automation to marketing customization. In recent years, the impact of AI on digital marketing has intensified, enabling marketers to more fully customise their sales and digital marketing strategies. The abundance of data accessible has allowed marketers to tailor their sales and marketing initiatives and go above and beyond what their clients had anticipated. AI engineering has the power to alter how goods and services are offered to clients. Finally, we can state that information used to analyse consumer behaviour produces extremely predicative findings, making digital marketing automation more dynamic than ever. Businesses can identify their target thanks to AI. Customers on digital marketing platforms are aware of the demands and preferences of their customers, and they increased openness. Live chat is connected with AI tools from digital marketing platforms via Chatbots that draw in customers by quickly and simply answering their questions using human-produced data and artificial digital marketing intelligence tools, Businesses may increase good, individualised client experiences by fostering trust on digital channels. Encounters gained by a deep dive.

As AI continues to advance, marketers must stay informed and adapt their strategies to harness its full potential. By embracing AI in their digital marketing efforts, businesses can stay ahead of the curve, drive better results, and create meaningful connections with their customers. So, buckle up and get ready to unleash the power of artificial intelligence in the future of digital marketing. The possibilities are endless!

VI. REFERENCES:

1. Bag, S., Gupta, S., Kumar, A., & Sivarajah, U. (2021). An integrated artificial intelligence framework for knowledge creation and B2B marketing rational decision making for improving firm performance. Industrial Marketing Management, 92, 178-189.
2. Chopra, S.S. (2020). Helping Entrepreneurs and Small Businesses Make the Digital Transformation. In The Evolution of Business in the Cyber Age (pp. 39-51). Apple Academic Press.
3. Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. California management review, 61(4), 5-14.
4. Hoff, K.A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. Human factors, 57(3), 407-434.
5. Jarrahi., M.H. (2018), 'Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making', Business Horizons, vol. 61, no. 4, pp. 577-586.
6. Keiningham, T., Ball, J., Benoit, S., Bruce, H.L., Buoye, A., Dzenkovska, J., & Zaki, M. (2017). The interplay of customer experience and commitment. Journal of Services Marketing.
7. Kim, Y., Kim, C.K., Lee, D.K., Lee, H.W., & Andrada, R.I.T. (2019). Quantifying nature-based tourism in protected areas in developing countries by using social big data. Tourism Management, 72, 249-256.
8. Merendino, A., Dibb, S., Meadows, M., Quinn, L., Wilson, D., Simkin, L., & Canhoto, A. (2018). Big data, big decisions: The impact of big data on board level decision-making. Journal of Business Research, 93, 67-78.
9. Ransbotham, S., Kiron, D., Gerbert, P., & Reeves, M. (2017). Reshaping business with artificial intelligence: Closing the gap between ambition and action. MIT Sloan Management Review, 59(1).
10. Singh, J., Flaherty, K., Sohi, R.S., Deeter-Schmelz, D., Habel, J., Le Meunier-FitzHugh, K., & Onyemah, V. (2019). Sales profession and professionals in the age of digitization and artificial intelligence technologies: concepts, priorities, and questions. Journal of Personal Selling & Sales Management, 39(1), 2-22.

ID: 58

BIO MEDICAL AND MEDICAL IMAGING

Ms.V.YOGASHRI
Teaching Assistant,
Department of Information Technology,
Rathinam College of Arts and Science (Autonomous),
Coimbatore, Tamil Nadu, India.

Mr.M.SURIYA
Assistant Professor,
Department of Computer Science,
Rathinam College of Arts and Science (Autonomous),
Coimbatore, Tamil Nadu, India.

ABSTRACT:
In this article, we examine the advancements in the field of biomedical imaging, which is constantly evolving and utilizes various techniques for generating images of the human body. These images are crucial for disease diagnosis, surgical planning, and treatment evaluation.

The paper primarily focuses on three important aspects:
1. Imaging modalities: We explore different types of imaging modalities, such as X-ray, ultrasound, MRI, and CT scans. And multimodal imaging techniques.
2. Image processing techniques: We delve into the various methods used for processing medical images, including segmentation, registration, and enhancement.
3. Applications of biomedical imaging: We discuss the numerous applications of biomedical imaging, ranging from cancer detection and cardiovascular disease diagnosis to surgical planning.

Lastly, we discuss the challenges and opportunities associated with biomedical imaging. We strongly believe that this field holds tremendous potential to revolutionize the healthcare industry by equipping doctors with essential tools for accurate and efficient medical care.

KEYWORDS: *Biomedical Advancements, Medical Imaging.*

I. INTRODUCTION

Biomedical imaging is an ever-evolving domain that employs various techniques to generate visuals of the human body. These visuals play a crucial role in diagnosing illnesses, organizing surgeries, and evaluating the efficacy of treatments.

Recent years have witnessed remarkable progress in the field of biomedical imaging, largely thanks to several contributing factors. These include the emergence of novel imaging methods like MRI and CT scans, enhancements made to established techniques like ultrasound and X-rays, the introduction of advanced image processing techniques such as segmentation and registration, and the incorporation of artificial intelligence (AI) to enhance imaging accuracy and efficiency.

The recent developments have enabled capturing detailed and accurate images of the human body like never before. This has resulted in various advantages for patients, including:
- Prompt detection of diseases.
- Enhanced assessment of diseases' severity.
- Improved surgical planning with greater precision.

- Better monitoring of treatment effectiveness.

The prospects for biomedical imaging appear promising. As the field continues to advance, we can anticipate even more revolutionary and exceptional uses of this technology. These applications have the capacity to completely transform healthcare and enhance the quality of life for countless individuals.

II. LATEST ADVANCEMENTS IN BIOMEDICAL IMAGING

1. **Multimodal imaging:** This is the use of multiple imaging modalities to create a more complete picture of the anatomy or physiology of a patient. For example, a patient with a brain tumour might be imaged with MRI, CT, and PET scans.
2. **Molecular imaging:** This is the use of imaging techniques to detect and visualize molecular markers of disease. In vivo imaging: This is the use of imaging techniques to image the body in its natural environment. In vivo imaging allows doctors to see how the body is functioning in real time.
3. **3D imaging:** This is the use of imaging techniques to create three-dimensional images of the body.
4. **Artificial intelligence (AI):** AI is being used to develop new imaging techniques and to improve the accuracy of existing techniques. For example, AI can be used to segment images, to detect tumors, and to predict the risk of disease.

Multimodal Imaging:

In neurology, multimodal imaging can be used to study the structure and function of the brain in patients with disorders such as Alzheimer's disease or traumatic brain injury. By combining structural imaging modalities like MRI scans with functional imaging modalities like PET scans, researchers can better understand the changes that occur in the brain and develop more targeted treatment approaches.

In oncology, multimodal imaging can be used to accurately diagnose and stage tumors. For example, combining CT scans and PET scans can provide information about both the location and activity of a tumour, helping doctors determine the best course of treatment.

In cardiology, multimodal imaging can be used to assess the structure and function of the heart. By combining imaging modalities such as echocardiography, MRI, and angiography, doctors can obtain a more comprehensive evaluation of the patient's cardiac health.

In orthopaedics, multimodal imaging can be used to evaluate musculoskeletal injuries or conditions. For example, combining X-rays, CT scans, and MRI scans can help doctors assess the extent of a fracture or the presence of soft tissue damage.

Overall, multimodal imaging plays a crucial role in modern medicine by providing a more complete picture of a patient's condition. It allows for more accurate diagnoses, personalized treatment plans, and better monitoring of treatment effectiveness. As technology continues to advance, the possibilities for multimodal imaging will only continue to expand, further improving patient care.

Molecular Imaging:

The integration of molecular imaging with other diagnostic modalities: Molecular imaging can complement other imaging techniques such as MRI or CT scans by providing additional molecular information. This integration will enhance the diagnostic capabilities of these techniques and lead to more comprehensive and accurate diagnoses.

Advancements in molecular imaging technologies: New technologies are being developed to improve the speed, resolution, and depth of molecular imaging. For example, advances in positron emission tomography (PET) scanners have allowed for higher resolution images, while developments in fluorescence imaging have enabled real-time imaging of molecular events at the cellular level.

The use of molecular imaging in assessing treatment response: Molecular imaging can be used to monitor the response of tumors or disease markers to treatment. This allows doctors to gauge the effectiveness of therapies and make individualized treatment decisions.

In conclusion, the future of molecular imaging holds great promise in revolutionizing healthcare. With the development of new probes, imaging agents, and technologies, as well as the integration of molecular imaging with other diagnostic modalities, doctors will have access to earlier and more accurate diagnoses, leading to more personalized treatments for patients.

3D Imaging:

Diagnosis, treatment planning, and monitoring of diseases and conditions. It allows physicians to visualize internal structures and abnormalities with greater detail and accuracy.

In addition to medical applications, 3D imaging is also widely used in engineering, architecture, virtual reality, and entertainment industries. It enables designers and engineers to create realistic models and simulations, aiding in product development and design optimization.

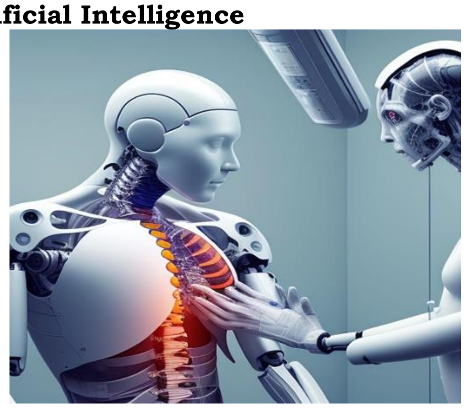
One notable advantage of 3D imaging is its ability to provide a more comprehensive understanding of an object or structure. By capturing multiple perspectives and layers, it offers a more holistic view, allowing for more accurate analysis and interpretation.

Furthermore, 3D imaging has revolutionized the field of archaeology and palaeontology. By digitally reconstructing artifacts, fossils, and ancient sites, researchers can study and preserve these objects without causing any damage.

With advancements in technology, 3D imaging techniques continue to evolve and improve. Higher resolution, faster processing, and improved algorithms contribute to more detailed and precise images.

In conclusion, 3D imaging is a powerful tool that has transformed various industries, including medicine, engineering, and archaeology. Its ability to create realistic, three-dimensional representations of objects and structures opens up new possibilities for research, analysis, and visualization. As technology continues to advance, the future of 3D imaging holds even greater potential for innovation in medical assistance and technology.

Artificial Intelligence



AI is being used in medical engineering for various purposes. It is used in medical imaging to develop new techniques and improve accuracy. AI can also be used to detect tumors and predict disease risk.

In drug discovery, AI helps in screening compounds and predicting drug toxicity. Personalized medicine is another area where AI is used to tailor treatments to individual patients' needs.

Robotics powered by AI are being developed for performing surgeries with precision. Virtual reality and augmented reality applications are being developed for medical training and education. AI is also improving healthcare administration by automating tasks and making better decisions. These are just a few examples of how AI is being used in medical engineering.

Companies are using AI for drug development, genetic analysis, surgical robotics, and VR/AR applications. As AI technology advances, we can expect more innovative applications in healthcare.

III. CONCLUSION:

On Conclusion, the biomedical imaging field is experiencing rapid advancement and has the capability to transform healthcare. Continuous development of imaging techniques enables us to capture detailed and precise images of the human body, leading to enhanced diagnosis, treatment, and disease monitoring.

The prospects for biomedical imaging are promising. As the field progresses, we anticipate witnessing more revolutionary and inventive applications of this technology. These applications possess the potential to positively impact millions of individuals' lives.

IV. REFERRENCE:

1. Shenghui, X., Hua, Y., Jingjuan, Q., Fan, P., Jie, J., & Kendra, H. (2015). Protein MRI contrast agent with unprecedented metal selectivity and sensitivity for liver cancer imaging. Proceedings of the National Academy of Sciences of the United States of America, 112(21), 6607–6612. doi:10.1073/pnas.1423021112 PMID:25971726
2. Thapa, D., Raahemifar, K., Bobier, W. R., & Lakshminarayanan, V. (2014). Comparison of super-reso-lution algorithms applied to retinal images. Journal of Biomedical Optics, 19(5), 056002. doi:10.1117/1.JBO.19.5.056002 PMID:24788371
3. Visage Imaging. (2015). Retrieved June 12, 2015, from www.visageimaging.com

/visage-imaging-receives-fda-clearance-visage-ease-pro/Wang, T., McElroy, A., Halaney, D., Vela, D., Fung, E., & Hossain, S. etal.. (2015).

4. Yamamoto, S., Maki, D., Korn, R. L., & Kuo, M. D. (2012). Radio genomic analysis of breast cancer using MRI: A preliminary study to define the landscape. AJR. American Journal of Roentgenology, 199(3), 654–663. doi:10.2214/AJR.11.7824 PMID:22915408
5. Ziegler, S. I. (2005). Positron Emission Tomography: Principles, Technology, and Recent Developments. Nuclear Physics. A., 752, 679c–687c. doi:10.1016/j.nuclphysa.2005.02.067F DA. (2004). Guidance for industry, Part III. Retrieved June 10, 2015, from http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DevelopmentResources/ucm092895.htm Gonzalez, R. C., & Woods, R. E. (2001). Digital Image Processing.
6. FDA. (2004). Guidance for industry, Part III. Retrieved June 10, 2015, from http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DevelopmentResources/ucm092895.htmGonzalez, R. C., & Woods, R. E. (2001). Digital Image Processing. Upper Saddle River, NJ: Prentice Hall Inc. Intermountain Medical Centre. (n.d.). New imaging tool to diagnose heart conditions is dramatically more accurate, less expensive and safer. ScienceDaily. Retrieved March 16, 2015, from www.science-daily.com/releases/2015/03/150316113323. htm
7. Joshua, E. (2012). Iron oxide nanoparticles for targeted cancer imaging and diagnostics. Nanomedicine; Nanotechnology, Biology, and Medicine, 8(3), 275–290. doi:10.1016/j.nano.2011.08.017 PMID:21930108
8. Kase, K. R. (2004). Radiation protection principles of NCRP. Health Physics, 87(3), 251–257. doi:10.1097/00004032-200409000-00005 PMID:15303061
9. Kulbago, T. (2013). Imaging in Medical Device Manufacturing. Retrieved June 15, 2015, from http://www.bonezonepub.com/component/content/article/742-the-use-of-imaging -analytics-in-medical-device-manufacturingLang, E. K., Thomas, R., Davis, R., Shore, B., Ruiz-Deya, G., & Macchia, R. J. (2004). Multiphasic helical CT diagnosis of early medullary and papillary necrosis. Journal of Endourology, 18, 167–171. doi:10.1089/089277904322959815 PMID:15006054
10. Nabeel, U. A. (2013). New Innovations in Medical Imaging Technology: Live from RSNA. Retrieved December 18, 2013, from http://in-training.org/new-innovations-in-medical-imaging-technology-live-from-rsna-3961Radiological Society of North America. (2012). Fetal alcohol exposure affects brain structure in children. Science Daily.
11. RetrievedNovember25,2012,fromwww.sciencedaily.com/releases/2012/11/121125103949.htmRadiological Society of North America. (2015). Retrieved June 12, 2015, from http://www.sciencedaily.com/releases/2015/05/150527095440.htm Rakow-Penner, R. A., White, N. S., Parsons, J. K., Choi, H. W., Liss, M. A., Kuperman, J. M., & Dale, A. M. etal. (2015). Novel technique for characterizing prostate cancer utilizing MRI restriction spectrum imaging: Proof of principle and initial clinical experience with extra prostatic extension.

ID: 59

E-PASSPORT SCHEME USING CRYPTOGRAPHIC ALGORITHMS WITH FACE AND IRIS BIOMETRICS

Dr.S.PRASATH
Assistant Professor,
School of Computer Science,
VET Institute of Arts and Science (Co-Edu) College,
Erode, Tamil Nadu, India.

Dr.V.K.NARENDIRA KUMAR
Assistant Professor, s
PG & Research Department of Computer Science,
Gobi Arts & Science College (Autonomous),
Gobichettipalayam, Erode, Tamil Nadu, India.

ABSTRACT:
Advancements in technology have created the possibility of greater assurance of proper travel document ownership, but some concerns regarding security and effectiveness remain unaddressed. Electronic passports have known a wide and fast deployment all around the world since the International Civil Aviation Organization the world has adopted standards whereby passports can store biometric identifiers. The use of biometrics for identification has the potential to make the lives easier, and the world people live in a safer place. The purpose of biometric passports is to prevent the illegal entry of traveler into a specific country and limit the use of counterfeit documents by more accurate identification of an individual. This paper analyses the face and iris biometric e-passport design. This papers focus on privacy and personal security of bearers of e-passports, the actual security benefit countries obtained by the introduction of e-passports using face and iris recognition systems. Researcher analyzed its main cryptographic features; the face and iris biometrics currently used with e-passports and considered the surrounding procedures. Researcher focused on vulnerabilities since anyone willing to bypass the system would choose the same approach. The finding suggests that a security benefit can be achieved border control procedures are merely augmented with authentication based on e-passport. On the contrary, solely relying on them may pose a risk that did not exist with previous passports and border controls. The paper also provides a security analysis of the face and iris biometric using e-passport that is intended to provide improved security in protecting biometric information of the e-passport bearer.

KEYWORDS: *E-Passport, Face, Iris, Biometrics, and Recognition.*

I. Introduction

An electronic passport (e-Passport) is an identification document which possesses relevant biographic and biometric information of its bearer. It also has embedded in it a Radio Frequency Identification (RFID) Tag which is capable of cryptographic functionality. The successful implementation of biometric technologies in documents such as e-Passports aims to strengthen border security by reducing forgery and establishing without doubt the identity of the documents' bearer. Biometrics is measurable characteristics of an individual used to identify him or her. Biometric systems can function in verification or identification modes depending on their intended use. In a verification task, a person presents an identity claim to the system and the system only needs to verify the claim. In an identification task, an unknown individual presents himself or herself to the system, and it must identify them. In general, there are three approaches to authentication. In order of least secure and least convenient to most secure and most convenient, they are: Something you have - card, token, key. Something you know- PIN, password. Something you are - biometric [1].

1.1. Contributions

Through this paper provide an introduction to the three constituent technologies in E-Passports: Biometric, RFID, and Public Key Infrastructure. Researcher also effectively summarizes the contents of three technical reports which describe the protocols and the functioning of the ICAO e-Passport specifications. This is the first work that analyses the protocols behind the e-Passport. Researcher also presents some feasible threats to the e-Passport protocol.

1.2. Purpose of the Study

The primary objective of the study is to produce new knowledge with respect to security of face and iris biometric techniques in a national border control setting. The results of the work should be useful for those making design decisions with respect to biometric technologies in an e-passport setting or other large-scale applications.

II. Literature Survey

Juels *et al* (2005) discussed security and privacy issues that apply to e-passports. They expressed concerns that, the contact-less chip embedded in an e-passport allows the e-passport contents to be read without direct contact with an IS and, more importantly, with the e-passport booklet closed. They argued that data stored in the chip could be covertly collected by means of "skimming" or "eavesdropping". Because of low entropy, secret keys stored would be vulnerable to brute force attacks as demonstrated by Laurie (2007). Kc and Karger (2005) suggested that an e-passport may be susceptible to "splicing attack", "fake finger attack" and other related attacks that can be carried out when an e-passport bearer presents the e-passport to hotel clerks. There has been considerable press coverage (Johnson, 2006; Knight, 2006; Reid, 2006) on security weaknesses in e-passports. These reports indicated that it might be possible to "clone" an e-passport.

2.1. Technical Challenges

A system of international travel requires some degree of interoperability, and ICAO guidelines have provided some of the framework for that interoperability. Individual states always remain cautious in entering into any international agreement, however, fearing that they will lose some of their individual rights. In its efforts to maintain national sovereignty over passports, the NTWG avoided touching on several implementation issues, particularly in the security arena. These questions are essential to ensuring interoperability between national systems now that advanced electronics are being deployed. In order for widespread deployment to occur, though, more vendors will need to reach similar performance levels with increased accuracy and detection rates [8].

2.2. Domestic Challenges

The electronic passport is secure will prove substantially more difficult than actually securing it biometric technology in passports. It is quite clear, however, that contactless chips offer significant advantages, including larger capacities and lower costs. The technology also has yet to experience widespread deployment in either the private or public sector, though such deployment can be expected in the private sector in the next few years. Contact-based chips simply lack the robustness of contactless technology. A lack of available barcodes, in addition to the fact that RFID is a superior tracking technology compared to virtually any available, has led major retailers like Walmart to investigate inclusion of RFID in its supply chain. As this deployment occurs, RFID may also become an integral part of numerous other everyday tasks, such as entering a place of work or making a credit card transaction.

2.3. International Challenges

Electronic passport rollout plans continue to move forward, but some many countries, continue to lag behind. Certainly, the ongoing debate over how best to protect data stored on the RFID tags may be preventing some nations from moving forward. Given the complex nature of the project and the need for it to remain static for a substantial number of years, the international community would be well served to take the time necessary to implement it correctly. Many countries to begin issuing e-passports were too soon, and delayed it by one year. However, such a delay will require Congressional action [8].

2.4. Face Recognition

The first task needed after the capture of an image is an initial alignment. The features commonly used to identify the orientation and location of the face is the eyes, nose, and mouth. This approach is the standard used on most facial biometric algorithms. After this stage, processing varies based on whether the application is identification or verification. Identification is the process of determining who someone is. Verification only needs to confirm that a subject is the person they claim to be [9]. In identification, the system compares the captured image (probe) to the gallery. The type of comparisons made depends both on the biometric used and on the matching algorithm in question. After the comparison, the system returns a rank ordering of identities.

2.5. Iris Recognition

Iris recognition technology is based on the distinctly colored ring surrounding the pupil of the eye. Made from elastic connective tissue, the iris is a very rich source of biometric data, having approximately 266 distinctive characteristics. These include the trabecular meshwork, a tissue that gives the appearance of dividing the iris radially, with striations, rings, furrows, a corona, and freckles. Iris recognition technology uses about 173 of these distinctive characteristics. Iris recognition can be used in both verification and identification systems. Iris recognition systems use a small, high-quality camera to capture a black and white, high-resolution image of the iris. The systems then define the boundaries of the iris, establish a coordinate system over the iris, and define the zones for analysis within the coordinate system [3].

3. Biometric System Modules

- ❖ Enrollment Unit: The enrollment module registers individuals into the biometric system database. During this phase, a biometric reader scans the individual's biometric characteristic to produce its digital representation **(see figure 1)**.
- ❖ Feature Extraction Unit: This module processes the input sample to generate a compact representation called the template, which is then stored in a central database or a smartcard issued to the individual.
- ❖ Matching Unit: This module compares the current input with the template. If the system performs identity verification, it compares the new characteristics to the user's master template and produces a score or match value (one to one matching). A system performing identification matches the new characteristics against the master templates of many users resulting in multiple match values (one too many matching).
- ❖ Decision Maker: This module accepts or rejects the user based on a security threshold and matching score [1].

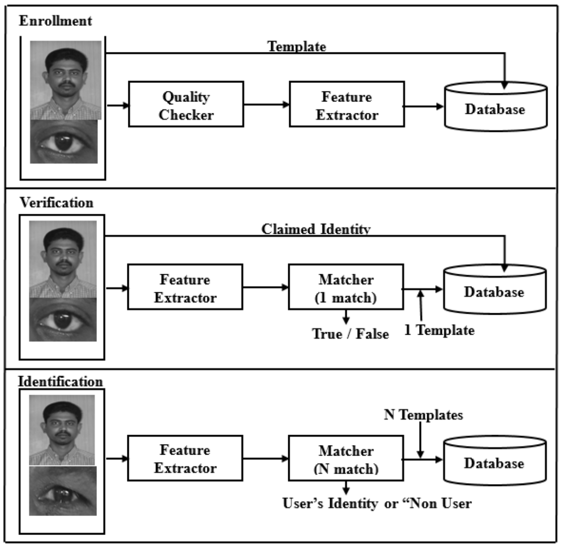


Figure 1 Block diagrams of Enrollment, Verification, an Identification tasks are shown using the modules of a face and iris biometric electronic passport system.

IV. E-Passport System Design

System design is a transition from a user-oriented document to a document oriented to programmers or database personnel. It goes through logical and physical design walkthrough before implementation.

4.1. Logical Data Structure

The ICAO issued a standardized data structure called Logical Data Structure (LDS) for the storage of data elements. This was to ensure that global interoperability for e-Passport Tags and Readers could be maintained. The specifications state that all the 16 data groups are write protected and can be written only at the time of issue of the e-Passport by the issuing state shown in table 1. A hash of data groups 1-15 are stored in the security data element (SOD), each of these hashes should be signed by the issuing state.

Table 1 E-Passport Logical Data Structure

Data Group	Data Element
DG 1	Document Details
DG 2	Encoded Headshot
DG 3	Encoded Face biometrics
DG 4	Encoded Iris biometrics
DG 5	Displayed Portrait

ISBN: 978-81-965283-7-9

DG 6	Reserved for Future Use
DG 7	Signature
DG 8-10	Data features
DG 11-13	Additional Details
DG 14	CA Public Key
DG 15	AA Public Key
DG 16	Persons to Notify
SOD	Security Data Element

4.2. E-Passport Certification

The Biometric authentication procedure for electronic passports involves two processes - Registration and Verification. During the registration phase, the e-Passport applicant registers their biometric at a secure location under human supervision. A feature extraction program is used to encode this biometric data after which it is stored on the user's e-Passport Tag. For user authentication and identity Verification at an inspection terminal, the user is made to supply a sample of their biometric. The same feature extraction algorithm is used to encode the freshly supplied biometric. A matching algorithm is run at the terminal to obtain the degree of similarity between the registered and supplied biometric. If the degree of similarity is deemed to be greater than a certain threshold value, the biometric is accepted and the user's identity is verified successfully. Unfortunately, without human supervision, it is not always possible to detect the use of prosthetics at the biometric registration or Verification stages. It is easy to see that biometric spoofing attacks will become easier to perform as automation increases and human supervision of the biometric process decreases [4].

4.3. Encryption

Since the data stored in the passport is highly confidential, the Contactless IC chip must have mechanisms for protection and integrity of the data. The ICAO recommended that the passport should have the following properties [6]:

- A cryptographic checksum is used to protect data integrity. The system can detect if data has been altered by comparing the checksum in the passport against the real-time computation of the stored data.
- A digital watermark is used to protect the integrity of facial and iris image. Some digital bits may be buried into an image for further verification purposes without degrading the quality of the image.
- Unique IC chip serial numbers are used to prevent cloning of chips.
- Symmetric or asymmetric secret keys can be used to ensure data privacy. Passport-issuing countries have the option not to encrypt the data.
- A Public Key Infrastructure (PKI) for generation and management is required.

V. Implementation of E-Passport System

In order to implement this biometric electronic passport system for person identification using face and iris recognition efficiently, ASP.NET program is used. This program could speed up the development of this system because it has facilities to draw forms and to add library easily.

5.1. Public Key Infrastructure

In normal situations, certificate-issuing organizations known as Certificates Authorities (CA's) are grouped in a trusted hierarchy, where the children CA's trust the parent CA's. All CA's directly or indirectly trust the top-level Root CA. Revoking one certificate means all its children CA's are no longer trusted. However, in ICAO, when a private key is compromised, the country cannot automatically invalidate all the passports issued with this key. The passport signed by any private key is expected to last for the issuing period. It is not feasible to ask hundreds or even thousands of passport holders to renew their passports every time a key is revoked. Instead, these passports should be used as normal, and a mechanism should notify the custom officials inspect the passport in greater detail. For each country such as the US, there is a Country Signing CA responsible for creating a public/private key pair, which is used to sign the Document Signer Certificates. This key pair should be generated and stored in a highly protected, offline CA infrastructure by the issuing country [5]. The lifetime of a Country Signing CA Key should be the longer of: The length of time the key will be used to issue passports. The lifetime of the passport issued by the key.

To ensure security, the ICAO recommended the countries to replace the CA key every 3-5 years [8]. Under each country, there are numerous passport-issuing offices. Each of them is a Document Signer with a public/private key pair and has a Document Signer Certificate. Each passport is signed by the Document Signer Certificate to ensure data integrity. In order to avoid large amount of passports with invalid keys when a Document Signer Certificate Key is revoked, the suggested lifetime of the key should be about three months, less if the office issue a lot of passports per period of time. If a key or a certificate needs to be revoked, the Country CA must communicate bilaterally to all other countries and to the ICAO Public Key Directory within 48 hours [8]. In addition, a full revocation list should be exchanged every 90 days. All the private keys of Document Signer is stored in the passport-issuing office, where as the public key is stored in the ICAO Public Key Directory. The directory is a central source used to distribute the public key to the participating countries. Each participant country is responsible for downloading the latest version of the keys and making sure passports are indeed signed by the Document Signer.

5.2. Passive Authentication

Passive Authentication is the only mandatory cryptographic protocol in the ICAO. Its primary goal is to allow a Reader to verify that the biometric face and iris data in the e-Passport is authentic. This scheme is known as passive authentication since the Tag performs no processing and is only passively involved in the protocol. One must note that Passive Authentication does not tie the Tag to a passport i.e. researcher can only establish that the face and iris data on the Tag is correct, not the authenticity of the Tag itself. The Inspection System retrieves the certificate of the issuing document verifier; using the public key from the certificate it verifies the digital signature and biometric used to sign the biometric face and iris data. Once the validity of the signature is established, the Reader computes the hash of each of these data elements and compares them with the hashed values stored. If there is a match, it can be established that the data on the Tag was not manipulated [7].

5.3. Active Authentication

Active Authentication is an optional protocol in the ICAO specifications. Using a simple challenge-response mechanism, it aims to detect if a Tag has been substituted or cloned. If Active Authentication is supported, the Tag on the e-Passport stores a public key (KP_{uAA}) in Data and its hash representation. The corresponding private key (KP_{rAA}) is stored in the secure section of Tag memory. In order for the Tag to establish its authenticity, it must prove to the Reader that it possess this private key.

- The Reader sends a randomly generated 64 bit string (R) to the Tag.
- The Tag signs this string using the key KP_{rAA} and sends this signature to the Reader.
- The Reader obtains the public key KP_{uAA} stored in biometric Data.
- The Reader verifies the correctness of the signed string using its knowledge of R and KP_{uAA}.

5.4. Basic Access Control

Basic Access Control (BAC) is an optional protocol that tries to ensure that only Readers with physical access to the passport can read Tag data. When a reader attempts to scan the BAC enabled e-Passport, it engages in a protocol which requires the Reader to prove knowledge of a pair of secret keys (called `access keys') that are derived from biometric data on the Machine Readable Zone (MRZ) of the passport. From these keys, a session key which is used for secure messaging is obtained [8].

5.5. Chip Authentication

The Chip Authentication protocol is aims to replace Active Authentication as a mechanism to detect cloned e-Passports. If CA is performed successfully it establishes a new pair of encryption and MAC keys to replace BAC derived session keys and enable secure messaging. It does this using the static key agreement protocol. Note that the e-Passport Tag already has a Chip Authentication public key and private key (in secure memory).

5.6. Terminal Authentication

The Terminal Authentication protocol is a protocol that is executed only if access biometrics data is required. It is a challenge-response mechanism that allows the Tag to validate the Reader used in Chip Authentication. The Reader proves to the Tag using digital certificates that it has been authorized by the home and visiting nation to read e-Passport Tags.

5.7. E-Passport Protocols

The ICAO e-passport is a complex protocol suite that consists of three sub protocols namely, BAC, PA and AA. Such a protocol suite is not only difficult to formalize, but also verification of such systems more often leads to an exponential state-space explosions. Researcher model the flow of e-passport protocol according to the following stages:

- When an e-passport is presented at a border security checkpoint, the chip and the e-passport reader execute the BAC protocol, in order to establish a secure (encrypted) communication channel between them.
- On successful completion of BAC, the e-passport reader performs PA.
- The chip and the e-passport reader execute the AA protocol.

The e-passport authentication heavily relies on PKI. Researcher model only one level of certification hierarchy, up to the document signer and researcher assume that document signer public key is certified by its root (country signing authority) and, is valid and secure. This does not weaken the verification process of the e-passport protocol suite, but only indicates that the model assumes the "ideal" PKI implementation. Researcher also supposes that cryptographic primitives and face and iris biometric used in the system like hash functions, MAC, and generation of keys are secure [8].

VI. CONCLUSIONS

The work represents an attempt to acknowledge and account for the presence on e-passport using face and iris recognition, towards their improved identification. The application of biometric in passports requires high accuracy rates; secure data storage, secure transfer of data and reliable generation of biometric data. The passport data is not required to be encrypted, identity thief and terrorists can easily obtain the biometric information. The discrepancy in privacy laws between different countries is a barrier for global implementation and acceptance of biometric passports. A possible solution to un-encrypted wireless access to passport data is to store a unique cryptographic key in printed form that is also obtained upon validation. The key is then used to decrypt passport data and forces thieves to physically obtain passports to steal personal information. More research into the technology, additional access and auditing policies, and further security enhancements are required before biometric recognition is considered as a viable solution to biometric security in passports. The adversaries might exploit the passports with the lowest level of security. The inclusion of biometric identification information into machine readable passports will improve their robustness against identity theft if additional security measures are implemented in order to compensate for the limitations of the biometric technologies.

VII. REFERENCES

1. A.K.Jian, "Biometrics personal identification in networked society" Technical report 1999.
2. C.Hesher, "A novel technique for face recognition using range images" in the Proceedings of 7th International Symposium on Signal Processing and Its Application, 2003.
3. HOME AFFAIRS JUSTICE (2006), "EU standard specifications for security features and biometrics in passports and travel documents", Technical report, European Union.
4. ICAO (2006), "Machine readable travel documents", Technical report, ICAO.
5. KLUGLER, D. (2005), "Advance security mechanisms for machine readable travel documents, Technical report", Federal Office for Information Security (BSI), Germany.
6. ICAO, "Machine Readable Travel Documents", Part 1 Machine Readable Passports, 5th Edition, 2003
7. Riscure Security Lab, "E-Passport privacy attack", Cards Asia Singapore, April 2006.
8. D. Wagner, "Security and privacy issues in e-passports", Cryptology ePrint, Report 2005.
9. ICAO, "Biometrics Deployment of Machine Readable Travel Documents", Version 2.0, May 2004.

ID: 60

IOT BASED A FARMER FRIENDLY AGRICULTURE AUTOMATION SYSTEM USING NODE-RED.

Dr.C.SIVAPRAKASAM
Associate Professor,
Depaertment of MCA,
Muthayammal Engineering College,
Kakaveri, Rasipuram, Namakkal, Tamil Nadu, India.

Mrs.C.RADHA
Associate Professor,
Depaertment of MCA,
Muthayammal Engineering College,
Kakaveri, Rasipuram, Namakkal, Tamil Nadu, India.

ABSTRACT

Smart agriculture is one of the Internet of Things' most important uses. Water, fertilizer, and crop yield waste are all reduced via smart agriculture. The manual detection of specifications like temperature, moisture, and humidity in the existing agricultural system drives up labor costs, and continuous monitoring is not possible. The irrigation procedure is carried out automatically in this study utilizing various sensors, which reduces manual work. It is suggested to utilize a sensor-based monitoring system for crop fields. It would entail gathering information on the soil moisture, humidity, and temperature. Automation of irrigation is possible by keeping an eye on all these variables. Unquestionably, smart farming is a key facilitator in providing more food with less resources for a growing global population. While this is essential to feeding the world's expanding population responsibly, smart farming also offers producers and communities throughout the world additional advantages. Farmers may raise yields and improve environmental management by using these strategies. By monitoring the field, IoT-based smart agriculture enhances the overall farming system. The Internet of Things in agriculture helps farmers save time and lessen the usage of resources like water thanks to sensors and connections. electricity, internet-connected temperature monitoring.

1. INTRODUCTION

The increase in world population requires improved production to provide food in all areas, especially in agriculture. However, there are instances when demand and supply are not balanced. Improving agricultural output still faces significant challenges in managing and retaining personnel and capital. A superior choice for enhancing food production, resource management, and labour is smart farming. The increase in world population requires improved production to provide food in all areas, especially in agriculture. However, at certain times, supply and Demand won't equal supply. Improving agricultural output still faces significant challenges in managing and retaining personnel and capital. A superior strategy for boosting food production, resource management, and labour is smart farming. The IoT based farming system helps the farmer to monitor various parameters in his field such as humidity, temperature, and soil moisture using certain sensors [1]. Farmers can use certain sensors to monitor various conditions in the field, such as soil moisture, temperature, and humidity, using an IoT based farming system [2]. Even when far from their fields, farmers can use the web or mobile app to track all sensor information. One of the essential tasks of farmers is to water their crops. By monitoring sensor parameters and managing the pump motor from the mobile app, they can decide whether to water crops or delay watering.

2. LITERATURE SURVEY

An introduction relating to the Internet of Things (IoT), its use in agriculture to improve productivity and quality by reducing costs is provided. The sensors used in the architecture are briefly discussed and the data transfer process from the agricultural domain to the central system is explained [3-4]. The advantages of the recommendation system are included. In addition, open research issues, challenges, and the future of IoT in agriculture are also highlighted. G.Suciu et.al[5-6] proposed the concept is basically developed based on an idea, in which many things or objects - such as Arduino, sensors, GSM models, LCD screens, etc., are connected to the Internet. Each object has a different address and can interact with other elements. Things or objects that cooperate with each other to achieve a common goal. Build a smart agricultural monitoring system that can collect important agricultural data and send it to an IoT platform called Thing speak in real time, where the data can be recorded and analyzed [7]. Data is recorded on Thing speak in a graphical format, a botanist or farmer with reasonable knowledge can analyze the data (from anywhere in the world) to make changes sensitive to the resources provided (for crops) to achieve productivity High caliber.

P. Padalalu et.al [12] explains an emerging technological idea known as "Smart Agriculture" refers to the collection of data from numerous small to large-scale agricultural fields as well as the surrounding environment utilizing sophisticated electronic sensors. Local experts and farmers examine the data collected to make short- and long-term predictions regarding weather patterns, soil fertility, the quality of the present crop, and how much water will be required for the crop. subsequent weeks, months, etc.

3. PROBLEM STATEMENT

An introduction to the Internet of Things (IoT) is given, along with information on how technology might be used in agriculture to increase production and quality while cutting costs [8]. A brief discussion of the sensors utilized in the design is followed by an explanation of how data is sent from the agricultural area to the central system. The advantages of the recommendation system are included [9]. In addition, open research issues, challenges, and the future of IoT in agriculture are also highlighted. This concept is basically developed based on an idea, where there are many things or objects - like Arduino, sensor, GSM model, LCD screen, etc., connected to the internet. Each object has a different address and can interact with other elements [10].

4. PRODUCT DESIGN

Schematics of Data Flow

A data flow diagram (DFD) is a common visual depiction of how information moves through a system. The right number of system needs can be graphically represented by a

clear and precise DFD. It demonstrates how information enters and exits a network, what alters information, and where information It demonstrates how information enters and exits the system, what modifies the data, and where it is kept [11].

Functional Requirement

Each item that can be controlled over the internet is considered an IoT device. IoT devices are becoming readily accessible to consumers because to wearable IoT products (Internet of Wearable Things), like smart watches, and home management solutions, like Google Home. use. More than 30 billion devices are expected to be connected to the Internet of Things by the year 2020. A few requirements are necessary for the IOT based intelligent agricultural system to function properly and decrease output loss [13]. Some of these demands are included in the list below. Farmland first had monitoring devices placed. The user may customize the range for each sensor using the web interface. Since, for instance, some crops need a certain temperature, each sensor is then set to a current value or range [14].

Solution and Technical Architecture

Different soil parameters, temperature, soil moisture, and then moisture are detected by different sensors and the obtained value is stored in the IBM cloud. Arduino UNO is used as a processing unit that processes data obtained from sensors and weather data from the Weather API [15-16]. NODE-RED is used as a programming tool to write hardware, software, and APIs. The MQTT protocol is followed for communication. All collected data is made available to users through a mobile application developed with the help of MIT app inventors [17]. The user can decide, through an app, whether to water the plants based on the sensor's values. Using the app, they can operate the motor switch remotely.The above processes are briefly explained as a block diagram in the Fig 1.Technical Architecture [18]. IOT Device is connected with the IBM Watson Platform which contains the unique username and passkey which are used to access the node points in the node red platform as in (Connections made in the preferred website of Node red) and send information from the web UI to the app controlled by the user as represented in Fig.1 Technical Architecture.

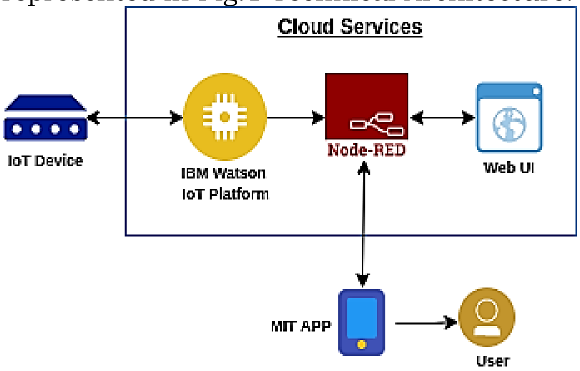


Fig. 1. Technical Architecture Flow

5. EXISTING SYSTEM

Several enabling technologies, including as wireless sensor networks, cloud computing, big data, embedded systems, protocols, and architectural frameworks, provide as a solid basis for the present Internet of Things methodologies. Internet, search engines, online services, communication-enabling protocols, and security mechanisms. Wireless Sensor Network (WSN) It is composed of a number of interconnected sensors and nodes that may monitor different types of data.

6. PROPOSED SYSTEM

To solve farmers' problems such as lack of modernization and mechanization. Invest in agricultural productivity and improve production, help them. respond to climate change and soil erosion. An app and a device are introduced to know various data remotely on their land where they can schedule certain events for a month or a day. It also provides recommendations to users based on the crops they have grown. Farmers can monitor and control their land, and make suggestions for the next crop and improve yields. It's a subscription model in which users pay for their Internet services for support. Reach customers through referrals, resellers and third-party apps. Our products are extensible with our devices (additional modules) as well as with third party devices. Ability to provide different functions in an application like generating reports etc.

Building- Iot based intelligent farming

With the help of autonomous watering systems, sensors (for light, humidity, temperature, soil moisture, etc.), farmers may reduce waste and boost productivity. Farmers may also remotely check field conditions thanks to these sensors. When compared to the traditional method, advanced agriculture based on the Internet of Things is highly effective. It may, however, also be a new instrument for advancing other emerging or well-liked agricultural practices, such as organic farming, family farming (complicated or confined spaces, livestock and/or specific crops, maintaining certain kinds or attributes, etc.), and promoting extremely transparent agriculture.

Precision livestock farming

Cattle farming can also be optimized with IoT technologies. Internet of Things devices allows each animal to be tracked and tracked individually, in terms of health and location. To optimize beef production, farmers can tailor the nutrition of individual animals, as well as monitor animal welfare and identify potential outbreaks. This would be helpful by allowing sick animals to be separated from the herd before the problem spreads, to treat the animal before its condition worsens. This helps farmers reduce costs for veterinarians and routine checkups. It may, however, also be a new instrument for advancing other emerging or well- liked agricultural practices, such as organic farming, family farming (complicated or confined spaces, livestock and/or specific crops, maintaining certain kinds or attributes, etc.), and promoting extremely transparent agriculture.

Functions

Smart Agriculture solutions provide an integrated IOT platform in agriculture that allows farmers to use sensors, smart gateways, and monitoring systems to collect information and control various parameters on the site. their camps and analyze the data in a timely manner. Smart farming includes the application of sensors and automated irrigation practices that can
help monitor soil, soil temperature, and humidity. To provide an efficient decision support system using a wireless sensor network that manages various farm operations and provides farmers with useful information related to the agricultural sector. It is based on four areas, namely monitoring, control, prediction, and logic.

Management and processing

On a local agricultural scale as well as in large-scale soil-atmospheric interaction models, soil moisture is the most significant part of the atmospheric water cycle. Crops and vegetation have traditionally been more reliant on the presence of moisture at the roots than on the frequency of precipitation. Information regarding the local soil moisture is necessary for the budgeting of water for irrigation planning as well as the actual planning for irrigation measures. Understanding the soil moisture content aids in predicting the likelihood of flash floods or the emergence of fog. While soil water potential is a statement of the energy state of soil water, soil water content may be regarded of as an indication of the mass or volume of soil water. It is unusual for content and potential to be correlated, and this connection is influenced by regional soil properties including soil density and soil texture. Simple methods for measuring soil water All items that can be controlled via the Internet are considered IoT devices. To fulfil fluctuating demand, Internet of Things applications in agriculture are shifting toward conventional farms. loss of productivity. Some prerequisites are essential for the IOT-based intelligent agricultural system. Some of these needs are listed below. Farmland first had sensors put for monitoring. The user may configure the range for each sensor via the web interface. For example, some crops need a certain temperature, thus each sensor is then set to a current number or range. The lamp's resistance value is set to 255. If the range decreases, the light will turn on automatically.

Similarly, the two blades of the humidity sensor are embedded in the ground to check the moisture level. If the field is wet, no water is supplied, if the field is dry, the water pump connects to the sensor to supply water to the field. In addition, an additional sensor is retained to measure the water level in the area of interest, if the area is sufficiently watered, the engine can be turned off using a web application via mobile phone or laptop. To save crops from Rot in the warehouse, a temperature sensor is used, which is set to a specific range, if that range is exceeded, the fan will automatically turn on, providing a cooling effect. Likewise, if smoke is detected, the smoke sensor will alert the user and the fan will automatically turn on. All these sensors are monitored using a web application developed in Python. User will get a URL where he can login with username and password and can view temperature and control sensors. This feature allows users to get better performance and results. IoT platform agnostic types or applications

- Connectivity: cellular, LoRa
- Location: GPS, satellite
- Robots: automated tractors, processing plants

Data as a Service

The data will be fed from the database to the dashboard of the automation system customer that is actually grown in the crop field. Information about all devices, systems, and processes is considered data, and data is provided to customers to run automation processes efficiently.

Management as a Service

Basic and advanced control system for connected devices in the field for the process of automation system with control buttons and sliders for automation. Customers can control processes and equipment at their discretion. In order to collect data for analysis, In remote sensing, sensors are used, such as weather stations installed on farms. They watch for changes in plants' shape, size, light, moisture, and temperature. To determine weather patterns in fields and nurture the proper crops, sensors collect data on humidity, temperature, precipitation, humidity, and dew detection. The amount of water required for irrigation and the most profitable type of farming is controlled in part by the soil's acidity or drainage capability. Analysis of the soil quality also aids in determining the farm's dry area and nutritional value.

7. CODE FOR SMART FARMING MONITORING

```
import wiotp.sdk.device
import time
import os
import datetime
import random
myConfig = {"iden ty":{
"orgId":"04gt4e ",
"typeId": "NodeMCU",
deviceId": "12345"
},
"auth":{
"token": "123456789" }
}
client = wiotp.sdk.device.DeviceClient(config=myConfig,
logHandlers=None)
client.connect ()
def myCommandCallback (cmd) :print ("Message received from IBM IoT Pla orm:
%s" %
cmd.data['command'])
m=cmd.data['command']
if(m=="motoron"):
print ("Motor is switched on")
elif(m=="motoroff"):
```

Fig.2. Code for Finding Environmental Parameters

Fig. 2 and Fig. 3 represents the Code for Finding Environmental Parameters such as Temperature, Humidity and Moisture of the registered fields at different timings in different cities.

```
logHandlers=None)
client.connect ()
def myCommandCallback (cmd) :print ("Message received from IBM IoT Pla orm:
%s" %
cmd.data['command'])
m=cmd.data['command']
if(m=="motoron"):
print ("Motor is switched on")
elif(m=="motoroff"):
("Motor is switched OFF")
print(" ")
while True:
soil=random.randint(10,100)
temp=random.randint(-20, 125)
hum=random.randint(0, 100)
my Data={'soil_moisture': soil, 'temperature':temp,
'humidity ':hum}
Client . publish Event (even tId="status", msg
Format="j son",data=my Data, q o s=0 , on Publish=None)
print ("Published data Successfully: %s", my Data)
time.sleep(2)
client. Command Call back = my Command Call back
client . disconnect ()
```

Fig.3. Code for finding Environmental Parameters.

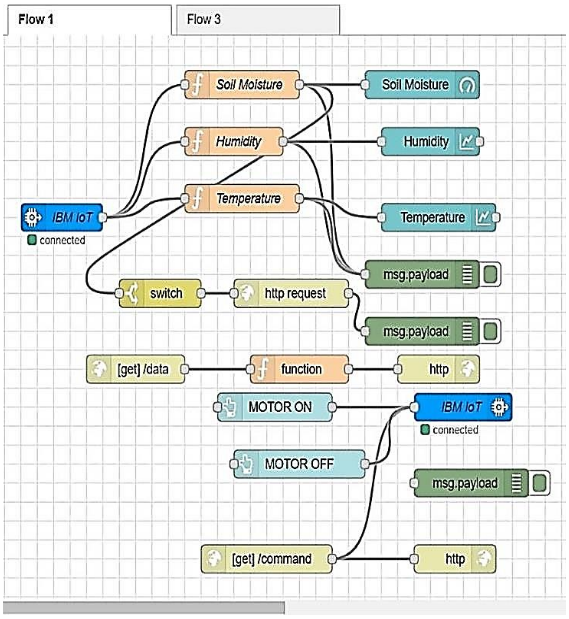


Fig.4. Connections made in preferred website of Node Red

Fig.4 represents the connections made in the Node red Platform which is an open-source visual programming tool called Node-RED is used to connect physical components, APIs, and web services. Users may design flows using the web-based flow editor by dragging and dropping nodes onto a canvas and connecting them.Node.js, a well-liked JavaScript runtime environment for developing server-side applications, forms the foundation of Node-RED. Node-RED is intended to be simple to use and quick to deploy, which makes it perfect for Internet of Things applications. Nodes from Node-RED may be used to communicate with a variety of hardware components, including sensors and actuators, and software applications, like Twitter and Slack.

8. ADVANTAGES

A remote control system can help operate the valves in the irrigation system on a schedule. Irrigation of remote farm properties can be particularly cumbersome and laborious. It is confusing when the valves are actuated and whether the ideal amount of water has been delivered. For situations where a quick response is required, manual valve activation is not always possible. Thus, remote monitoring and control of irrigation systems, generators or wind turbines, or any other motorized equipment becomes the next logical step.There are various solutions for monitoring engine statistics and starting or stopping the engine. When the customer chooses to start or stop the engine, the program transmits a signal to the unit within seconds via the mobile phone system. Submersible gravity sensors or ultrasonic sensors can monitor the levels of tanks, lakes, wells, and different types of liquid storage such as fuel and compost. The volume of the product figures depends on the shape of the reservoir or lake after a certain time. It transmits alarms based on different conditions.

Disadvantages does not meet this requirement. Also, the internet connection is slower. Smart farming-based equipment requires farmers to understand and learn how to use technology. This is a big challenge for the large-scale adoption of smart agriculture in all countries.

9. CONCLUSION

Farmers can greatly benefit from an IoT-based smart farming system. Agriculture is affected by a lack of irrigation water. Climatic factors such as humidity, temperature, and humidity can be adjusted for local environmental variables. The technology also detects animal encroachment, a major cause of crop failure. This technology facilitates irrigation planning based on field data and climate source records. This helps the farmer decide whether to irrigate or not. A permanent Internet connection is required for continuous monitoring of sensor data. This can also be overcome by using another GSM device for the mobile application. Thanks to GSM, text messages can be sent to farmers' phones.A smart farm monitoring system can be used as a farm's destiny item. It will be a relief for farmers as it will ease the burden of manual efforts. A soil moisture filter has been built and the mission offers the possibility of examining the dominant structures, from the perspective of their characteristics and disadvantages. The specified utility can be used to turn on/off sprinklers based on soil moisture, thereby automating irrigation engineering, one of the most time- consuming activities in agriculture. Farming is one of the most demanding hobbies. The device uses statistics from the soil moisture sensor to irrigate the soil.

10. FUTURE SCOPE

In the current project, we have implemented a project that can protect and maintain culture. In this project, the farmer monitors and controls the field remotely. In the future, we may add or update some more things to this project. We can create several more models of the same project so that the farmer has the overall information. We can update this project with the solar mechanism. So, the power supply of the electric poles can be replaced by solar panels. It reduces the cost of power lines. It will be a one-time investment. We can add solar fence technology to this project. We can use GSM technology for this project so farmers can receive information directly at home by SMS. This helps farmers get information if there is a problem with the internet. We can add camera functionality so farmers can monitor their fields in real time. This helps prevent theft. An economically friendly mechanism will be invented further and its our hope.

11. REFERENCES

1. S. Prathibha, A. Hongal and M. Jyothi, *IOT Based Monitoring System in Smart Agriculture*, in International Conference on Recent Advances in Electronics and Communication Technology (ICRAECT), (2017)
2. C. Yoon and M. Huh and S. Kang, J. Park and C. Lee, *Implement smart farm with IoT technology*, in 20th International Conference on Advanced Communication Technology (ICACT), (2018)
3. B.N.B. Ekanayake, M.N. Halgamuge, A. Syed, *Review:Security and Privacy Issues of Fog Computing for the Internet of Things (IoT)*, Lecture Notes on Data Engineering and Communications Technologies Cognitive Computing for Big Data Systems Over IoT, Frameworks, Tools and Applications, Springer, **14** (2018)
4. S. Takekar, S. Takekar, *Plant and taste to reap with Internet of Things implementation of IoT in agriculture to make it a parallel* industry, in International Conference on I- SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), (2017)
5. D. Sreekantha, Kavya A.M., *Agricultural crop monitoring using IOT - a study*, in 11th International Conference on Intelligent Systems and Control (ISCO), **4** (2017)
6. M. Marcu, G. Suciu, C. M. Balaceanu, A. Banaru, *IoT based System for Smart Agriculture*, 2019 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), 1-4, (2019)
7. S. Rajendrakumar, Rajashekarappa, V.Parvati, B.Parameshachari, K. Soyjaudah, R. Banu, *An intelligent report generator for efficient farming*, in International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEEECCOT), (2017)
8. P. Padalalu, S. Mahajan, K. Dabir S. Mitkar, D. Javale, *Smart water dripping system for agriculture/farming*, in 2nd International Conference for Convergence in Technology (I2CT), (2017)
9. G. Raghudathesh, D. Deepak, G. Prasad, A. Arun, R.Balekai, V. Yatnalli, S. Lata, B. Kumar, *Iot based intelligent poultry management system using Linux embedded system*, in International Conference on Advances in Computing, Communications and Informatics (ICACCI), (2017)
10. Walter, R. Finger, R. Huber, N. Buchmanna, Proc Natl Acad Science, **114**, 6148–6150, (2017).

ID: 61

STRESS MANAGEMENT: CONCEPT, APPROACHES, AND ANALYSIS

Dr.P.GANAPATHI
Associate Professor & Head,
Department of Management Studies,
Muthayammal Engineering College (Autonomous),
Rasipuram, Namakkal, Tamil Nadu, India.

S.K.SENDHILSABRIVISWANATH
Assistant Professor,
Department of Management Studies,
Muthayammal Engineering College (Autonomous),
Rasipuram, Namakkal, Tamil Nadu, India.

ABSTRACT
Purpose: Enhance well-being; acquire adaptive decision-making skills; Relationship breakdown, mental health enhancement, and unemployment problem-solving are all focus areas.
Design/Methodology/Approach: Secondary data for this study came from a wide range of places, such as case studies, books, periodicals, journals, articles, and online searches.
Findings/Results: Stress has many facets, both in terms of its causes and its effects. It's more of a personal journey than a general truth to determine what helps you cope with pressure.

The client will be taught techniques for dealing with everyday and unexpected tension sources. Cognitive and behavioral strategies are helpful for stress management.

Mindfulness-based stress reduction therapies have received much research and attention recently. Due to its association with numerous diseases, stress management is paramount.

Outcome:Theoutcomes include the concept of stress management, various stages of stress and how to overcome the stress, the causes and effects of individual stress, the management of stress effectively through the use of cognitive behavioral techniques, and how to use mindfulness-based interventions for stress management.

Originality/Values: This research sheds light on the many sources of stress in the lives of humans, including but not limited to increased workload, increased risk of physical and mental illness, and strategies for coping with stress at different levels. How one chooses to spend one's life determines the particular sources of stress that they will face.

Type of Paper: Literature review

KEYWORDS: Stress Management, individual and social life, stressors, personality, and cognitive behavioral.

1. INTRODUCTION:

There is much pressure in today's society. Having lofty goals and feeling pressure to reach them are well-known sources of tension. A person's incapacity can also cause stress to make decisions or adjust to new circumstances. How stressed you feel is influenced by various individual and environmental factors. Those prone to stress have more difficulty keeping relationships strong and worry more about their health. Although stress is unavoidable, it can be controlled to some degree. It might help keep connections with people on an upbeat track. It's essential to talk about practical ways of coping with stress in humans.

One definition of stress by Lazarus and Folkman (1984, p. 19) [1] suggests that stress and coping are dynamic phenomena: "particular interaction between the person and the environment, that the individual perceives as tiring or exceeding their resources and damaging their well-being." A developing brain is more vulnerable to stress while moving along a continuum between two poles, such as "does" and "does not." Long-term stress, which can lead to mental and physical illness, rises with the volume and length of perplexity. Psychological stress, as defined by Coffer and Appley (1967) [2], is the perception that one's physical or social well-being is threatened and that one must take steps to mitigate this threat. According to Vingoi (1981) [3], prolonged exposure to aggravating activities or conflictual circumstances is a significant contributor to mental health issues. Stress is a standard component of life because of how common stressful events are among humans.

Learning positive coping strategies can reduce your chances of developing psychopathology. One's environment greatly influences the ability to perceive and react to stressful events. Bower (1973) [4] argues that people experience emotional pain when they are placed in situations that are inappropriate for them.

Stressors:

The pressures of daily life, whether at the job or home, can be overwhelming at times. The decision to live alone places stress on everyone. There are numerous origins for both short-term and long-term stresses. Researchers Holmes and Rahe (1967) [5] found that disruptions in a married couple's routine—such as the death of one partner, a divorce, or a trip abroad—can be just as stressful as other significant life changes. Cohen and colleagues examined both short-term and long-term stress in their 1998 study [6]. Compared to marital discontent or unemployment, they found that even a severe reprimand at work or a disagreement with the spouse resulted in less chronic stress. Selye (1950 [7], 1956 [8], and 1958 [9]) identified three stages of stress that make up the general adaption syndrome: alert, resistance, and weariness. A person under stress may go through each of these stages. In certain situations, overcoming each stress level may be doable or impossible.

2. OBJECTIVES OF THE PAPER:

(1) To review the concept of stress management.
(2) To study the various stages of stress and how to overcome the stress.
(3) To know about the causes and effects of individual stress.
(4) To manage stress effectively through the use of cognitive behavioral techniques.
(5) To focus on mindfulness-based interventions for stress management.

3. REVIEW OF LITERATURE / RELATED WORKS:

Weiss M. (1983)[10] -The author looked at what causes job stress, which is associated with being unhappy at work, experiencing stress and anxiety on the job, and being less productive and successful overall. He tried to lessen his stress levels to avoid any potential health problems. His research showed that having a strong social network can help mitigate the adverse effects of stress.

Singh A.P. & Singh S.(2009) [11] – His research centers on employee happiness on the job. He argues that Working conditions and stress levelssignificantly contribute to whether or not an employee feels fulfilled. He divided stress into two primary categories, eustress, and distress, and pinpointed its genesis in three distinct areas. Furthermore, he emphasized the significance of positive stress and happy occurrences for enhanced productivity and worker happiness.

Gladies J.J. & Kennedy V (2011)[12]- The author found a robust connection between an organization's culture and female IT workers' stress levels on the job in India. He thinks that IT firms should prioritize training their employees to deal with stress if they want to lessen its effects on productivity and morale in the workplace.

Charu M (2013)[13] -According to his findings, IT professionals' quality of life at work decreases as their stress levels rise. He listed a few elements that directly impactthe quality of life at work, including fair pay structure, steady role demands, supervisory support, a friendly work environment, capacity fit of the job, role autonomy, and stress. IT workers' primary source of anxiety is the industry's penchant for constant technological upheaval.

Khalid A (2012)[14] -Stress and work performance go hand in hand in any workplace. Employees who receive strong support from their superiors are more likely to succeed. Therefore, even in dire circumstances, an employee's performance can be enhanced by having a leader who is encouraging.

4. RESEARCH METHODOLOGY:

The majority of the data in this qualitative investigation came from secondary resources. Case studies, books, periodicals, journals, papers, and online searches or publications were consulted to compile the data used in this study. Google Scholar, Research Gate, and SSRN were used to locate the necessary data for the investigation.

Sources of Stress:

The two most common types of stress are high-frequency and low-frequency. The increased workload is an example of low-frequency stress that can have a considerably more significant impact than high-frequency pressures like those encountered in daily living (Delongis et al., 1988 [15]; Kenner et al., 1981[16]; Lazarus et al., 1985 [17]). Because of this, we may argue that stress is an inevitable part of being human and can manifest itself in any facet of an individual's existence, including but not limited to personal relationships, work, and school. Remember that the things that stress you out are different for everyone. Depending on one's lifestyle choices, stress can have a wide range of causes.

5. PERSONALITY AND STRESS:

The effects of stress and one's ability to handle them vary significantly from person to person. Based on the research of Jung and Sheldon, Mohanty (1991) [18]concluded that ectomorphs are less vulnerable to stress than endomorphs and geomorphs. Carl Jung found that introverts and the geomorphic personality type were very similar. Jung categorizes extroverts and introverts similarly to Sheldon's endomorphic and ectomorphic personality types.

Scheier and Carver's (1988) [19] study found that pessimists generally have better stress tolerance. According to research by Scheier and Carver (1992) [20], pessimists use multiple coping mechanisms. Optimists, on the other hand, are less competent in this setting. Segerstrom and coworkers found 1998 [21] that law students with a positive outlook performed better academically than their more pessimistic peers. Because of these dissimilar responses to stress, it is evident that people of varying personalities utilize various coping strategies. The fact that introverts worry more and hold themselves to higher standards makes them more susceptible to stress is not debatable.

Women are more easily overwhelmed by stressful situations than men. Potter and Stone (1995) [22] found that males and females approach stress differently. Gross's (1992) [23] medical research demonstrated that while male and female doctors experienced pressure from the time commitments required by work, female doctors had the added burden of handling family duties. Female family physicians are more likely than their male counterparts to show signs of mental illness, per research by Rout (1999) [24]. Male doctors, on the other hand, are more prone to show signs of anxiety and depression. His studies found that men generally have lower "floating anxiety" levels than women. In contrast to women, men are more likely to experience somatic anxiety. The depression rate is considerably higher in men than in women.

6. STAGES OF STRESS:

The range of possible degrees of stress is quite extensive. There are four distinct stages of stress, as described by Seaward (1999, p.7) [25]. The senses are the first to send information to the brain. The second level is some mental decoding. The brain makes an instantaneous assessment of risk. The stress reaction would be expected if the brain didn't sense a threat. Failure to do so triggers stage three, during which our bodies stay on high alert until the threat has gone. The last step is to bring the body and mind back into homeostasis or balance. First, the body experiences what Hans Selye (1976) [26] calls "alarm," the sympathetic nervous system is activated, and hormones are released to prepare the organism to deal with the threat. If the stress is prolonged, the body enters a second stage called resistance, characterized by a lower arousal level than the first. Physical and mental weariness will set in if the tension continues for a long enough time.

The strategies suggested during the resistance and fatigue phases will lead to the normalization phase.

Stages of Stress:

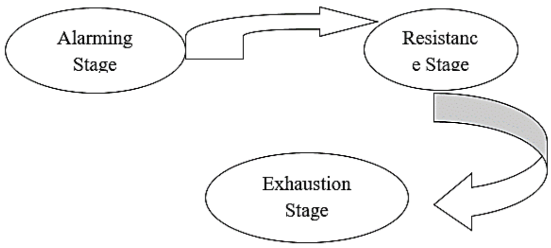


Fig. 1: Various Stages of Stress

(1) Alarm Stage: The initial exposure to the stressor occurs during the Alarm phase. The brain's nervous system immediately sends up a red flag. When we're under pressure, our bodies release more adrenaline hormones. So, breathing speeds get up. As stress levels grow, the body responds by increasing heart rate, blood pressure, oxygen delivery to the brain and muscles, slowed digestion, improved vision, increased perspiration, and a dry mouth. The "fight or flight" response is triggered by the quick release of adrenaline in response to stress.

(2) ResistanceStage: The resistance stage occurs when an individual's body cannot adapt to a stressor during the alarm stage.

Increased irritation, exhaustion, and a diminished capacity to deal with stressful events are all symptoms of stress because the body is still trying to deal with the stressor. The process of adapting to a stressor by fighting it off until it can be removed is known as resistance. If the stressful situation continues, the body will maintain the adaptations it established previously. When no longer under stress, the body returns to its pre-stress state.

(3) Exhaustion Stage: The "exhaustion stage" is when a person has been subjected to a stressful situation for a long time, and their bodies have finally given up (weeks, months, or years). The third and final stress stage of exhaustion develops when stress continues past the second stage, and no corrective action is performed. Everyone involved has run out of steam and is spent.

7. STRESS MANAGEMENT:

Everyone is unique in their ability to recognize and respond to stressful situations, and everyone has unique resources. Although people utilize a vast range of coping strategies, a select handful could be helpful to everyone. The existence of coping strategies raises an interesting question. Specifically, Lazarus and Folkman define coping as "the process where the individual constantly alters his/her cognitive and behavioral attempts to manage certain external and internal demands that are assessed as stressful or beyond the person's resources" (1984) [27].

Stress originates in the individual's thought and activity processes, making the cognitive and behavioral functions vital in combating it. When dealing with stress and the accompanying emotions, "coping mechanisms essentially rely on the cognitive and behavioral of the person," as Halahan et al. (2004) [28] put it.

Coping Strategies:

There are other approaches to stress management beyond the Western-based stress reduction (CBSR) techniques, where cognitive intervention plays a central role. Dysfunctional evaluations keep stress levels high and even increase them. Recognizing and changing such maladaptive evaluations or other cognitive processing problems is the main component of cognitive intervention. Mindfulness-based stress reduction (MBSR) is a modern field that draws inspiration from ancient Eastern practices to manage stress. In a small pilot study with 50 participants, Smith, Shelley, Dalen, Wiggins, Tooley, and Bernard (2008) [29] compared CBSR with MBSR. The Mindfulness-Based Stress Reduction (MBSR) programme aims to help its participants become more self-aware and attentive through deep breathing, body scans, meditation, gentle yoga, and open discussion. Stress management strategies, both cognitive and behavioral, were taught at the CBSR. The solution mainly consisted of cognitive retraining and behavioral methods for reducing stress. Across eight outcomes, the MBSR group's impact magnitude was more significant than double that of the CBSR group (i.e., perceived stress, depression, psychological well-being, neuroticism, binge eating, energy, pain, and mindfulness). Stress from everyday life has been related to recurring bouts of depression. Britton, Shahar, Szepsenwol, and Jacobs conducted research in 2012 [30] on 52 individuals whose depression had only partially subsided. Those who performed MBCT had less of an emotional ability to withstand the effects of social stress. Happiness and stress are reciprocal (Carlson et al., 2003) [31]. Nykl&Kuijpers (2008) [32] examined this issue by surveying 40 distressed females and 20 distressed males. Emotional discomfort was found to decrease when participants participated in mindfulness-based therapies.

(1) Social support: A study including 232 senior patients found that the chance of dying within six months after open heart surgery was reduced when the patients participated in a social and community group program (Oxman et al., 1995) [33]. Multiple research (e.g., Berkman and Syme, 1979 [34]; Cohen and Wills, 1985 [35], 1988 [36]; Dunkel/Schetter and Skokan, 1990 [37]; Hobfollet al., 1990 [38]; House et al., 1988 [39]; Segerstrom, 2007 [40]; Strine et al., 2008 [41]) have found that people's mental and physical health improve after experiencing a supportive relationship.

(2) Relaxation Exercise: Relaxation is good for the practitioner's mental and physical health. Progressive muscular relaxation is one method of relaxation training among many—methods like meditation and autogenic training. Alternating periods of muscle tension and relaxation, as in progressive muscle relaxation, helps relax specific groups of muscles. However, various meditation techniques aim to induce a calm, reflective frame of mind. Meditation is a technique for calming the mind and body that has gained popularity far beyond its Eastern beginnings. References: Shapiro (1985) [42]; Dimatteo & Martin (2012) [43]. Autogenic training is a behavioral technique for controlling anxiety and learning to unwind at will. Autogenic training uses guided visualization and sounds to help people relax their bodies and mind. This technique can affect various physiological reactions, including breathing, blood pressure, heart rate, and skin temperature (e.g., Ernst and Kanji, 2000 [44]; Crowther, 1983 [45]).

(3) Time Management: Time management courses help participants plan their days more efficiently and set priorities. Increasing productivity is a must for effective time management. To assist people in modifying their actions and experiencing less stress, Lakein (1973) [46] suggested time management techniques like compiling a list and prioritizing tasks. When Richards (1987) [47] started using time management strategies, he felt less pressure and produced more. Robinson and Godfrey (1997) [48] confirmed the close relationship between time management and stress relief.

8. APPROACHES TO MANAGE STRESS:

- Treat your body with respect.
- Finding Peace of Mind.
- Establishing a system of social assistance.
- Physical and mental withdrawal.

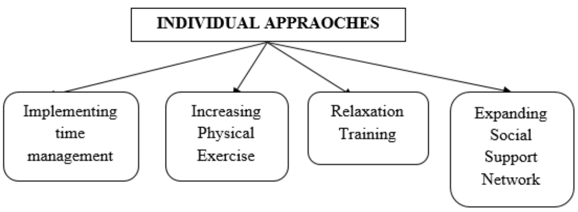


Fig. 2: Block Diagram of Managing the Stress by Individual Approaches

Reasons behind high levels of stress:

- When workers' workloads increase, their stress levels rise.
- The absence of growth and development opportunities
- An unacknowledged society
- Inadequate resources to complete regular tasks
- A lack of open dialogue and a toxic workplace atmosphere led to this.

EFFECTS OF JOB STRESS ON PHYSICAL HEALTH:

- Chest pain and rapid heartbeat
- improved breathing
- Muscle soreness, tightness, and stiffness
- poor health
- Headaches
- Abdominal pain and digestive issues that won't go away
- Insomnia
- Persistent viral infections and colds
- Fear and shaking
- Lipped dryness
- Clenched teeth
- Obesity

Effects of Job Stress on Mental Health

- Possessing a propensity for outbursts of rage and frustration
- Weary of worrying about everything and overanalyzing everything
- Neglecting or avoiding one's responsibilities
- Having a hard time relaxing and calming down
- Confidence level dropping
- dislike of human company
- Lack of focus and memory
- Carelessness and poor decision-making
- Having a poor attitude
- Increase in Anxiety and Depression
- The prevalence of drug, alcohol, and cigarette use rises

Effects of Stress on Employee Performance

- Business operations become less efficient
- Worker output as a whole suffers.
- Not very precise
- Poor output and a high rate of staff turnover.
- Absenteeism

9. PRIMARY TECHNIQUES FOR EMPLOYEE STRESS MANAGEMENT:

(1) Improving communication: This can be achieved by ensuring everyone in the company knows what they're supposed to do. Meeting consistently and having defined roles can help with this.

(2) Increasing productivity: This is accomplished by clearly communicating what is expected of workers and why and giving themthe tools they'll need to succeed.

ISBN: 978-81-965283-7-9

(3) Reducing costs: One way to accomplish this is to lessen or eliminate sources of unneeded stress in the workplace. Such factors could include, for example, impossible deadlines, high workloads, or unfavorable working environment.

10. DISCUSSION AND SUGGESTIONS:

(1) Take a break from news coverage on TV, radio, or social media.
(2) If you're feeling overwhelmed by stress, remember to take care of yourself by eating well, exercising regularly, getting plenty of sleep and water, and giving yourself a break.
(3) Maintain your physical health with frequent deep breathing, meditation, and exercise.
(4) Don't get into the habit of drinking, smoking, or using other drugs excessively.
(5) If your doctor has prescribed regular preventative care, such as immunizations or cancer screening, keep up with them.
(6) Protect yourself from contracting COVID-19 by being vaccinated.
(7) Share your situation with people you trust, including your parents, friends, a counselor, or a doctor, and let them know how you feel.
(8) Practicing self-care at work is especially important because it directly impacts productivity and health.
(9) Get more shut-eye; impaired cognition is a direct result of sleep deprivation.
(10) Boost workers' sense of psychological security to mitigate anxiety about potential dangers on the job.
(11) Foster an environment where employees may work in peace.
(12) The lines between time spent on the job and with friends and family tend to blur.
(13) Develop accommodating procedures for employees who must juggle work with caring for their families.
(14) Employees are most productive when engaged in meaningful work, receiving adequate support, and given opportunities to grow.

11. CONCLUSION:

Stress has many facets in terms of its causes and effects. It's more of a personal journey than a general truth to determine what helps you cope with pressure. The client will be taught techniques for dealing with everyday and unexpected tension sources. Cognitive and behavioral strategies are helpful for stress management. Mindfulness-based stress reduction therapies have received much research and attention recently. Due to its association with numerous diseases, stress management is paramount.

REFERENCES:

1. Baron, R. A. (2004). Health,stress and coping: Psychology, (5theds), Low PriceEdition, published by Pearson Education (Singapore), Pte, Ltd, Indian branch, 482, F.I.E, Delhi-92. India, pp-488-530. Google Scholar↗
2. Bower, K.S. (1973). The situation in Psychology: An Analysis and Critics. Psychology Review, 80, 307-335. Google Scholar↗
3. Brown, J.D. & McGill, K.L. (1989). The cost of good fortune: When a positive life events haves negative health consequences. Journal of Personality and Social Psychology, 57, 1103-1110. Google Scholar↗
4. Bruhn, J.G. & Phillips, B. U., (1987). A developmental basis for social support. Journal of Behavioural Medicine, 10, 213-229. Google Scholar↗
5. Coffer, C.NandAppley, M.H. (1967). Motivation: Theory and Research, Willey, NewYork. Google Scholar↗
6. Cohen, S., Frank,E., Doyle,W.J., Skoner,D.P., Robin,B.S. & Gwaltney, J.M. (1998). Types of stress that increase us acceptability to the common cold in healthy adults. Health Psychology, 3, 301-304. Google Scholar↗
7. Delongis, A., Folkman, S. & Lazarus, R .S. (1988). The impact of daily stress on health and mood: Psychological and social resources as mediators. Journal of Personality and Social Psychology, 54, 486-495. Google Scholar↗
8. Gerber, R.(1991).Personalcommunication,November25.Google Scholar↗
9. Gerber, R.(1998). Vibrational Medicine. BearandCo. Santa Fe,N. M.Google Scholar↗
10. Gross, E.B. (1992). Gender difference sinphysicians' stress. Journal of the American Womens Association, 47(4), 107-112. Google Scholar↗
11. Halahan, C.J., Roger, J.D. and Moos, R.H.(2004). Stress, Encyclopedia of Applied Psychology, firstedition, Edited by Charles, D.Spielberger, volume3, pp-485-495. Google Scholar↗
12. Holmes, T.H.& Rahe,R.H. (1967). The social readjustment rating scale.

Journal of Psychosomatic Research, 11, 213-218.Google Scholar

13. Kenner, A.D., Coyne, J.C., Schafer, C., Lazarus, R.S. (1981). Comparison of two modes of stress measurement: daily has sles and up lift svers us major life events. Journal of Behavioural Medicine, 4,1-39.Google Scholar
14. Karasek, R. & Theorell, T., (1990). Healthy work: Job stress, productivity and the reconstruction of working life. NewYork, Basic Books. Google Scholar
15. Lazarus, R.S., Opton, E.M.Nomikos, M.S. & Rankin, N.O. (1985). The principle of short-circuiting of threat: Further evidence. Journal of Personality, 33, 622-635. Google Scholar↗
16. Lazarus, R.S. and Folkman, S.(1984) Stress: Appraisal and Coping, NewYork, Springer. Google Scholar↗
17. Meichenbaum, D.H., (1977). Cognitive Behavioural Modification., NewYork: Plenum. Google Scholar
18. Mohanty, G.B. (1991). Personality: A Text Book of General Psychology. Kalyani publisher, NewDelhi, PP-355-385. Google Scholar
19. Myerson, S. (1992). Problems in UK general practice since the new contract (1990) and general practitioners' strategies for dealing with them. Medical Science Research, 20(1), 461-463. Google Scholar
20. Myss, C. (1996).Anatomy of the spirit. Harmony Book, New York. Google Scholar
21. Porter, L. S. & Stone, A. A. (1995). Arethere gender differences in coping? A reconsideration of previous data and results from a daily study. Journal of Social and Clinical Psychology, 14(1), 184-202. Google Scholar
22. Rodin, J. & Salovey, P. (1989). Health psychology. Annual Review of Psychology, 40(1), 533-580. Google Scholar
23. Scheier, M.F. & Carver, C.S., (1988). Perspective on Personality, Boston: Allyn and Bacon. Google Scholar
24. Scheier, M.F. & Carver, C.S.(1992). Effect of optimism on psychological and empirical update. Cognitive Therapy and Research, 16(1), 201-228.Google Scholar

ISBN: 978-81-965283-7-9

ID: 62

IMPACT OF WORK-LIFE BALANCE AND STRESS MANAGEMENT ON JOB SATISFACTION AMONG THE WORKING WOMEN IN HIGHER EDUCATIONAL INSTITUTIONS IN NAMAKKAL DISTRICT

Dr.P.GANAPATHI
Associate Professor & Head,
Department of Management Studies,
Muthayammal Engineering College (Autonomous),
Rasipuram, Namakkal, Tamil Nadu, India.

Mrs.P.KEERTHI
Assistant Professor,
Department of Management Studies,
Muthayammal Engineering College (Autonomous),
Rasipuram, Namakkal, Tamil Nadu, India.

ABSTRACT

Purpose: Job Satisfaction**,** Stress management, Organizational success, Competitive Workplace, Fair compensation, teamwork, peer communication, ambition, Workload, and Well-being.

Design / Methodology / Approach: Case studies, books, periodicals, journals, articles, and online searches are all examples of secondary sources used in this investigation.

Findings / Results: Effective stress management helps people feel better about themselves and do better in the predictable world. It gave people something to look forward to, which boosted their motivation and confidence.

As a woman has to take care of her family, she should know the techniques to manage the family and the professional work she is undergoing.

She has to maintain good health because she will be the backbone of her family.

She has a sole responsibility to take care of herself by eating a healthy diet, exercising, taking deep sleep, maintaining a friendly social network, listening to music, reading books, having positive thinking ability, etc.,

Outcome:

- To review the concept of work-life balance and stress management.
- To know the steps to achieve a successful Work-Life Balance.
- To have the awareness about how to find a perfect Work-Life Balance.
- To develop the skills that require managing stress.

Originality/Values: Irritability, impatience, family pressure, anger, frustration, fear, anxiety, self-doubt, panic, and melancholy, feelings of inadequacy, insecurity, hopelessness, dissatisfaction, emotional withdrawal, and depression are some of the symptoms of stress that this study focuses on, particularly as they pertain to women. Every lady needs to understand how to triumph over the obstacles above.

Type of Paper: Empirical Research

KEYWORDS: Job satisfaction, Family, Employee satisfaction, Stress, Performance, Productivity, and Rewards.

1. INTRODUCTION:

Women in the workforce can better manage their stress and strike a work-life balance if they are happy and have a secure home life. Everywhere you turn, people are struggling to find a way to combine their work and personal life successfully. If you're not enjoying your job, this is a pressing problem. Finding a means to keep your workload consistent at home and work is essential to strike a good work-life balance. Employee production is crucial to the success of any business, and it can be affected by a wide range of factors. There may be a connection between job satisfaction and the following factors. An empirical study was conducted among the working women of public sector banks to examine the impact of work-life balance and stress management on job satisfaction. The data and conclusions in this study were derived from secondary resources. Employees report less stress when their workplaces have positive features such as social support, work that challenges them mentally, reasonable pay, and protections for their rights as workers. Findings suggest that work-life balance and stress management are not issues that can be solved but rather persistent difficulties that require constant attention.

A "work-life balance" exists when one's professional and personal responsibilities are equally important. An imbalance between work and personal life might have several root causes. Both the hours worked, and the amount of labour required has increased.

The economy of India is on the rise, and most women there are now working outside the home. Therefore, the women who work in the public sector banking industry are the primary focus of this research. The primary goal of this research was to determine if and how stress management and work-life balance affected job satisfaction. One's mental identity and sense of well-being are affected by one's work, which affects one's social worth and status. When we talk about "work,"we refer to formal, salaried employment. A job is a great way to become part of the human family. It connects people, furthers cultural goals, and offers individuals a sense of meaning in life. Humans engage in work because it helps them meet basic requirements and pursue personal goals. Having a job that one enjoys doing is essential for a healthy working relationship between an individual and a company.

Work-life balance and stress management are becoming increasingly discussed topics in the Indian setting, particularly as they relate to women's levels of job satisfaction. A work-life balance is achieved when work and personal life demands are roughly equal. When the requirements of one role overlap with those of another, it could be more transparent and more manageable. When people's mental and emotional resources are stretched to the breaking point, when they feel overwhelmed and powerless in the face of the many demands and obstacles, they describe this as stress. Stressors, or the things that produce stress, are plentiful, numbering in the hundreds. People can be stressed out by anything they perceive as dangerous, difficult to deal with, or excessively pressuring. The reality is that work and personal commitments overlap, interact, and contribute to stress. The effects of stress on productivity and morale are well-documented. However, if women aren't happy in their jobs, it can lead to stress and an unbalanced lifestyle. Employers should consider employee dedication as a differentiator between successful and unsuccessful businesses when formulating stress and work-life policies. Women must spread their attention and resources to live a balanced existence.

The people who work for a company are its most valuable asset—success in business hinges on attracting and keeping talented employees. The human environment looks at the factors unique to people that can boost productivity and happiness on the job. Employees are said to be satisfied with their jobs when they feel a positive emotional pull toward their employment. Worker creativity increases when they enjoy their work. People may develop a stronger sense of loyalty to the company.

Workers will be happier if their work-life balance and stress reduction needs are met. Multiple studies have examined the correlation between work-life balance, stress levels, and job satisfaction. While there have yet to be many studies on the interplay between these three factors, what has been done provides evidence that public sector bank employees' female counterparts are likelier to be honest and hardworking. There will be a severe lack of intellectuals (both in quantity and quality) if bank administrations cannot effectively manage their female workforce when creating policies. Logic dictates that a contented employee is likelier to exhibit high productivity levels[1,5,7,8].*

The three significant habits for better Work-life balance are:

(1) Putting forth maximum effort.

(2) Please focus on the positive and remember what makes us happy.

(3) Work, play, love, and health all need obligations.

Programs that aim to help employees strike a better work-life balance include those that allow for more time off, help with caring for family members, provide paid time off or vacation, teach new skills, encourage healthy habits, and organize group activities.

2. OBJECTIVES OF THE PAPER:

(1) To enrich the knowledge of theWork-life balance and stress management concept at educational institutions.

(2) To know the female worker's condition within the educational environment

(3) To concentrate on happiness among the female workers within the educational institutions.

(4) To set boundaries for women who are working in educational institutions.

(5) To check whether the women employees maintain healthy relationships among

the other employees working in the educational institutions.

3. REVIEW OF LITERATURE / RELATED WORKS:

Greenhaus and Kossek, 2014 - Earlier voices of WLB concern came from the UK's working mothers in the 1960s and 1970s. The US government finally gave this matter full attention in the middle of the 1980s. WLB became a primary concern in human resource management worldwide in the 1990s (Bird, 2006). The growing importance of women in the workforce, technological advancements, societal shifts in attitudes regarding the interaction between work and family, and the variety of family arrangements have all contributed to a rise in scholarly publications addressing WLB. Several theoretical work-family models are included in the WLB literature. Even though there has been more study of WLB, we still know relatively little about balancing work and family responsibilities.

Kalliath and Brough (2008)- Researchers have defined WLB as "The individual's perception that work and non-work activities are compatible and promote growth by an individual's current life priorities," but this definition and its accompanying measurements vary across studies that investigate WLB and related aspects. For example, the Canadian Department of Labor defines WLB as "a self-defined, self-determined state of well-being that a person can reach, or can set as a goal, that allows them to manage effectively multiple responsibilities at work, at home, and in the community; it supports physical, emotional, family, and community health and does so without grief, stress, or negative impact" (Waters &Bardoel, 2006).

Vijay V. Raghavan, (2010) - The potential stress-relieving effects of telecommuting, staff support and training, and adaptable work schedules. Potential sources of stress for professionals include perceived workload, role uncertainty, work facilitation, and decision latitude. Workplace stress can be mitigated by clarifying employees' roles, increasing work facilitation, and granting them more leeway in setting their schedules.

SahanaCharan (2007), - If you already know that working long hours at a computer and living a hectic lifestyle are bad for your health, here's more evidence that you shouldn't do either of them: experts in mental health are increasingly convinced that more and more people in the IT-enabled services sector succumb to depression as a result of the extreme stress they face on the job.

4. RESEARCH METHODOLOGY:

In this qualitative study, secondary sources accounted for most of the information. This study is based on a literature review, and information was taken from case studies, books, magazines, journals, articles, and internet searches or articles dealing with the subject matter. The data required for the study was found using Google Scholar, Research Gate, and SSRN.

5. WORTH OF COLLISION OF WORK-LIFE BALANCE AND STRESS WITH JOB SATISFACTION TO THE WOMEN EMPLOYEES:

Dissatisfaction with one's employment is the focus here, specifically as it relates to the interaction of work-life balance and stress with one's performance on the job. Since there isn't a stress-management or work-life balance policy that works for everyone. Stress, work-life balance, and job satisfaction requirements vary with one's age, stage of life, and professional trajectory. It's essential to tailor your commitments and necessities for your business. Discord between one's professional and private lives can lead to emotional distress. Furthermore, it is highly stressful for female workers if they achieve their goals but must be adequately rewarded for their efforts.

Due to factors such as the gradual lowering of trade barriers, cutting-edge technological advancements, an internationally organized marketplace, fierce competition in business, and shifting family and population patterns, women's fields of employment in India are rapidly evolving. When both husband and wife have to work to support their increasing family and aging parents, the strain on women's mental health is amplified. Constant stress can disrupt women's mental health by making them feel powerless and giving them the impression that they'll never have enough time to find a healthy work-life balance. Reduced proficiency, increased absenteeism, low employee morale, ineffective teamwork, and health-related issues result from ineffective stress management and work-life policies. Aside from that, several other elements contribute to women's favorable or unfavorable attitudes toward their jobs. In addition, some workers may be pleased with some aspects of their jobs while unhappy

ISBN: 978-81-965283-7-9

with others. There is a relationship between how you feel about your work-life balance and how satisfied you are with your job.

When women experience mental stress, it often manifests physically in the form of ailments like headaches, gastritis, sore muscles, a lack of enthusiasm, low morale, etc., and it can even lead to more severe conditions like high blood pressure, diabetes, depression, and poor job performance. These problems cause anxiety and dissatisfaction in the workplace, especially for women. Banks can differentiate themselves from mediocre organizations by developing healthy work-life management and stress programs effectively detecting pressures, causes for work-life imbalance, and job discontent.

The problems we describe here are manageable but require constant work and examination. It's easy to let work consume you. Managing stress and striking a work-life balance is possible if you can identify what is most important to you and work toward achieving it. Applying management skills greatly enhances an individual's ability to deal with stress, professional demands, and family obligations.

Many things might cause stress at work, and it would be impossible to eliminate them. It could be hazardous as well. However, there are situations when stress can help you. The ability to deal with stress is the determining factor in whether it positively affects motivation and productivity. Organizational methods to manage or lessen some of the critical causes of stress in the banking industry might be helpful to working women. It's safe to say that people's levels of job satisfaction vary from person to person, just as the factors contributing to that happiness shift with time. [3, 9, 10, 12-16].

Employees with a healthy work-life balance report higher levels of job satisfaction, reduced stress, and decreased rates of absenteeism and illness.

While the ideal work-life balance may look different for various people, being constantly available at odd hours is only viable in the short term.

Maintaining a healthy work-life balance is crucial to your happiness and success at home and in the workplace.

6. IMPACT OF WORK-LIFE BALANCE AND STRESS ON JOB SATISFACTION:

(1) A significant component of being happy in your career is being appreciated. It's a statement of approval or disapproval made by a higher-up peer, colleague, manager, client, or general public member—stress and discontent at work result from being overlooked for awards too often.

(2) Employees' timeliness, teamwork, customer service, work direction responsibility, group conduct, peer communication, and managerial abilities all suffer when they have a low sense of self-worth and morale due to stress and conflict in their work-life balance.

(3) The promotion is necessary due to the employee's seniority and provides a psychological benefit. Promotion means an actual increase in one's level of employment. Getting a raise in rank means more responsibility and less micro-management.

(4) When employees get disenchanted with their work, it devastates their ability to think creatively, learn new things, and be original.

(5) Employees who struggle to maintain a healthy work-life balance can negatively influencetheir co-workers' corporate culture, work environment, and morale.

(6) When employees complain, the turnover, operating profit, and the balance sheet all take a hit.

(7) Monthly wage or compensation provides the most financial security for workers, especially women. The role of compensation in these chains of events is crucial. There's no denying that money can be a significant factor in whether or not an employee is happy. Employees may become dissatisfied if their pay is not competitive.

(8) Female workers' happiness in the workplace is just as dependent on their physical working conditions and the availability of amenities. However, firm administration and policy also significantly affect employees' happiness. These should be shaped with an eye on what the staff wants and needs.

(9) Working primarily with other women, women employees emphasize building solid relationships within the office. Interactions with superiors, reports, and co-workers all figure into this. The employee's morale, job satisfaction, and output will all increase if they feel they have good working relationships with their co-workers.

(10) If left unchecked, stress can lead to serious health problems like high blood pressure, heart disease, obesity, and even diabetes.

(11) Moderate stress can help us focus and get things done, but too much can cloud our judgment. There's a chance that your mind will get all mixed up. The work itself may be satisfying, but it still leads to frustration.
(12) The way people react to pressure might vary considerably. Emotional repercussions include but are not limited to irritability, impatience, rage, irritation, fear, worry, self-doubt, panic, melancholy, feelings of inadequacy, insecurity, hopelessness, discontent, emotional withdrawal, and depression.
(13) The way people interact with one another can shift when they're under stress. We may become less sociable, less sympathetic, more antagonistic, and less empathetic. We become less tolerant, lose our cool quickly, and start fights when stressed out, all of which lower our job satisfaction [2, 4, 6, 11].
(14) You may improve your mood, health, and productivity by learning to control your stress and freeing yourself from its grip.
(15) Mental or bodily tension is what we mean when we talk about stress. Anything that causes nervousness, frustration, or anger could set it off—the body's physical reaction to an environmental stressor. Stress can be beneficial in tiny doses, such as when it motivates you to take precautions that keep you safe or pushes you to complete a task on time.
(16) A person's unhappiness with their employment can be significantly affected by their working conditions, co-workers, job security, and duties.
(17) The quality of one's working life improves along with one's level of job satisfaction.
(18) If an employee is happy, he is likelier to perform well.

7. FIVE STEPS TO ACHIEVING A SUCCESSFUL WORK-LIFE BALANCE:

(1) Create must-do lists and find the fun in accomplishing them in the educational sector.
(2) Maintain a healthy work-life balance by prioritizing physical activity like games and exercise in the educational sector.
(3) We must put in the time and effort required to complete the tasks without interruption, which may happen in the educational sector, like school and college.
(4) It would help if you learned to assign tasks effectively to the educational sector's staff.
(5) Schedule downtime as and when it's needed.

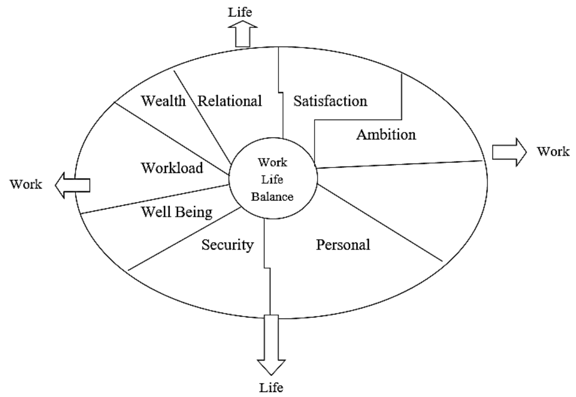


Fig.1: Work-Life Balance

SOME OF THE MAJOR TIPS TO FIND PERFECT WORK-LIFE BALANCE:

(1) Analyze your work and personal life in detail:

If you don't prepare, you're intending to fail. To get things done on schedule, you need to examine your daily habits and make a strategy.

Sixty percent of respondents to a poll conducted by Work Front blamed overbearing managers for their inability to strike a healthy work-life balance. However, you can make your supervisor look good by meeting all your deadlines and preparing beforehand.

How long do you typically stay in the office each day? In what ways are you able to spend your free time? Please write down your observations as you make them. Answer the following questions using this paper:

Q. Are you overwhelmed at work? Are you satisfied with your performance?

Q. Can you devote quality time to your family and personal pursuits?

The information you submit here will help you assess your current level of work-life harmony.

If you can't handle your workload, talk to your manager about setting up a more manageable timetable. These days, employees at most organizations have considerable leeway in their schedules. Many other options exist if your current employer is not receptive to your proposal.

Consider if there is room to enhance how you currently operate. Working more

ISBN: 978-81-965283-7-9

innovative rather than more complex is the key.

(2) FIND A WAY TO REJUVENATE YOURSELF:

A survey of more than 2300 workers found that only 25% took advantage of their vacation days. Sixty-one percent of these people reported working while on vacation. People are sure to feel overwhelmed and exhausted.

Organizations frequently request time off from their top workers torefresh them before a big project. Similarly, you and the rest of the world can benefit significantly from taking a short break from your duties.

How busy is it where you work? Do you feel the urge for a little break? Take a break from work for a few days or weeks to get your life back on track. Make time for introspection and prioritize your own needs. And please don't take any work calls while you're out!

(3) PREPARE A TO-DO LIST FOR THE HOME AND OFFICE:

People who are self-controlled and well-organized tend to fare best in life. Most of them have made a list of things they want to accomplish and when they want to accomplish them.

While many people keep a to-do list or use some task management software at work, fewer do so at home, which makes it simple to disregard your obligations.

Please keep a personal to-do list as well as a professional one. And it need not be made up entirely of monotonous choral pieces. In your spare time, how do you like to spend it? Play golf or bowl occasionally? Make sure to schedule in time to indulge in your preferred pastimes.

(4) SET BOUNDARIES AT WORK:

Eighty-nine percent of workers feel it's essential that their bosses only call or text them during regular business hours.

Most people often fail to separate their personal lives from their professional ones. It can be challenging to take care of oneself if you're always available to your co-workers, even after hours.

As a result, there may be times when others require you to step in and help out immediately. But that frequency is unacceptable. Having more time on your hands will allow you to improve many aspects of your life.

(5) MASTER THE ART OF DELEGATION:

Learning to delegate effectively is a crucial skill for every leader.

Caring for every team member's issues can sometimes be enticing and convenient. Since you have been in his shoes before and know what to do, directly assisting him may seem natural.

Although it may appear complicated initially, things would have gone more smoothly if he had outsourced some of his responsibilities to capable staff members.

Training new people and giving them responsibility inside the group is crucial. You may finally relax a little bit at the office. Furthermore, you can inculcate a sense of responsibility in them through delegation.

(6) Get rid of unnecessary tasks:

The first piece of advice is to take stock of your routine. Do any of the items on the list seem superfluous to you? It could be a straightforward procedure that doesn't require participation or a meeting you're not required to attend.

It should only be a priority if it is crucial to your daily life. We must now proceed.

(7) IGNORE YOUR SMART PHONE AND EMAILS WHEN NECESSARY:

Half of the Work Front poll respondents said that being available 24/7 has cut into their quality time with their families. We need help with our reliance on smart phones, even outside a call center capacity.

You could be responding to emails, reading Facebook messages, researching the stock market, etc. There is no problem with this at all. However, it becomes problematic if you do this frequently, interfering with your time with loved ones.

If you can wait till tomorrow to respond to any emails, put down your smartphone and do so.

(8) Spend your weekends wisely:

Weekends off are a perk of working a typical 9 to 5. It's the weekend, so kick back, have fun, and recharge to take on the week with renewed vigor.

Take a nap, picnic with the family, play golf with your pals, or do whatever brings you joy. Put down the computer and stop thinking about the office. Your loved ones are counting on you, so don't shortchange them for the sake of your career.

Maintaining their happiness will have a positive effect on your own life.

(9) LEARN TO SAY NO WHEN REQUIRED:

Mastering the art of saying no is essential if you're constantly being bombarded with requests and inquiries at work. Your boss, co-workers, classmates, or customers could all be potential sources of requests. Whoever is asking should only accept responsibilities that they can handle.

Say no when you have to, but be sure you can back it up by expressing how busy you already are. Please accomplish this and avoid devoting less time to the things that truly matter to you in favor of tasks that seem more urgent but have little bearing on your life.

(10) AVOID MULTI-TASKING OFTEN:

We all know that multi-tasking isn't our strong suit. Our team shifts its attention swiftly from one task to the next. Once you lose attention on a task, studies suggest it might take 15 to 20 minutes to get it back.

Think about shifting your attention 10–20 times a day; you'll see why this is a recipe for catastrophe. In addition, there are countless things to consider while seated at a computer. Mastering focus is challenging, but there are some easy steps you can do to cut down on your distractions. You can take practically instantaneous action, such as moving to a separate conference room, closing the email tab, or switching to Skype or an internal messenger.

8. TIPS FOR STRESS MANAGEMENT:

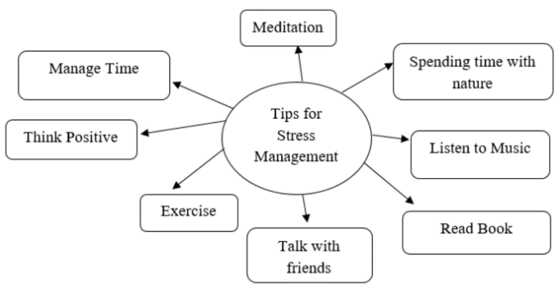


Fig. 2: Tips for Stress Management

9. Stress Management Skills

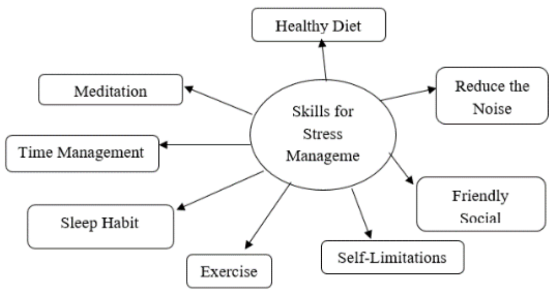


Fig. 3: Skills for Stress Management

10. DISCUSSIONS AND SUGGESTIONS:

(1) Allow for Telecommuting and Flex Time.
(2) Get your staff to concentrate on getting work done.
(3) Activate the rest periods.
(4) Check in on the staff's workload regularly.
(5) Lead the employees by example.
(6) Give employees time to volunteer.
(7) Reconsider time off.

11. CONCLUSION:

The term "work-life balance" refers to the harmony between work and other essential aspects of one's life, such as family and hobbies. Meanwhile, stress management provides numerous methods for coping with pressure and hardship (adversity).

Women's participation in the labor force now demonstrates no inherent distinction between the sexes. Some businesses even claim that their success is mainly due to the contributions of their female employees. It's encouraging to see more women actively participating in public life. However, every woman also has her own home and personal life to keep track of.

The importance of the boundary between one's professional and private life has grown in recent years, along with the pressures of the modern workplace. Research has shown that such a circumstance can negatively affect a person's physical and mental health. Workers must maintain their work-life balance by minimizing stress and enjoying their employment.

Employers will want to consider issues like work-life balance and rising employee stress when formulating policies for their staff. The secret to success is a positive attitude; master your tension and anxiety by learning efficient coping mechanisms.

Those who have chosen a line of employment that is both physically and mentally taxing often feel dissatisfied with their jobs. Many things might go wrong on the job, and it's hard to eliminate them all. The results of being dissatisfied can be beneficial at times. Depending on the person, this can invigorate and inspire them to push themselves to greater heights of productivity. In conclusion, the level of job satisfaction varies as the satisfaction factors change.

Burnout, stress, and absences all hurt productivity, starting with a lack of work-life balance. Meanwhile, those who can maintain a healthy equilibrium between their professional and personal lives are more likely to believe they have control over both.

ISBN: 978-81-965283-7-9

REFERENCES:

1. Gururaja, Maiya Umesh, Sanatombi Elsa, George (2013). Anice, International Journal of Humanities and Social Science Invention 2(3), 52-54.
2. Hughes Jaanna, Bozionelos Nikos, (2007). Personnel Review 36(1), 145-154.
3. Jamal Nazrul Islam, Haradhan Kumar Mahajan, Rajib Datta (2012). A Study on Job Satisfaction and Morale of Commercial Banks in Bangladesh, IJER, x(4), 153-173.
4. Latif Khalid, Shahid Muhammad Naeem, Sohail Naeem, Shahbaz Muhammad, Inter-disciplinary Journal of Contemporary Research in Business 3 (8)(2011), 235-242.
5. KumariGeeta, PandeyK.M., (2012). Xxxxxx. International Journal of Innovation Management and Technology 2 (5), 360-367.
6. Kumari Lalita, (2012). Xxxxxx. IJEMR, 2(2), 03-13.
7. KamalRaj, Sengupta Debashish, (2008). Prajnan, 37(3), 229-245.
8. Rao, V, S, P, (2009). Human Resource Management–Text and cases. Excel books – NewDelhi, pp. 299-306.