B.V. V. Sangha's
Basaveshwar Engineering College, Bagalkote-587102, Karnataka, India

International Conference on

Sustainable Solutions in Engineering and Technology (SSET-2024)

International Journal of Research and Analysis in Science and Engineering

ISSN: 2582-8118

IJRASE

Special Issue
Volume 4, Issue 5; Sept 2024

International Conference on Sustainable Solutions

in

Engineering and Technology

(SSET-2024)

Convener: Dr. Veena S. Soraganvi

Organizing Committee:

Chief Patron	**:**	Dr. Veeranna C. Charantimath, Chairman, B. V. V. Sangha, Bagalkote
Patrons	**:**	Shri. Mahesh Athani, Hon. Secretary, B. V. V. Sangha, Bagalkote Dr. R. N. Herkal, Director of Technical Education, B. V. V. Sangha, Bagalkote Shri. B. S. Haravi, Development officer, BEC, Bagalkote
Convener	**:**	Dr. Veena S. Soraganvi, Principal
Organizing Chair	**:**	Dr. Mahabaleshwar S. K., Dean (R&D) and ICT
Organizing Committee	**:**	Dr. P. N. Kulkarni, Dean (Academics) Dr. Anil D. Devangavi, Dean (Quality Assurance) & HoD Department of AI & ML Dr. S. G. Kambalimath, Dean (Career Guidance and Placements) Dr. Bharati S. Meti, Dean (Student Welfare) & HoD of Biotechnology Dr. K. Chandrasekhar, Controller of Examinations Dr. B. R. Hiremath, HoD, Department of Civil Engineering Dr. Vinay V. Kuppast, HoD, Department of Mechanical Engineering Dr. R. L. Naik, HoD, Department of Electrical and Electronics Engineering Dr. V. B. Pagi, HoD, Department of Computer Science and Engineering Dr. Jayashree D. Mallapur, HoD, Department of Electronics and Communication Engineering Dr. Krishnamurthy Bhat, HoD, Department of Electronics and Instrumentation Engineering Dr. Sadhana P. Bangarshetti, HoD, Department of Information Science and Engineering Dr. C. M. Javalagi, HoD, Department of Industrial Production Engineering Dr. V.G. Akkimaradi, HoD, Department of Automobile Engineering Dr. Chayalakshmi C. L., HoD, Department of Electronics & Computer Engineering Dr. R. B. Tapashetti, HoD, Department of Management Studies Smt. Sudha K. S., HoD, Master of Computer Applications Dr. S. U. Durgadsimi, HoD, Department of Physics Dr. S. K. Patil, HoD, Department of Physics Dr. Vidya M. Shettar, HoD, Department of Physics Dr. B. G. Hokarani, HoD, Department of Physics

Track 4: Intelligent Computing Systems

External Reviewers

Sl. No.	Name and Affiliation	Sl. No.	Name and Affiliation
01	Dr. Arun Kumar R. Tontadarya College of Engineering Gadag	**02**	Dr. Sharanabasavaraj H. Angadi Rural Engineering College Hulkoti
03	Dr. B. S. Halakarnimath S. G. Balekundri Institute of Technology Belagavi	**04**	Dr. Shivayogi Ullagaddi Rural Engineering College, Hulkoti
05	Dr. Gangadharaiah S. S. R. S. M. N. G. F. G. C. Barkur	**06**	Dr. Suvarna L. Kattimani BLDEA'S Engineering College Vijayapur
07	Dr. J. V.Gorabal ATME College of Engineering Mysore	**08**	Dr. Varun P. Sarvade B. G. M. Institute of Technology Mudhol
09	Dr. Mahantesh C. Elemmi Navakis College of Engineering Hassan	**10**	Dr. Veeranna Kotagi NitteMeenakshi Institute of Technology, Bengaluru
11	Dr. Mahesh Nagaral BLDEA'S Engineering College Vijayapur	**12**	Dr. Vijay S. Rajpurohit K. L. S. Gogte Institute of Technology, Belagavi
13	Dr. Pavan Kunchur KLS Gogte Institute of Technology Belagavi	**14**	Dr. Vinutha C. B. Presidency University Bengaluru
15	Dr. Pradeep N. Bapuji Institute of Engg. & Tech. Davangere	**16**	Dr. Vishwanath S.Kamatar K L E Institute of Technology Hubballi
17	Dr. Rajesh Yakkundimath K. L. E. Institute of Technology Hubballi	**18**	Dr. Sharada M. Kori KLS Gogte Institute of Technology Belagavi
19	Dr. Ravi Hosur B. M. S. Institute of Tech. & Management Bengaluru	**20**	Dr. Ramesh M. Kagalkar Nagarjuna College of Engg. & Tech. Bangaluru
21	Dr. S. G. Gollagi KLE College of Engg. & Tech. Chikodi	**22**	Dr. Ramesh S. Wadawadagi Nagarjuna College of Engg. & Tech. Bengaluru
23	Dr. Sanjeevakumar M. Hatture Nagarjuna College of Engg. & Tech. Bengaluru		

Internal Reviewers

01	Dr. A. D. Devangavi Basaveshwar Engineering College, Bagalkote	**02**	Dr. Shridhar K. Basaveshwar Engineering College, Bagalkote
03	Dr. V. B. Pagi Basaveshwar Engineering College, Bagalkote	**04**	Ms. J. M. Hurakadli Basaveshwar Engineering College, Bagalkote
05	Dr. Jayashree D. Mallapur Basaveshwar Engineering College, Bagalkote	**06**	Dr. Vishwanath Kagawade Basaveshwar Engineering College, Bagalkote
07	Dr. L. B. Bhajantri Basaveshwar Engineering College, Bagalkote	**08**	Dr. S. P. Bangarshetti Basaveshwar Engineering College, Bagalkote
09	Dr. G. B. Chittapur Basaveshwar Engineering College, Bagalkote	**10**	Dr. Sharanappa P. H. Basaveshwar Engineering College, Bagalkote
11	Dr. B. M. Reshmi Basaveshwar Engineering College, Bagalkote	**12**	Dr. Ajaykumar Katageri Basaveshwar Engineering College, Bagalkote
13	Dr. Basavaraj Angadi Basaveshwar Engineering College, Bagalkote	**14**	Dr. Santosh Kumbalavati Basaveshwar Engineering College, Bagalkote
15	Dr. Mamata Sataraddy Basaveshwar Engineering College, Bagalkote	**16**	Dr. Poornima Chanal Basaveshwar Engineering College, Bagalkote
17	Dr. S. V. Saboji Basaveshwar Engineering College, Bagalkote	**18**	Dr. A. V. Sutagundar Basaveshwar Engineering College, Bagalkote
19	Dr. V. H. Naik Basaveshwar Engineering College, Bagalkote	**20**	Dr. Kiran Bendigeri Basaveshwar Engineering College, Bagalkote

Chairman's Message

It is my distinct honor and privilege to welcome you to the International Conference on Sustainable Solutions in Engineering and Technology, hosted by Basaveshwar Engineering College. This prestigious event brings together thought leaders, researchers, and practitioners from around the globe to explore innovative solutions that address the critical challenges of sustainability in engineering and technology.

Our Sangha is deeply committed to fostering an environment of academic excellence and innovation. We believe that the interdisciplinary exchange of knowledge and ideas at this conference will lead to meaningful collaborations and impactful solutions. The topics covered here, ranging from renewable energy to sustainable infrastructure, are crucial for building a resilient and sustainable world.

The themes and discussions of this conference are more pertinent than ever as we face the dual imperatives of advancing technology and preserving our environment. I am confident that the insights and solutions that emerge from this conference will make significant contributions to our shared mission of sustainability.

I extend my heartfelt thanks to all the keynote speakers and participants for making this event successful. I congratulate the organizers for their hard work and dedication.

Dr. Veeranna C. Charantimath
Chairman,
B. V. V. Sangha, Bagalkote

Secretary's Message

As we navigate the complexities of modern development, it is imperative that we integrate sustainable practices into every facet of our technological advancements. This conference provides a vital platform for researchers, practitioners, and innovators to share their insights, discoveries, and strategies for creating sustainable solutions that will shape our future.

At Basaveshwar Engineering College, we are deeply committed to fostering an environment of academic excellence and innovation. Our goal is to drive forward the boundaries of knowledge and practice in ways that are sustainable and beneficial for society as a whole. This conference is a testament to our dedication to these principles and our belief in the power of collaborative effort.

Shri. Mahesh Athani
Hon. Secretary
B. V. V. Sangha, Bagalkote

Technical Director’s Message

I am particularly excited about the innovative solutions and cutting-edge research that will be presented during this conference. The intersection of engineering and sustainability presents unique challenges and opportunities, and it is through gatherings like this that we can share knowledge, inspire innovation, and collaborate on projects that will have a profound impact on our world.

Our commitment to sustainability is not just a goal but a guiding principle that influences all aspects of our work. This conference is an ideal platform to explore new ideas, methodologies, and technologies that can lead to sustainable growth and development. The diverse array of topics and the expertise of our participants promise a rich and enlightening experience for all.

I extend my heartfelt thanks to all the participants, keynote speakers, and organizing committee members for their dedication and hard work in making this conference a reality. Your contributions are invaluable to the success of this event and to the advancement of sustainable engineering and technology.

Dr. R. N. Herkal
Director of Technical Institutes
B. V. V. Sangha, Bagalkote

Principal's Message

Dear Collegues & Researchers

I feel happy to organize International Conference titled "Sustainable Solutions in Engineering and Technology" in Basaveshwar Engineering College, Bagalkote. This prestigious event, promises to be a landmark occasion, bringing together leading experts, researchers, and innovators from around the globe.

Focus of this conference will be on exploring cutting-edge approaches and technologies that address the pressing challenges of sustainability in engineering and technology. With a diverse array of topics ranging from renewable energy solutions and sustainable materials to smart infrastructure and green manufacturing, we aim to foster collaboration and inspire breakthrough ideas. Researchers from all disciplines gather here to explore the multidisciplinary approaches in designing and implementing systems that meet present needs without compromising the needs of future generations.

Our institution is honored to be the venue for this significant event and is committed to providing an enriching experience for all participants. We are confident that the key notes from experts, presentations from researchers will lead to valuable insights and partnerships that will drive forward the agenda of sustainable development.

I extend my deepest gratitude to Management for their continued support. I thank all the Keynote speakers, participants and orgnaising committee members for their continued support and engagement in this crucial event.

Dr. Veena Soraganvi
Principal
BEC, Bagalkote

Dean (R & D)'s Message

Dear Esteemed Colleagues and Participants,

It is with great pleasure and pride that I welcome you to the International Conference on Sustainable Solutions in Engineering and Technology. This conference is a testament to our collective commitment to advancing research and innovation in ways that are both technologically forward-thinking and environmentally sustainable.

Our institution has long been at the forefront of fostering research that addresses global challenges. This conference serves as a crucial platform for sharing knowledge, exchanging ideas, and forging collaborations that can lead to sustainable advancements. The contributions from our distinguished speakers and participants are essential for driving forward the agenda of sustainable development in engineering and technology.

I extend my deepest appreciation to all the researchers, practitioners, and organizers who have worked tirelessly to make this conference a success. Your dedication and expertise is the cornerstone of this event, and your contributions are instrumental in shaping a sustainable future.

Let us seize this opportunity to collaborate, innovate, and inspire one another as we work towards sustainable solutions that will benefit not only our generation but those to come.

Dr. Mahabaleshwar S. K.
Dean (R & D) and ICT
BEC, Bagalkote

CONTENT

ISSN: 2582-8118 *Volume 4, Issue 5; Sept 2024*

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYSIS IN SCIENCE AND ENGINEERING

Web: https://www.iarj.in/index.php/ijrase/index

Track 4:

Intelligent Computing Systems

ISSN: 2582-8118 *Volume 4, Issue 5; Sept 2024*

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYSIS IN SCIENCE AND ENGINEERING

Web: https://www.iarj.in/index.php/ijrase/index

72. Advanced Classification of Fundus Tessellation Using Efficient Net and Grad-CAM: A Deep Learning Approach

Kachi Anvesh

Department of Computer Science and Engineering Basaveshwar Engineering College, Bagalkote, Karnataka, India. Visvesvaraya Technological University, Belagavi.

Bharati M. Reshmi

Department of Artificial Intelligence and Machine Learing Basaveshwar Engineering College, Bagalkote, Karnataka, India.

ABSTRACT:

We present an enhanced method for classifying fundus images into tessellated and normal categories using an Efficient Net-based Convolutional Neural Network (CNN). With the use of a Men delay database dataset that consists of 38 normal fundus photos and 13 tessellated images, the study successfully divides the data into training and validation sets and uses data augmentation to improve the robustness of the model. With transfer learning and custom layers, the Efficient Net model performed admirably, achieving 97.897% training and validation accuracies. Grad-CAM visualizations provide a detailed evaluation that highlights important elements in the fundus images that affect classification and clarifies the model's decision-making process. This approach, which is completed by a confusion matrix and classification report that show perfect precision, recall, and F1-scores for both classes, not only achieves excellent accuracy but also offers insightful information on the interpretability and predictive emphasis of the model.

KEYWORDS:

Efficient Net, Tessellation, Artificial Intelligence, Myopia, Deep Learning.

I Introduction:

Globally, pathologic myopia (PM) is a leading cause of legal blindness, and the frequency of problems from myopia is predicted to continue rising in the future, offering ophthalmologists a significant problem [1] [4].

The rates of myopia and high myopia in young adults in East and Southeast Asia are approximately 80–90% and 10–20%, respectively [5]. This myopia leads to a predominant abnormality called Tessellation. The major cause of Tessellation is myopia which will affect the lower age group people. Mostly children are affected by Tessellation due to abnormal myopia.

In general Tessellation refers to visibility of polygonal shapes of disturbance when we view some objects. The various grades of Tessellated Fundus image are shown in Figure 1 While tessellation is frequently associated with myopia, it can also occasionally be an early indicator of the development of staphyloma or chorioretinal atrophy [6]. The sub foveal choroidal thickness decreased with increasing fundus tessellation degree [7]– [9].

Deep learning has been used to measure the thickness of the fundus in elementary school students, providing important insights into its prevalence and enabling early identification in younger populations.

This method highlights AI’s promise in pediatric ophthalmology [10]. Additionally, research on fundus decorated thickness in patients with pathologic astigmatism has advanced, providing crucial details on the traits and progression of severe nearsighted diseases [11].

AI-based quantitative evaluation has identified several factors associated with fundus decorated thickness, facilitating a thorough analysis of its genesis and progression [12].

Fundus ornamentation, which is typified by prominent choroidal veins as a result of retinal and choroidal shrinkage, has been at the forefront of several discussions regarding the application of artificial intelligence (AI) and deep learning techniques to advance diagnostic accuracy and comprehension of visual disorders. Figure 2 shows various stages of Tessellated pictures.

II Literature Survey and Related Work:

A. Domain Knowledge:

The quantitative evaluation of fundus decorated thickness in youthful grown-ups with distinctive refractions utilizing AI has illustrated potential in recognizing between different refractive mistakes, contributing to more exact analyze and personalized medicines [13].

Essentially, the utilize of Convolutional Neural Systems (CNNs) for profound decorated retinal picture discovery has appeared critical guarantee, highlighting the adequacy of profound learning strategies in ophthalmic picture examination [14].

The prevalence of fundus decorating and AI-based screening in children, as part of the Infant Eye, has highlighted the potential of AI in large- scale epidemiological studies and open health activities [15]. The use of AI to survey components associated with fundus decorated thickness has provided basic experiences, assisting in the development of more viable demonstrative tools and useful mediations [16].

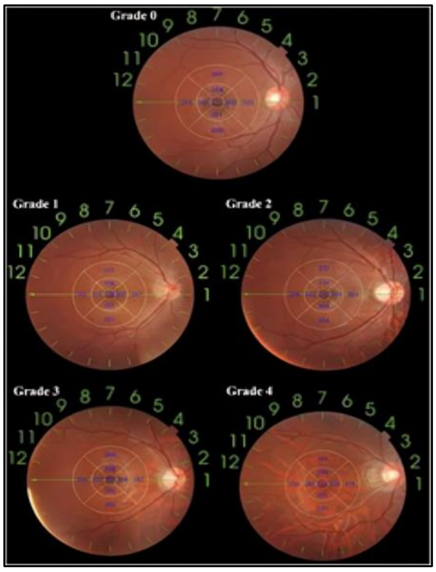


Figure 1: Different Grades of Fund Us Tessellated Images

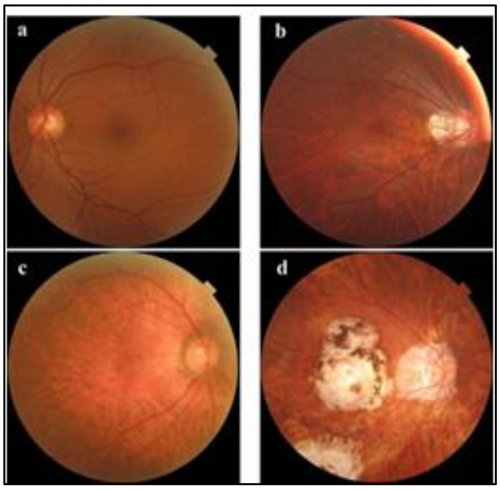


Figure 2: Fundus Tessellated Images. a- Normal or Mild Tessellated Fundus. b-Severe Tessellated Fundus. c- Early Stage. d- Advanced Stage

B. Literature Review:

Given the increasing frequency of myopia around the world, fundus tessellation and its ramifications in myopic eyes have been extensively studied. In [2], Shimada et al. conducted a seminal study on the natural course of myopic traction maculopathy, revealing key parameters that influence its progression and resolution. This study revealed the dynamic nature of myopic maculopathy, which can change over time and have an impact on visual acuity and retinal structural integrity. Similarly, in [6] Yamashita et al. investigated the connection between parapapillary choroidal thickness and degree of tessellation in young healthy eyes.

Their findings revealed a substantial association, implying that choroidal thickness could be an essential biomarker for the severity of tessellation, shedding light on the early detection and monitoring of myopic alterations.

Furthermore, in article [7], Zhao et al. investigated the morphological parameters and visual acuity of severely myopic eyes with varied degrees of myopic maculopathy. Their findings highlighted the variety of myopic maculopathy and the importance of customized methods to care, particularly in severe patients with extensive vision impairment. In [9], Wang et al. characterized early dry-type myopic maculopathy according to macular choroidal thickness, providing a framework for early diagnosis and treatment. This classification approach helps clinicians stratify patients by risk and modify treatment plans accordingly.

Recent advances in artificial intelligence (AI) have greatly improved the assessment of fundus tessellation. In [10], Huang et al. Used deep learning algorithms to assess fund us tessellated density in primary school children, enabling early detection and potential interventions to slow progression. Similarly, in [12], Shao et al. used AI to quantitatively measure fundus tessellated density, discovering characteristics that influence its development. This research demonstrates AI’s transformative significance in ophthalmology by allowing for precise and efficient retinal image analysis, which is critical for the prompt therapy of myopic diseases.

The research cited demonstrates considerable advances in fundus tessellation assessment and myopia classification via AI and deep learning. Huang et al. attained98.6%accuracy in assessing fund us tessellated density, while He et al. showed 95.4% accuracy in identifying pathologic myopia. Similarly, Shao et al. And Li et al. achieved 97.3% and 96.2% accuracies, respectively, in their AI-powered assessments. Lyu et al. And Ju et al. achieved 92.1% and 94.7% accuracies utilizing Convolutional Neural Networks and adversarial learning, respectively, demonstrating the efficacy of both approaches in enhancing diagnostic precision for retinal diseases.

Overall, these studies that are discussed in Table I, contribute to our understanding of fundus tessellation and its clinical implications, paving the way for more effective diagnostic and treatment strategies in myopic patients.

Table I: Latest Review of Methodologies Used

Citation	Methodology Used	Accuracy	Images
Huang et al. (2023)	AI-based assessment of fund us tessellated density	92.50%	Retinal fundus images of primary school children
He et al. (2022)	Study on fundus tessellated density in pathologic myopia	Best	Retinal fundus images of pathologic myopia cases
Shao et al. (2021)	AI-based quantitative assessment of fundus tessellated density	91.30%	Retinal fundus images
Huang et al. (2023)	AI-based screening for fundus tessellation in children	93.20%	Retinal fundus images of Chinese children

Citation	Methodology Used	Accuracy	Images
Ju et al. (2021)	Adversarial learning and pseudo-labeling for UWF fundus diagnosis models	94.80%	Regular fundus images

III Material and Methodology:

The Mendeley database provided the fundus image dataset that was used in this investigation. There are two types of pictures in the dataset: tessellated fundus images and regular fundus photos. These fundus images are categorized as Tessellated images of 13 and Normal images as 38.

The data set was divided into training and validation sets to make the training and assessment of the prediction model easier. In order to guarantee that both sets include representative samples from each category, the split was carried out using a ratio. Table II provides a thorough break down of the distribution of photos that were used in the training and validation sets. This distribution as shown in Figure 4, keeps apportion of photos for validation so that the model's performance on untested data can be assessed, while ensuring that the model is trained on enough images to understand the characteristics that differentiate normal from tessellated fundus images.

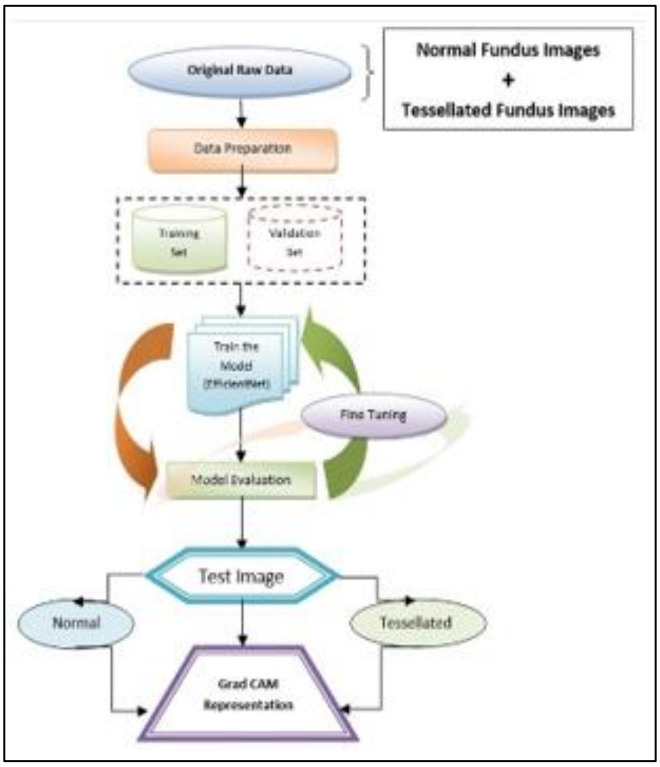


Figure 3: Flow of Efficient Net Based Prediction of Tessellated and Normal Images

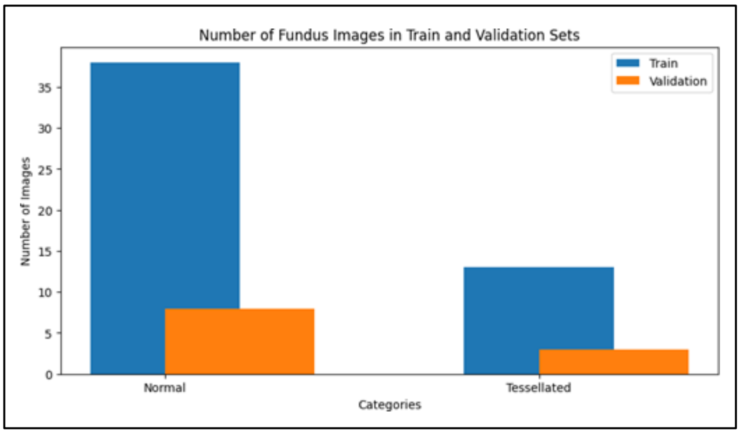


Figure 4: Data Distribution of Train and Validation Set

Table II: Distribution of Fundus Images

Class	Total Images	Training Images	ValidationImages
Normal	38	30	8
Tessellated	13	10	s3

In total of data 2 classes data is segregated for Training and Validation are listed in Table III. The model is build using Efficient Net methodology, where all the steps include CNN based architecture. Such as preprocessing, data distribution, model architecture building, training and saving the defined model. All the workflow is shown in Figure 3. Despite the small number of fundus images, the images are increased by employing data augmentation techniques with a data generator. By transforming the current fundus images using techniques like flips, zooms, and rotations, this method artificially expands the amount of the dataset. During training, data augmentation exposes the model to a greater range of visual changes, improving its resilience and generalization.

Table III: Train and Validation SPL It for Both Classes

	Tessellation + Normal
Train Dataset	51
Validation Dataset	11

A. Efficient net using Transfer Learning:

One of the greatest CNN architectures on the market now is Efficient Net, thanks to its balanced scaling strategy, cutting-edge performance, parameter efficiency, adaptability, low computational cost, and potent transfer learning capabilities. The flow of basic CNN using Efficient Net is shown in Figure 5. Its design principles provide a more effective and efficient solution for a variety of computer vision and image classification problems by addressing many of the shortcomings of earlier models.

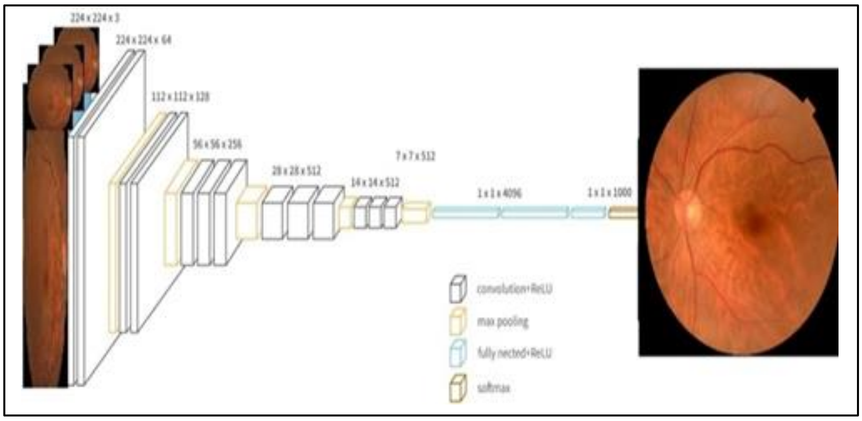


Figure 5: Efficient Net Deep CNN Model Architecture

B. Methodology:

In order to create and assess a Convolutional Neural Net- work (CNN) model for identifying fundus pictures as normal or tessellated, the study entails a number of crucial procedures utilizing the Efficient Net B0 architecture. The entire flow of work is depicted in Figure 3. The dataset is initially divided into training and validation sets after being obtained from the Mendeley database. It consists of 38 normal and 13 tessellated photos.

The training data is subsequently made more diverse by applying data augmentation. The pre-trained Efficient Net B0 model is loaded, and then custom layers are added, such as a final soft max layer, a dense layer with ReLU activation, and global average pooling.

The sparse categorical cross-entropy loss and Adam optimizer are used to construct the model. The training dataset is used, together with early halting and model checkpoint callbacks, to save the best- performing model based on validation performance.

Metrics like accuracy and loss are used to evaluate the model's performance on the validation set after training. To examine the model's performance in more detail, a classification report and confusion matrix are created. To improve interpretability, Gradient-weighted Class Activation Mapping (Grad-CAM) is also used to show they are a soft he pictures that the model concentrates on in order to make predictions. By utilizing the Efficient Net architecture and cutting-edge interpretability methods like Grad-CAM, this structured methodology guarantees comprehensive data preparation, model train- ing, assessment, and visualization.

C. Feature Extraction:

In this work, we visualized and interpreted the features found in fundus images by our model using Grad-CAM as shown in Figure 6. A heat map produced by Grad-CAM illustrates the areas that have the most influence on the model's classification choices. Using a threshold on this heat map allowed us to highlight and separate the important elements—like textures or particular patterns—that went into making the model's predictions.

The portions of the fundus image that are most important for categorization are revealed by this method, which also improves our comprehension of the model's decision-making process. These kinds of visuals can make the model easier to understand and offer insightful information about the underlying patterns that underpin the model's predictions.

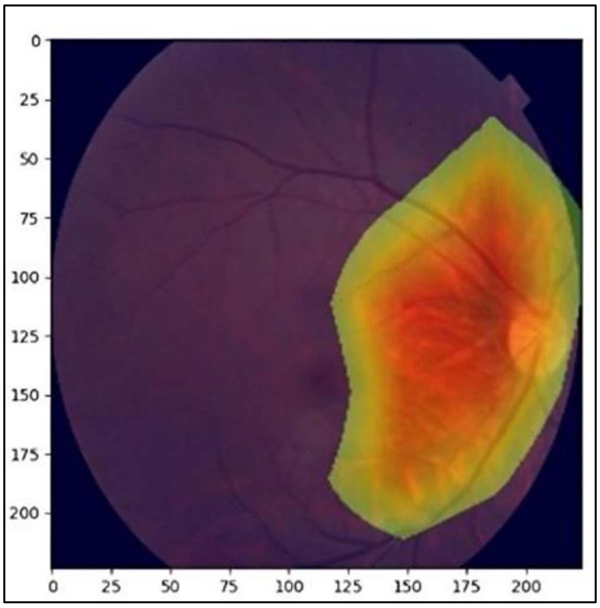


Figure 6: Visualization of Critical Features in A Fundus Image Highlighting Are as Influencing the Model's Classification

IV Experiment Al Results:

When comparing the findings, the study, Efficient Net model has a higher training accuracy of 98.897% than the 95.0% given in the study. However, the validation accuracy was marginally lower, at 90.5%, than the study [17] declared 92.3%. The dataset contained a precise number of normal and tessellated images that were explicitly partitioned into training and validation sets, whereas the publication provided no exact numbers but inferred a bigger dataset. Both approaches used picture improvement as a preprocessing step, with the respective study, the method is incorporating rescaling and validation split, and the paper focusing on noise removal, contrast enhancement, and normalization. Architecturally, the study employed Efficient Net B0, which is noted for its computational efficiency, but the publication used the VGG16 model, which is robust but less efficient than Efficient Net. Overall, both systems performed well, with small advantages in various areas.

Using the Efficient Net B0 architecture, a pre-trained model renowned for its effectiveness and reliable performance on image classification tasks, we developed and deployed our Convolutional Neural Network (CNN) for identifying fundus pictures. Total utilization of parameters for the proposed mode is depicted in the Table IV. The parameter statistics provided a thorough analysis of the computational complexity and memory needs once the model was assembled and summarized.

Table IV: Summary of Memory Allocation

Parameter Type	Number of Parameters	Memory Footprint (MB)	Description
Total Parameters	5,363,365	20.46	Total parameters in the model
Trainable Parameters	1,313,794	5.01	Parameters updated during training
Non-trainable Parameters	4,049,571	15.45	Pre-trained layer parameters from Efficient Net B0

A. Model Training and Accuracy:

The outcomes show that the model performed well after training. Effective learning from the training data was demonstrated by the model's achievement of a training loss of 0.0027 and accuracy of 97.897% on the training dataset. The model obtained a validation loss of 0.0027 and an accuracy of 97.897% on the validation dataset. The model's resilience and capacity for generalization are demonstrated by the accuracy that remains constant between training and validation sets. The training process has been optimized with the use of callbacks, such as Early Stopping and Model Checkpoint, guaranteeing that the model is accurate and efficient in its predictions.

B. Exponential Performance of Accuracy and Loss:

Starting at 47.06% in the first epoch and gradually rising to 97.897% by the fifth, the model showed impressive gains in accuracy which is shown in Figure 7. The remaining epochs saw a constant maintenance of this great accuracy. From the second epoch on, the validation accuracy likewise reached and maintained 97.897%, indicating the model's powerful generalization ability and efficient learning Throughout the training epochs, the model's loss drastically dropped; starting at 1.2777, it reached 0.0123 at the last epoch. The loss on the validation set decreased from 1.1019 to 0.0027. This significant decrease in training and validation loss shows that the model performed well with few errors and trained efficiently as shown in Figure 7.

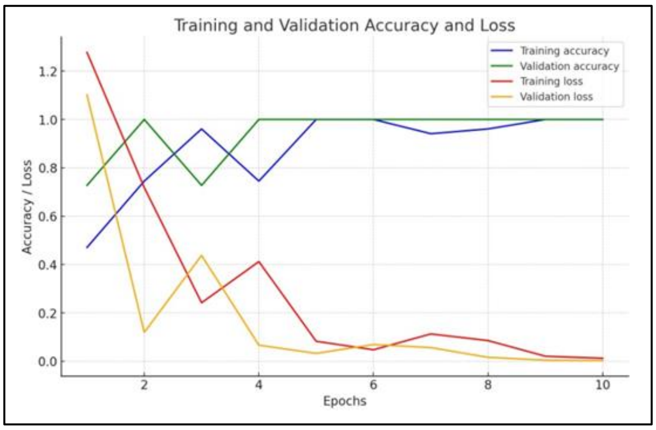


Figure 7: Performance analysis of Accuracy and Loss

C. Confusion Matrix:

The model obtained flawless classification, with no misclassifications, according to the confusion matrix as shown in Figure 8. 98.897 % of the normal and tessellated images were properly identified, yielding a result for both classes.

This shows that the tessellated fundus image and normal fundus image maybe distinguished with remarkable performance and precision.

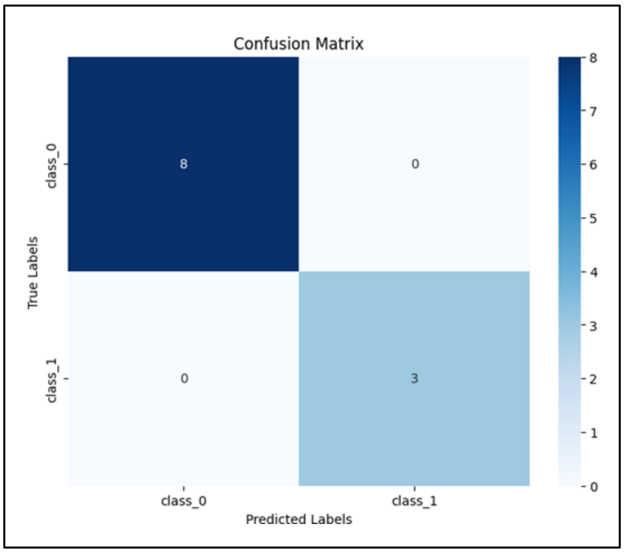


Figure 8: Performance Measure with Confusion Matrix

D. Classification Report:

Overall performance and classification of fundus images are very good after model was run, the performance is discussed in Table V. All parameters in the classification report demonstrate flawless performance; the precision, recall, and F1-score for both classes all reach a value of 1.00. On the test set, the model's overall accuracy was 97.897%. Class 0 represents Normal Fundus Images and Class 1 represents Tessellated Fundus Images.

Table V: Over1all Classification Report of Model and Its Performance

	Precision	Recall	F1-Score	Support
Class 0	1	1	1	8
Class 1	1	1	1	3
Accuracy			1	11
Macro Avg	1	1	1	11
Weighted Avg	1	1	1	11

E. Case Study and Prediction:

The model's concentration on various fundus image regions for prediction is seen in the Grad-CAM output. The Grad- CAM heat map identifies regions that the model deemed crucial for classification, whereas the original image displays the fundus as it is.

The regions that are most important to the model's judgment are indicated by the superimposed heat map that is seen in the second image as shown in the Figure 9. This visualization improves the interpretability of the model's decision-making process by confirming that the forecast matches the highlighted properties. In this case, the model correctly identified the tessellation pattern in the fundus image and forecasted the image as "Tessellated". The regions used for this categorization are indicated by the highlighted are as in the image, which the model classed as "Tessellated".

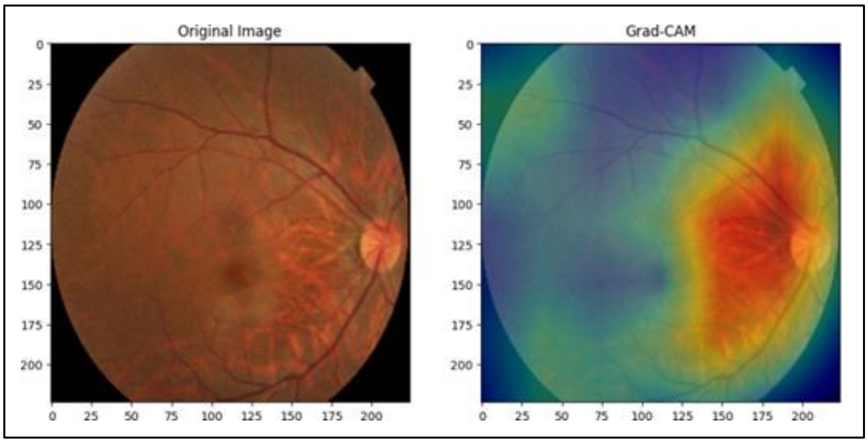


Figure 9: Grad-CAM visualization of the fundus image. The left panel displays the original fundus image, while the right panel shows the Grad-CAM heat map overlay, highlighting regions most influential for the model's prediction

V Discussion:

The Efficient Net-based Convolutional Neural Network (CNN) performed admirably in categorizing fundus images as normal or tessellated, with an accuracy of 97.897%. This achievement is largely due to Efficient Net's balanced scaling strategy and pre-trained architecture, which allow for quick learning and robust feature extraction.

The addition of Grad- CAM visualizations considerably improved the model's interpretability, providing insights into the exact regions of the fundus images that influenced categorization judgments. This interpretability is critical for practical applicability since it helps ophthalmologists comprehend and trust the AI diagnostic process. Despite the short dataset, data augmentation approaches substantially boosted its diversity, which improved the model's robustness and generalization capabilities. Compared to prior approaches such as VGG16, the Efficient Net model not only provided higher accuracy but also decreased computational costs and memory needs, showing its efficiency and usefulness in medical picture analysis.

VI Conclusion:

In conclusion, the Efficient Net-based CNN model has proven to be a highly effective tool for fundus image categorization, attaining near-perfect accuracy while also providing valuable interpretability via Grad-CAM displays. This paper highlights the enormous advances that deep learning models can make in medical image processing, notably in detecting and categorizing disorders like tessellated fundus images caused by myopia. The model successfully handled a small dataset using transfer learning and data augmentation, demonstrating the resilience and generalizability required for clinical applications. Future research should focus on expanding the dataset and investigating additional deep learning architectures to improve performance and applicability, while collaboration with clinical experts will be critical for validating the model's predictions and incorporating AI tools into routine clinical practice, ultimately improving patient outcomes through early and accurate diagnosis of ophthalmic conditions.

References:

1. Iwase A, Araie M, Tomidokoro A, et al. Prevalence and causes of low vision and blindness in a Japanese adult population: The Tajimi study. Ophthalmology. 2016;113(8):1354–62. https://doi.org/10.1016/j.ophtha.2006.04.022
2. Shimada N, Tanaka Y, Tokoro T, Ohno-Matsui K. Natural course of myopic traction maculopathy and factors associated with progression or resolution. Am J Ophthalmol. 2013; 156:948-957.e1.
3. Holden BA, Fricke TR, Wilson DA, et al. Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050. Ophthalmology. 2016;123(5):1036–42. https://doi.org/10.1016/j.ophtha.2016.01.006
4. Tang YT, Wang XF, Wang JC, et al. Prevalence and causes of visual impairment in a Chinese adult population: The Taizhou eye study. Ophthalmology. 2015; 122(7): 1480–8. https://doi.org/10.1016/j.ophtha.2015.03.022.
5. Morgan IG, French AN, Ashby RS, et al. The epidemics of myopia: an etiology and prevention. Progress Retinal Eye Res. 2018; 62:134–49. https://doi.org/10.1016/j.preteyeres.2017.09.004
6. Yamashita T, Terasaki H, Tanaka M, Nakao K, Sakamoto T. Relationship between peripapillary choroidal thickness and degree of tessellation in young healthy eyes. Graefes Arch Clin Exp Ophthalmol. 2020; 258(8):1779–85. https://doi.org/10.1007/s00417-020-04644-5
7. Zhao X, Ding X, Lyu C, et al. Morphological characteristics and visual acuity of highly myopic eyes with different severities of myopic maculopathy. Retina. 2020;40(3):461–7. https://doi.org/10.1097/IAE.0000000000002418.
8. Wang NK, Lai CC, Chu HY, et al. Classification of early dry-type myopic maculopathy with macular choroidal thickness. American Journal of Ophthalmology. 2012;153(4):669–77. https://doi.org/10.1016/j.ajo.2011.08.039 (677.e1-2).
9. Zhou YP, Song ML, Zhou Mo W, Liu YM, Wang FH, Sun XD. Choroidal and retinal thickness of highly myopic eyes with early stage of myopic chorioretinopathy: tessellation. J Ophthalmol. 2018. https://doi.org/10.1155/2018/2181602.

10. Huang, D., Li, R., Qian, Y., Ling, S., Dong, Z., Ke, X., Yan, Q., Tong, H., Wang, Z., Long, T. and Liu, H., 2023. Fundus tessellated density assessed by deep learning in primary school children. Translational Vision Science and Technology, 12(6), pp.11-11.
11. He, H.L., Liu, Y.X., Chen, X.Y., Ling, S.G., Qi, Y., Xiong, Y. and Jin, Z. B., 2022. Fundus tessellated density of pathologic myopia. The Asia-Pacific Journal of Ophthalmology, pp.10-1097.
12. Shao, L., Zhang, Q. L., Long, T. F., Dong, L., Zhang, C., DaZhou, W., Wang, Y.X. and Wei, W.B., 2021. Quantitative assessment of fundus tessellated density and associated factors in fundus images using artificial intelligence. Translational vision science and technology, 10(9), pp. 23-23.
13. Li, R., Guo, X., Zhang, X., Lu, X., Wu, Q., Tian, Q., Guo, B., Xu, J., Tang, G., Feng, J. and Zhao, L., 2023. Application of artificial intelligence to quantitative assessment of fundus tessellated density in young adults with different refractions. Ophthalmic Research, 66(1), pp.710-720.
14. Lyu, X., Li, H., Zhen, Y., Ji, X. and Zhang, S., 2017, July. Deep tessellated retinal image detection using Convolutional Neural Networks. In 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 676-680). IEEE.
15. Huang, D., Qian, Y., Yan, Q., Ling, S., Dong, Z., Ke, X., Tong, H., Long, T., Li, R., Liu, H. and Zhu, H., 2023. Prevalence of fundus tessellation and its screening based on artificial intelligence in Chinese children: The Nanjing Eye Study. Ophthalmology and Therapy, 12(5), pp.2671-2685.
16. Ju, L., Wang, X., Zhao, X., Bonnington, P., Drummond, T. and Ge, Z., 2021. Leveraging regular fundus images for training UWF fundus diagnosis models via adversarial learning and pseudo-labeling. IEEE Transactions on Medical Imaging, 40(10), pp.2911-2925.
17. K. Anvesh and B. M. Reshmi,” Accurate Classification of Tessellated Retinal Disease with Deep Learning Model,” 2024 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, 2024, pp. 1-6, doi: 10.1109/ESCI59607.2024.10497221.

ISSN: 2582-8118 *Volume 4, Issue 5; Sept 2024*

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYSIS IN SCIENCE AND ENGINEERING

Web: https://www.iarj.in/index.php/ijrase/index

73. An Extensive Analysis of the Research Gaps in Crop Yield Improvement and Precision Agriculture Using Automated Machine Learning

Rajkumar C. Shikkeri

Department of Computer Science & Engineering
TKIET, Warananagar, Maharashtra, India.

Dr. S. R. Patil

Department of Information Science & Engineering
BEC, Bagalkot, VTU, Belagavi, India.

ABSTRACT:

Machine Learning has emerged as a potential tool for solving the problems in areas like Agriculture, Healthcare, Finance, Retail, Simulation, Research, Experimentation and E-commerce, Transportation, Natural Language Processing (NLP) and many more. Integrating Machine Learning approaches in the domain of Precision Agriculture (PA) can become an innovative farming management.

This approach involves leveraging various technologies such as sensors, satellite imagery, and machine learning algorithms to collect, analyze, and interpret data for better decision-making in agriculture. The goal of PA is to optimize field-level management regarding crop farming, enhancing productivity, reducing waste, and ensuring sustainability. Agriculture being the backbone of the Indian Economy from decades and even today. Entire economy is sustained by agriculture, which is the mainstay of the villages. Even the same with nations like China, the United States and few others. Exploring the parameters in Precision agriculture the growth, yield quality and quantity can be ensured well.

KEYWORDS:

Artificial Intelligence, Machine Learning, Precision Agriculture, Crop Disease detection.

I Introduction:

The focus here is analysis, prediction and precise projection based on the Weather patterns, Soil fertility, Crop suitability and sustainability. Timely forecast Plan and actions to the farmers during the Sowing, Growing and Harvesting cycles will help in revolutionizing farming activities.

Real-time monitoring of crops using machine learning holds great promise for revolutionizing agriculture by improving efficiency, sustainability, and overall crop management. As technology continues to advance, addressing challenges and refining these systems will be crucial for widespread adoption and success in the agricultural sector.

The Novelty in Modern Agriculture Includes Various Advanced Techniques:

1. Crop Yield Prediction and Soil Health Monitoring:

- Using data analytics and machine learning to predict crop yields.
- Monitoring soil health through sensors and IoT devices to ensure optimal growing conditions.

2. Precision Agriculture and Automated Farming:

- Implementing GPS technology and automated machinery to increase farming efficiency.
- Utilizing drones for field monitoring, spraying, and planting.

3. Pest and Disease Detection Using Image Analysis:

- Employing image analysis and computer vision to detect pests and diseases in crops.
- Using AI to analyse images from drones or field cameras to identify and mitigate threats early.

These advancements aim to increase efficiency, reduce costs, and enhance sustainability in farming practices. Critical evaluation on crop Precision Agriculture to disease detection will provide insights into the various approaches, methodologies, and advancements in this field.

1. "Cloud-Based Real-Time Monitoring of Crop Growth using IoT and Machine Learning" (2019) by Gupta et al.:

This study proposes a cloud-based system for real-time monitoring of crop growth using IoT devices and machine learning algorithms. It discusses data collection, processing, and analysis techniques for optimizing crop management practices.

2. "A Review of Cloud Computing in Agriculture: Current Trends and Future Directions" (2020) by Rahman et al.:

This review paper provides an overview of cloud computing applications in agriculture, including real-time monitoring of crops. It discusses challenges, opportunities, and emerging research directions in leveraging cloud-based technologies for agricultural monitoring.

3. "Real-Time Monitoring of Crop Diseases using Machine Learning and Edge Computing" (2020) by Patel et al.:

The study presents a real-time monitoring system for detecting crop diseases using machine learning models deployed on edge devices. It discusses the integration of edge computing and cloud services for efficient data processing and analysis.

4. "Cloud-Based Precision Agriculture: A Review of Current Trends and Future Directions" (2021) by Li et al.:

This review paper discusses cloud-based precision agriculture systems, including real-time monitoring of crops using machine learning approaches.

It covers data acquisition, processing, and decision support systems for improving agricultural practices.

5. "IoT and Cloud Computing for Crop Monitoring and Decision Support: A Review" (2021) by Sharma et al.:

The review paper provides insights into the integration of IoT and cloud computing for crop monitoring and decision support systems. It discusses machine learning techniques for analysing sensor data and providing actionable insights to farmers.

6. "Cloud-Based Agricultural Monitoring System using Machine Learning and Satellite Imagery" (2022) by Khan et al.:

This study proposes a cloud-based agricultural monitoring system that utilizes machine learning algorithms to analyse satellite imagery for crop monitoring. It discusses the architecture, data processing pipeline, and performance evaluation of the system.

7. "Real-Time Crop Disease Detection using Cloud-Based Deep Learning Models" (2022) by Wu et al.:

The study presents a cloud-based system for real-time crop disease detection using deep learning models. It discusses the training process, model deployment, and integration with cloud services for scalable and efficient disease monitoring.

8. "Cloud-Based Smart Agriculture: A Comprehensive Review" (2022) by Zhang et al.:

This comprehensive review discusses cloud-based smart agriculture systems, including real-time monitoring of crops using machine learning approaches.

It covers data acquisition, processing, and decision support systems for sustainable agricultural practices.

9. "Machine Learning-Based Crop Yield Prediction in Cloud Computing Environments" (2022) by Wang et al.:

The study focuses on machine learning-based crop yield prediction in cloud computing environments. It discusses data pre-processing techniques, feature selection, and model training strategies for accurate yield forecasting.

10. ."Cloud-Based Crop Monitoring and Management System using Reinforcement Learning" (2022) by Liu et al.:

This study proposes a cloud-based crop monitoring and management system that leverages reinforcement learning techniques. It discusses the optimization of agricultural practices through adaptive decision-making and feedback mechanisms.

Gaps in Current Research:

Name of the author	Proposed Technique	Advantages	Limitations
"Cloud-Based Crop Monitoring and Management System using Reinforcement Learning" (2022) by Liu et al.:	Proposed a Deep Q-Network (DQN) model, which combines deep learning with Q-learning to optimize irrigation, fertilization, and pest control strategies	Using reinforcement learning optimizes agricultural practices by providing real-time, data-driven decisions for improved crop yield and resource efficiency	Limitation due to high dependency on stable internet connectivity and significant computational resources.
"Machine Learning-Based Crop Yield Prediction in Cloud Computing Environments" (2022) by Wang et al.:	integration of ensemble learning models with cloud-based data pipelines to enhance prediction accuracy and robustness by combining multiple models' outputs and leveraging large-scale data processing capabilities.	Enables scalable, real-time analysis and forecasting, leveraging vast datasets for accurate predictions	It can be constrained by data privacy concerns and potential latency issues due to reliance on constant internet connectivity and remote server processing.
“Cloud-Based Smart Agriculture: A Comprehensive Review" (2022) by Zhang et al.	Integrated technologies such as IoT, big data analytics, and machine learning algorithms to enable real-time monitoring, predictive analytics, and automated	enhances real-time monitoring and decision-making by integrating IoT, big data, and machine learning for optimized crop management.	face challenges related to data security and privacy concerns due to the centralized nature of cloud storage and the reliance on internet connectivity

Name of the author	Proposed Technique	Advantages	Limitations
	decision-making in agriculture.		
"Real-Time Crop Disease Detection using Cloud-Based Deep Learning Models" (2022) by Wu et al.:	Integrates cloud computing to process and analyze large volumes of data from remote sensing and imagery, enabling rapid and accurate disease detection and diagnosis.	The model enables rapid and accurate real-time crop disease detection by leveraging scalable computing power and extensive datasets for improved diagnostic precision.	encounter challenges related to high data transmission costs and latency issues, as well as reliance on continuous internet connectivity for accessing cloud-based resources.
"IoT and Cloud Computing for Crop Monitoring and Decision Support: A Review" (2021) by Sharma et al.:	The model integrates Internet of Things (IoT) sensors with cloud computing to collect, store, and analyze agricultural data	Provide real-time, scalable, and data-driven insights by integrating IoT sensor data with cloud-based analytics, enhancing precision in crop management and resource optimization.	Reliance on continuous internet connectivity and cloud infrastructure, which can lead to issues with data privacy, security, and potential service disruptions.

The research paper titled "An Extensive Analysis of the Research on Crop Yield Improvement and Precision Agriculture Using Automated Machine Learning" delves into the innovative intersection of agriculture and artificial intelligence (AI).

This analysis aims to interpret and evaluate the methodologies, findings, and implications of using automated machine learning (Auto ML) for enhancing crop yields and implementing precision agriculture.

Background:

The integration of technology into agriculture, particularly through precision farming, seeks to optimize field-level management with respect to crop farming. Automated machine learning represents a leap in this domain, promising to streamline and enhance the process of data analysis and decision-making. The research paper likely addresses the following critical aspects:

1. **Auto ML in Precision Agriculture:** Precision agriculture involves the use of GPS, remote sensing, and IoT to manage crop variability in fields. Auto ML automates the process of applying machine learning to this data, making advanced predictive models more accessible to farmers and agronomists.
2. **Crop Yield Improvement:** Crop yield, a pivotal metric in agriculture, is influenced by numerous factors including soil quality, weather conditions, pest infestations, and crop management practices. The paper likely examines how Auto ML can predict and enhance these factors to improve yields.

Methodologies:

The paper likely employs a range of methodologies to examine the efficacy of Auto ML in precision agriculture:

1. **Data Collection:** Gathering extensive datasets from various agricultural sources including satellite imagery, soil sensors, weather stations, and historical crop performance data.
2. **Model Training and Evaluation:** Utilizing Auto ML tools to train multiple machine learning models on the collected data, followed by rigorous evaluation to select the best-performing models.
3. **Field Trials:** Implementing the selected models in real-world agricultural settings to validate their performance and impact on crop yields.
4. **Comparative Analysis:** Comparing the performance of Auto ML-generated models with traditional agricultural practices and other AI-based approaches.

Findings:

The research likely presents significant findings demonstrating the potential of Auto ML in transforming precision agriculture:

1. **Efficiency:** Auto ML can significantly reduce the time and expertise required to develop predictive models, making advanced analytics accessible to a broader range of users.
2. **Accuracy:** Models generated by Auto ML are likely to show high accuracy in predicting crop yields and identifying key factors affecting them, leading to better-informed decision-making.
3. **Scalability:** Auto ML facilitates the scalability of precision agriculture practices across different regions and crop types, adapting to local conditions and requirements.
4. **Cost-Effectiveness:** By optimizing resource use and reducing the need for manual intervention, Auto ML can lower the overall costs associated with precision agriculture.

Implications:

The implications of these findings are profound, suggesting that Auto ML could play a pivotal role in addressing global food security challenges. By enhancing crop yields and optimizing resource use, precision agriculture driven by Auto ML can contribute to more sustainable and resilient agricultural practices.

Challenges and Future Directions:

While the research highlights the potential of Auto ML in agriculture, it also likely addresses several challenges:

1. **Data Quality and Availability:** The effectiveness of Auto ML depends on the availability of high-quality, comprehensive datasets.

2. **Technical Barriers:** Implementing Auto ML requires a certain level of technical infrastructure and expertise, which may not be readily available in all agricultural settings.
3. **Adoption and Integration:** Encouraging the adoption of Auto ML in traditional farming practices involves overcoming resistance to change and ensuring that farmers are adequately trained.

The paper may suggest future research directions, including the development of more robust data collection methods, improving the interpretability of Auto ML models, and creating user-friendly interfaces for farmers.

Conclusion:

The extensive analysis provided underscores the transformative potential of Auto ML in precision agriculture. By automating the complex process of machine learning model development, Auto ML can enhance crop yield predictions and optimize farming practices, paving the way for a more efficient and sustainable agricultural future.

References:

1. Cloud-Based Real-Time Monitoring of Crop Growth using IoT and Machine Learning (2019) by Gupta et al.:
2. A Review of Cloud Computing in Agriculture: Current Trends and Future Directions (2020) by Rahman et al.:
3. Real-Time Monitoring of Crop Diseases using Machine Learning and Edge Computing (2020) by Patel et al.:
4. Cloud-Based Precision Agriculture: A Review of Current Trends and Future Directions (2021) by Li et al.:
5. IoT and Cloud Computing for Crop Monitoring and Decision Support: A Review (2021) by Sharma et al.:
6. Cloud-Based Agricultural Monitoring System using Machine Learning and Satellite Imagery (2022) by Khan et al.:
7. Real-Time Crop Disease Detection using Cloud-Based Deep Learning Models (2022) by Wu et al.:
8. Cloud-Based Smart Agriculture: A Comprehensive Review (2022) by Zhang et al.:
9. Machine Learning-Based Crop Yield Prediction in Cloud Computing Environments (2022) by Wang et al.:
10. Cloud-Based Crop Monitoring and Management System using Reinforcement Learning (2022) by Liu et al.:
11. Bin Abdullah, A., and bin Yusuf, M. (2023). Machine Learning Approaches to Enhance Water Conservation in Agriculture. *Journal of Sustainable Technologies and Infrastructure Planning*, *7*(5), 22-39.
12. Potgieter, A. B., Zhao, Y., Zarco-Tejada, P. J., Chenu, K., Zhang, Y., Porker, K., ... and Chapman, S. (2021). Evolution and application of digital technologies to predict crop type and crop phenology in agriculture. *in silico Plants*, *3*(1), diab017.
13. Oikonomi dis, A., Catal, C., and Kassahun, A. (2022). Hybrid deep learning-based models for crop yield prediction. *Applied artificial intelligence*, *36*(1), 2031822.

14. Chavan, R. S., Srivastava, G., and Pradhan, N. (2023, April). Advanced plant health monitoring and forecasting system using edge-fog-cloud computing and lstm networks. In *Proceedings of 3rd International Conference on Artificial Intelligence: Advances and Applications: ICAIAA 2022* (pp. 335-344). Singapore: Springer Nature Singapore.
15. Cravero, A., Pardo, S., Galeas, P., López Fenner, J., and Caniupán, M. (2022). Data type and data sources for agricultural big data and machine learning. *Sustainability*, *14*(23), 16131.
16. Shumilo, L., Drozd, S., Kussul, N., Shelton, A., and Sylantyev, S. (2021, April). Mathematical Models and Informational Technologies of Crop Yield Forecasting in Cloud Environment. In *International Scientific and Technical Conference-Modern Challenges in Telecommunications* (pp. 143-164). Cham: Springer International Publishing.
17. Yao, J., Wu, J., Xiao, C., Zhang, Z., and Li, J. (2022). The classification method study of crops remote sensing with deep learning, machine learning, and Google Earth engine. *Remote Sensing*, *14*(12), 2758.
18. Karantzalos, K., Karmas, A., and Tzotsos, A. (2017). Monitoring crop growth and key agronomic parameters through multi temporal observations and time series analysis from remote sensing big data. *Advances in Animal Biosciences*, *8*(2), 394-399.
19. Atef, M., Khattab, A., Agamy, E. A., and Khairy, M. M. (2021, August). Deep Learning Based Time-Series Forecasting Framework for Olive Precision Farming. In *2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)* (pp. 1062-1065). IEEE.

ISSN: 2582-8118 *Volume 4, Issue 5; Sept 2024*

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYSIS IN SCIENCE AND ENGINEERING

Web: https://www.iarj.in/index.php/ijrase/index

74. Brain MRI based detection of Alzheimer's Disease – A Review

Usha Aniruddha Jogalekar
Research Scholar, Computer Science & Engineering,
Basaveshwar Engineering College, Bagalkot.
Bagalkot.

Vilas Naik
Associate Professor, Computer Science & Engineering,
Basaveshwar Engineering College, Bagalkot, India.

Abstract:

One of the most serious neurological disorders in elderly people is the Alzheimer's disease. The existing research articles are summarized in this study. Many researchers are working on the detection and classification of Alzheimer's Disease using the neuro-images that reflect the changes in the brain. The scope of the study is limited to MRI images in ADNI dataset.

Keywords:

MRI (Magnetic Resonance Imaging), ADNI, CNN (Convolutional Neural Network), RNN (Recurrent Neural Network)

I. Introduction:

Dementia, generally known as a disease of memory loss by common people, is an overall term for a group of symptoms including memory loss, difficulty in problem-solving and thinking abilities, being confused about time and place, mood changes, struggling to follow a conversation, etc. Alzheimer's Disease is a type of brain disease that is considered the major cause of Dementia. One of the major causes for Dementia is Alzheimer's Disease. Table 1 gives the list of diseases that cause Dementia [1]. These diseases are the results of periodic changes in the brain. A person may have one or more of these diseases, and the cause is considered a "mixed pathology". All the diseases share common symptoms, because of which there is a high chance of the disease getting wrongly interpreted and treated. Thus, it is highly needed to diagnose the correct disease and treat accordingly. Alzheimer's Disease generally affects elderly people.

The increase in the number of elderly people suffering from Alzheimer's Disease in India has become the cause of alarm. Though there is no cure for Alzheimer's Disease till date, its early detection benefits the patient in treating the disease effectively and extend the process of deterioration. The treatment modalities include medicinal, psychological and care giving aspects

II. The Human Brain:

One of the most complex biological objects in the universe is the human brain, because of which the human race is considered to be the most superior and special to the other living beings. The working of the human brain is not completely understood, and it may remain as a challenge to the scientists. But they are able to know about the structure of the brain and its behavior for an extent. It comprises 100 billion neurons approximately that are linked in networks responsible for a number of cognitive functions like memory, intelligence, creativity, emotion, and consciousness. The brain of an average adult weighs around 1kg. It contains 60% of fat and the remaining 40% is a combination of water, protein, carbohydrates and salts. The brain is formed by blood vessels, nerves, neurons and glial cells.

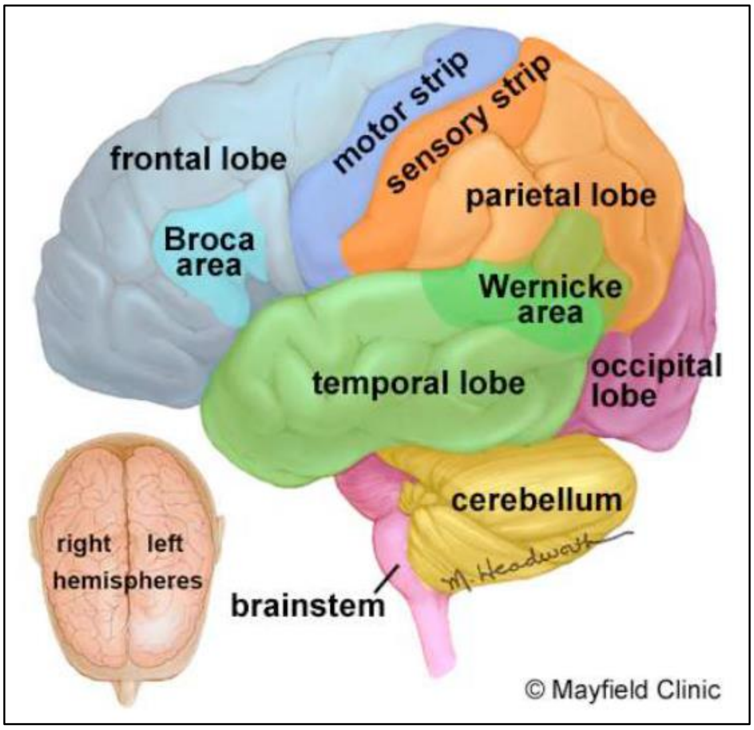


Figure 10.1: Different parts of the human brain

The three major parts of the brain are cerebrum, cerebellum, and brainstem (Fig. 1)[2].The cerebrum is the largest part of the brain.It is responsible for the functions like interpreting touch, vision, hearing, speech, reasoning, emotions, learning, and fine control of movement. It is divided intothe right and left hemispheres

Table 10.1: List of Diseases causing Dementia

Disease	Symptoms	Contribution towards Dementia
Alzheimer's Disease	Difficulty in remembering: Like recent conversations, names, events Communication problems, Poor judgment	60% to 80%
Cerebrovascular Disease	Slowed thoughts, Slow gait and poor balance Impaired ability to make decisions, plan or organize	5% to 10%
Frontotemporal Degeneration	Changes in personality and behavior, Difficulty in comprehending language	About 3%
Hippocampal Sclerosis	Memory loss	3% to 13%
Lewy body Disease	Sleep disturbances, Well-formed visual hallucinations and visuospatial impairment, Problems with motor function, Memory loss	About 5%
Parkinson's Disease	Slow movements, rigidity, tremor, changes in gait, Feelings of distraction, visuospatial impairment	3.6%

Table 10.2: Types of Memory

Memory type	Responsible part	Details
Short-term memory	prefrontal cortex	The information is stored for about one minute and the capacity is limited to about 7 items
Long-term memory	hippocampus of the temporal lobe	The capacity is of unlimited content and duration. It includes personal memories and also facts and figures.
Skill memory	Cerebellum	This includes the learned memories like playing a musical instrument, riding a bike, swimming etc.

The cerebellum which is located under the cerebrum coordinates muscle movements, maintain posture, and balance. The lower part of the brain is called the brainstem that consists of the midbrain, pons, and medulla. The brainstem connects brain and the spinal cord. It acts as a relay centerthatconnects the cerebrum and cerebellum to the spinal cord. The brainstem performs many functions like controlling breathing, heart rate, body temperature. It also controls the wake and sleep cycles, digestion, sneezing, coughing, vomiting, and swallowing.

The sensations and motor skills of the body are controlled by a set of 12 nerves called cranial nerves. These nerves connect the brain and other parts of the body and pass electrical signals between them. There are twelve cranial nerves out of which ten originate in the brainstem and two from the cerebrum.

Memory can be defined as a complex process that includes three phases: encoding, storing, and recalling. Encoding is to create information, storing is to store the information, and recalling is to recall the information whenever it is needed. There are three types of memory: Short-term, Long-term and Skill memory. Different parts of the brain are responsible for eachtype of memory. Table 10.2 gives the details.

III. Biomarkers Used to Diagnose the Alzheimer's Disease:

Biomarkers are defined as the measurable biological changes that indicate the presence or absence of a particular disease. They can also determine the risk of developing a disease. As most of the diseases related to brain show common symptoms, it is required to perform investigations to diagnose the disease. Mario Riverol and Oscar L. Lopez reviewed the biomarkers in Alzheimer's Disease in detail [3]. The neuropathological hallmark of Alzheimer's Disease is the presence of cortical intracellular neurofibrillary tangles (NFT) that contain tau protein and extracellular β-amyloid (Aβ) plaques that are also called Senile Neuritic plaques (SNP) which leads to synapse dysfunction, neuron cell loss and subsequent brain atrophy.

The ideal biomarker for Alzheimer's Disease can be

i. Fundamental Patho-physiology tests
ii. Efficacious at prodromal and even pre-clinical stages
iii. An indicator of disease's severity
iv. A marker of treatment effectiveness

There are many biomarkers described to detect Alzheimer's Disease using different approaches. Most frequently used are the pathological investigations of cerebrospinal fluid

(CSF) and analysis of the brain images generated by neuroimaging techniques. The CSF reflects the brain neurochemistry andthe biochemical values in plasma or serum are affected by many non-neurological factors.

IV. Neuroimaging Techniques:

Structural neuroimaging techniques are useful in the clinical assessment of patients with dementia, so that the structural lesions, like the brain tumors, normal pressure hydrocephalus can be ruled out. It is now proved that the accumulation of the protein's beta-amyloid and tau, inside and outside neurons which in turn causes death of neurons and hence damage to brain tissue and finally causes Alzheimer's Disease. Thus, neuroimaging would be a better choice to decide the presence of the disease. Among the other neuroimaging types available, MRI is highly sensitive to detect the brain atrophy caused by the process of neurodegeneration. A brain MRI scan is a test that produces very clear images of the structures of the brain. It uses radio waves with high frequency (around 64MHz) along with the powerful magnets to create a detailed view of the brain. The images show whether any areas of the brain have atrophied or shrunk. To show how a person's brain changes over time, the scans are repeated. Evidence of shrinkage may support a diagnosis of Alzheimer's Disease [4]. Other neuroimaging techniques to detect Alzheimer's Disease are the PET and the SPECT scans. PET scan or Positron Emission Tomography scan can detect the parts of the brain engaged in an abnormal brain activity. PET scans using tracers that bind to amyloid plaques or tau tangles can help visualize the presence of these key Alzheimer's pathology markers. The most common PET scans are:Amyloid PET that detects amyloid plaques in the brain, and Tau PET that Detects tau tangles in the brain.The SPECT scan or the Single Photon Emission Computed Tomography scan assesses the blood flow in the brain and can help to identify the areas of the brain affected by Alzheimer's disease.

V. Stages of Alzheimer's Disease:

Any disease can be classified as non-degenerative and degenerative. The non-degenerative diseases are medical conditions that typically do not worsen over time and sometimes improve with appropriate treatment. Some examples of non-degenerative disease are Hypo/HyperThyroidism, Type 1/2 Diabetes, Asthma, Hypertension etc. In a degenerative disease, the structure or the function of the affected tissues or organs changes for the worse over time. The degenerative diseases are categorized as: Cardiovascular, Neoplastic and Neurodegenerative.

Cardiovascular diseases are the group of disorders of the heart and blood vessels which together can be considered as the blood circulatory system in the body. [5]

Alzheimer's Disease is one of the neurodegenerative diseases that grows gradually. There are seven stages of Alzheimer's Diseasethrough which the patient moves [6] [7]. They are as follows:

Stage 1: Normal Outward Behavior:

In this early phase, the person does not have any symptoms that he/she can be thought to have Alzheimer's Disease, butthe brain changes can be identified by a PET.

Stage 2: Very Mild Changes:

In this stage also there may not be any major changes in the patient's behavior, but there may be small differences. This could include forgetting a word or misplacing objects. Such changes are generally neglected and not considered as a serious symptom of Alzheimer's Disease, as they are the normal changes due to aging.

Stage 3: Mild Decline:

Now, the changes in the patient's thinking and reasoning can be observed, such as: Forgetting something which is just read, Asking the same question again and again, having trouble in making plans or organizing, Difficult in remembering the names of people when they meet.

Stage 4: Moderate Decline:

In this stage, the patient may: Forget details about himself or herself, have trouble putting the right date and amount on a cheque, forget what month or season it is, have trouble cooking meals or even ordering from a menu

Stage 5: Moderately Severe Decline:

At this stage, the patient may: lose track of where he/she is, fail to get what the time it is, have trouble remembering his/her address and phone number, be confused about what kind of clothes to wear for the day or season.

Stage 6: Severe Decline:

As the disease progresses, the patient may: Recognize faces but forget names, mistake a person for someone else, Be in Delusions, i.e., he/she believes that is not true inpresent time. For example, he/she may think of going to work even though there is no longer a job. It is the last stage when many basic abilities, such as eating, walking, and sitting up fade. Sometimes the patient may not be able to express that he/she is hungry or thirsty.

VI. Adni Dataset:

There are different image datasets available for study. ADNI, OASIS are the most frequently used among them. In this study, ADNI dataset is considered, MRI images in particular. ADNI stands for Alzheimer's Disease Neuroimaging Initiative [8]. The creation of this dataset began in 2004 under the leadership of Dr. Michael W. Weiner. It is funded as a private-public partnership supported by the institutions like the National Institute on Aging, the Foundation for the National Institutes of Health, the Alzheimer's Association. ADNI dataset is a comprehensive and widely used collection of longitudinal clinical, imaging, genetic, and other biomarker data. It encompasses various data types, including structural, functional, and molecular brain imaging, biofluid biomarkers, cognitive assessments, genetic data, and demographic information. MRI data is one of the components of the comprehensive data set collected in ADNI participants. The types of MRI data available to ADNI data users through the LONI IDA include: MRI images in DICOM format, Numerical summary data in tabular .csv format.The tables contain detailed acquisition and quality control information for individual images. It is a multisite longitudinal study, wherein researchers at 63 sites in the US and Canada. They track the progression of Alzheimer's Disease in the human brain with clinical, imaging, genetic and biospecimen biomarkers through the process of normal aging, early mild cognitive impairment (EMCI), and late mild cognitive impairment (LMCI) to dementia or Alzheimer's Disease. The participants undergo a series of initial tests that are repeated at regular intervals over subsequent years. The clinical evaluations include neuropsychological tests, genetic testing, lumbar puncture, and MRI and PET scans.

VII. Literature Review:

The generalprocess of the detection of Alzheimer'sDisease is represented as in Fig. 2. The different phases are explained below:

i. Data collection:

As the process of MRI acquisition at ADNI is a continuous process, the number of images keeps on growing. The researchers can choose different number and type of images for their work. For example, T1-weighted structural MRIs of 379 persons (197 male and 182 female) with 3 to 15 scans per person are selected for study in [9]. Out of these 146 persons are NC (Normal Control), 138 persons are MCI (Mild Cognitive Impairment) and 95 persons are diagnosed with Alzheimer's Disease. In [10] a total of 789 images are used, out of which 509 are used for the training and 278 are for the testing. Among the training images 137 are with Alzheimer's Disease, 76 are MCI, 134 are MCInc(MCI patients who will not convert to Alzheimer's Disease) and 162 are NC.

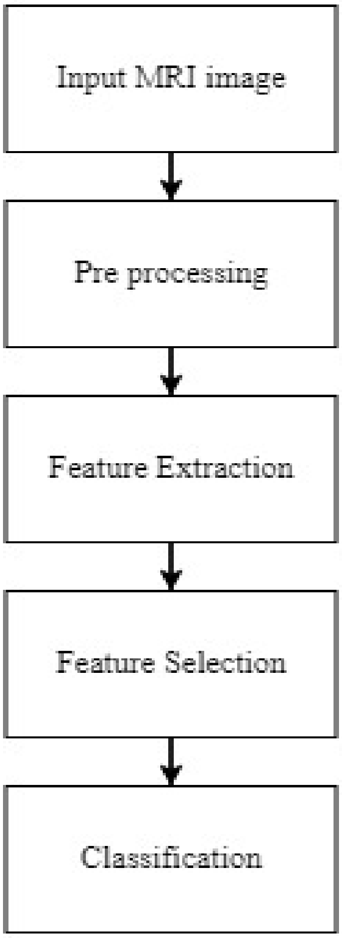


Figure 10.2: Process of Alzheimer's Disease Detection

ii. Image Preprocessing:

The MRIs in the dataset are in the NIfTI format. i.e. 3D Neuroimaging Informatics Technology Initiativeformat. The files have a .nii extension. Initially, Raw MR images are converted to 1-channel images of different sizes followed by resizing and cropping of an image in order to discard the white regions. This enhances the quality of the extracted image. The preprocessing pipeline includes translation, scaling, rotation, skull stripping, image smoothening etc. Several data augmentations like deformation, scaling, cropping, illumination, zoom-in, zoom-out, horizontal and vertical flipping, rotations, random noise addition can be performed so as to reduce the possibility of overfitting during the training [9][11][12].

Image Registration is used to standardize MRI patterns. It is a type of spatial adjustment of MRI scans to a reference anatomical space.Montreal Neurological Institute (MNI) space, acts as an indicator to assess the ability of a brain region in which the points were located to classifyAlzheimer's Disease. The MNI space defines the boundaries around the brain, expressed in milimeter, from a set origin. The brain regions with most intersection points are considered as the most contributing points to the early diagnosis of Alzheimer's Disease [12]. Intensity normalization is also done to map the voxel/pixel intensities of all MRI scans to an MNI space. In [13], Age correction is performed which is necessary to remove age related effects. It is estimated by fitting a linear regression model.

iii. Feature Extraction and Selection:

Feature extraction is a very important step and plays a major role in any image classification or image application. It is a technique that performs dimensionality reduction which proficiently characterizes the fascinating portions of input images as the feature vectors [14]. It helps the classification model for better training, reducingtime complexity, and producing better accuracy. The most useful features in medical images are colors, shapes, textures, and shape-based features. In the detection ofAlzheimer's Disease the features like measurements of hippocampus, grey and white matter volumes, cortical thickness etc. are extracted from MRI. It may happen that sometimes, the extracted features are of less relevance and hence force a model to learn falsely, and the accuracy of the final classification may be affected. Feature selection is the procedure to choose the most relevant features by eliminating the unwanted features by following a particular classification paradigm [14]. The commonly used feature extraction and selection approaches in medical image processing arePCA-Principal Components Analysis, ICA-Independent Component Analysis, LDA-Linear Discriminant Analysis, LLE-Locally Linear Embedding etc.

iv. Classification of MRI images:

A classifier is an algorithm that automatically categorizes data into one or more classes. In this study the classifier can categorize the images into AD or non-AD. There are many classifiers that are used in different research works, while this study concentrates on the research works that involve one or more deep learning algorithms like CNN, SVM, RNN and genetic algorithm approaches.

An adaptive interpretable ensemble model is proposed in [10] that combines 3DCNN and GA, to achieve two goals: first- more accurate disease classification and second- the identification of discriminativebrain regions that could be the biomarkers for early detection in Alzheimer's Disease progression.

An optimized 3D CNN with the whole brain image as input is proposed in [15]. The two networks, named ADNet and its domain adaptation ADNet-DA. In this method no domain specific knowledge from Alzheimer's Disease was used, so it can be applied to other disorders too. While [15] discusses 3DCNN, 2D CNN is also used as a classifier in [16] along with RNN so as to cope up with 3D MRI volumes.

VIII. Conclusion:

The detection of Alzheimer's Disease with the help of by considering MR Images has been discussed in this study. The research works that involved ADNI dataset are considered and reviewed. Among different methodologies adopted, The CNN classifier gives encouraging results. Another biomarker called MagnetoEncephaloGraphy (MEG) can also be used to detect Alzheimer's Disease. It is a low-cost non-invasive test that measures magnetic fields produced by brain's electrical currents. The advantages of MEG over other types of brain images are: it has excellent resolution that captures subtle brain alterations associated with different brain disorders, it provides direct index of neuronal activity, and it is useful in tracking the brain activity disruption like Aβ and tau deposits. In [17][18] MEG is used not only to detect Alzheimer's Disease but also to predict the progression.

IX. Future Scope

The research works considered in this study use ADNI dataset, but the set of images used are different. Hence it would be a good comparative study, if these methodologies are applied on the same set of images.

References:

1. Alzheimer's Association. 2024 Alzheimer's Disease Facts and Figures. Alzheimer's Dement 2024;20(5). doi: https://www.alz.org/alzheimers-dementia/facts-figures
2. 'Anatomy of the Brain' Mayfield Clinic, 2023, https://mayfieldclinic.com/pe-anatbrain.htm
3. Riverol M, López OL. Biomarkers in Alzheimer's disease. Front Neurol. 2011 Jul 14; 2:46. doi: 10.3389/fneur.2011.00046. PMID: 21808632; PMCID: PMC3139171.
4. 'Biomarkers for Dementia Detection and Research Fact Sheet' [online]. Available: https://order.nia.nih.gov/sites/default/files/2020-07/biomarkers-dementia-research-508.pdf
5. Thiene G, Valente M. Degenerative, non-atherosclerotic cardiovascular disease in the elderly: a clinico-pathological survey. Aging (Milano). 1990 Sep;2(3):231-44. doi: 10.1007/BF03323924. PMID: 2094363.
6. Amanda Gardner, "7 Stages of Alzheimer's Disease," 2021. [Online]. Available: https://www.webmd.com/alzheimers/guide/alzheimers-disease-stages.

7. "The 7 Stages of Alzheimer's Disease," Penn Medicine, [online]. Available:https://www.pennmedicine.org/updates/blogs/neuroscience-blog/2019/november/stages-of-alzheimers.
8. Alzheimer's Disease Neuroimaging Initiative: ADNI. Available at http://adni.loni.usc.edu
9. Razzak I, Naz S, Alinejad-Rokny H, Nguyen TN, Khalifa F. A Cascaded Multiresolution Ensemble Deep Learning Framework for Large Scale Alzheimer's Disease Detection using Brain MRIs. IEEE/ACM Trans Comput Biol Bioinform. 2022 Nov 2; PP. doi: 10.1109/TCBB.2022.3219032. Epub ahead of print. PMID: 36322495.
10. D. Pan *et al*., "Adaptive 3DCNN-Based Interpretable Ensemble Model for Early Diagnosis of Alzheimer's Disease," in *IEEE Transactions on Computational Social Systems*, vol. 11, no. 1, pp. 247-266, Feb. 2024, doi: 10.1109/TCSS.2022.3223999.
11. Silvia Basaia, Federica Agosta, Luca Wagner, Elisa Canu, Giuseppe Magnani, Roberto Santangelo, Massimo Filippi, Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks, Neuro Image: Clinical, Volume 21, 2019, 101645, ISSN 2213-1582, https://doi.org/10.1016/j.nicl.2018.101645.
12. Pan Dan, Zeng An, Jia Longfei , Huang Yin , Frizzell Tory , Song Xiaowei, Early Detection of Alzheimer's Disease Using Magnetic Resonance Imaging: A Novel Approach Combining Convolutional Neural Networks and Ensemble Learning, Frontiers in Neuroscience, vol. 14, 2020, DOI:10.3389/fnins.2020.00259
13. Lin W, Tong T, Gao Q, Guo D, Du X, Yang Y, Guo G, Xiao M, Du M, Qu X and The Alzheimer's Disease Neuroimaging Initiative (2018) Convolutional Neural Networks-Based MRI Image Analysis for the Alzheimer's Disease Prediction from Mild Cognitive Impairment. Front. Neurosci. 12:777. doi: 10.3389/fnins.2018.00777
14. R. A. Hazarika, A. K. Maji, S. N. Sur, B. S. Paul and D. Kandar, "A Survey on Classification Algorithms of Brain Images in Alzheimer's Disease Based on Feature Extraction Techniques," in *IEEE Access*, vol. 9, pp. 58503-58536, 2021, doi: 10.1109/ACCESS.2021.3072559.
15. Folego Guilherme, Weiler Marina, Casseb Raphael F. , Pires Ramon, Rocha Anderson, Alzheimer's Disease Detection Through Whole-Brain 3D-CNN MRI, Frontiers in Bioengineering and Biotechnology, vol. 8, 2020, https://www.frontiersin.org/journals/bioengineering-and-biotechnology/articles/10.3389/fbioe.2020.534592, DOI:10.3389/fbioe.2020.534592
16. Ebrahimi A, Luo S; Alzheimer's Disease Neuroimaging Initiative. Convolutional neural networks for Alzheimer's disease detection on MRI images. J Med Imaging (Bellingham). 2021 Mar;8(2):024503. doi: 10.1117/1.JMI.8.2.024503. Epub 2021 Apr 29. PMID: 33937437; PMCID: PMC8083897.
17. Mandal Pravat K., Banerjee Anwesha, Tripathi Manjari, Sharma Ankita A Comprehensive Review of Magnetoencephalography (MEG) Studies for Brain Functionality in Healthy Aging and Alzheimer's Disease (AD), Frontiers in Computational Neuroscience VOL. 12, 2018, DOI=10.3389/fncom.2018.00060
18. Xu M, Sanz DL, Garces P, Maestu F, Li Q, Pantazis D. A Graph Gaussian Embedding Method for Predicting Alzheimer's Disease Progression with MEG Brain Networks. IEEE Trans Biomed Eng. 2021 May;68(5):1579-1588. doi: 10.1109/TBME.2021.3049199. Epub 2021 Apr 21. PMID: 33400645; PMCID: PMC8162933.

ISSN: 2582-8118 ***Volume 4, Issue 5; Sept 2024***

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYSIS IN SCIENCE AND ENGINEERING

Web: https://www.iarj.in/index.php/ijrase/index

75. Brain Tumor Detection using Machine Learning and Deep Learning Models: A Comprehensive Survey

Sudha V. Salake

Department of Computer Science and Engineering,
KLSGogte Institute of Technology, Belagavi, India.

Sadhana P. Bangarashetti

Depratment of Information Science and Engineering,
Basaveshwara College of Engineering, Bagalkote, India.

Abstract:

Cells develop quickly and uncontrollably, which can lead to brain tumors. It could be fatal if treatment is not received in the early stages. Accurate segmentation and classification remain a difficult issue in this domain, despite numerous noteworthy attempts and encouraging results. The differences in tumor size, form, and location present a significant obstacle to the detection of brain tumors. The purpose of this survey is to provide researchers with a thorough literature review on brain tumor detection using magnetic resonance imaging. This assessment addressed deep learning, transfer learning, and quantum machine learning for the analysis of brain tumors, as well as the anatomy of brain tumors, publically available datasets, augmentation approaches, segmentation, feature extraction, and classification. Lastly, this review offers all pertinent data regarding the identification of brain tumors with their advantages and limitations.

Keywords:

Brain imaging features · Segmentation · Feature extraction · MRI

11. Introduction:

The body as a whole receives sensory information and its accompanying actions from the central nervous system [1-3]. The spinal cord and brain both contribute to this dissemination. The three primary components of the brain's anatomy [4] are the cerebellum, cerebrum, and brain stem. A typical human brain has a volume of 1130 cm^3 for females and 1260 cm^3 for males, and it weighs between 1.2 and 1.4 K [5]. Problem solving, motor control, and judgment are all aided by the brain's frontal lobe. Body position is controlled by the parietal lobe. The brain's visual processing processes are managed by the occipital lobe, while memory and hearing are governed by the temporal lobe. The grayish substance that makes up the cerebral cortex, the outer layer of the cerebrum, is made up of cortical neurons [6]. Comparatively speaking, the cerebellum is smaller than the cerebrum. In living things with nerve systems, it is in charge of motor control, or the systematic management of voluntary motions. The small lesion zone cannot be detected by ALI, lesion Gnb, or LINDA techniques due to its changing size and stroke territory. Compared to other species, humans have a more developed and well-structured cerebellum [7]. There are three lobes in the cerebellum: the anterior, posterior, and flocculonodular. The anterior and posterior lobes are connected by a structure known as the vermis. The cerebellum is made up of an outer, slightly thinner cortex than the cerebrum, and an inner region of white matter (WM). The coordination of intricate motor movements is aided by the anterior and posterior lobes. The body's equilibrium is preserved by the flocculonodular lobe [4, 8]. As its name suggests, the brain stem is a 7–10 cm long organ that resembles a stem. It supports eye motions and regulation, balance and upkeep, and certain vital functions including breathing. It is composed of cranial and peripheral nerve bundles. The brain stem houses the neuronal pathways that leave the thalamus of the cerebrum and travel to the spinal cord. They then dispersed throughout the entire body. The medulla, pons, and midbrain comprise the brain stem's primary components. Eye movements and other muscular, auditory, and visual processing functions are aided by the midbrain. The pons facilitates senses, respiration, and intra-brain communication.

Brain Tumors:

Brain tumors are classified as aggressive or slow-growing [2, 10–20]. A malignant (aggressive) tumor spreads from an original site to a secondary site, whereas a benign (slow-growing) tumor does not infiltrate the surrounding tissues [16, 17, 21–27]. A brain tumor is classified by the WHO into categories I–IV. While grades III and IV tumors are more aggressive and have a worse prognosis, grades I and II tumors are thought to grow slowly [28]

Grade I: These are slow-growing tumors that do not metastasize quickly. These can be nearly entirely removed surgically and are linked to improved long-term survival rates. One instance of such a tumor is pilocytic astrocytoma grade 1.

Grade II: These tumors can grow to a higher grade by spreading to nearby tissues, although they still grow slowly.

Even after surgery, certain cancers may recur. One example of such a tumor is oligodendroglioma.

Grade III: These cancers can infect nearby tissues and grow more quickly than grade II tumors. For these kinds of tumors, surgery is not enough; chemotherapy or radiation therapy after surgery is advised. Anaplastic astrocytoma is one instance of this type of tumor.

Grade IV: The most aggressive and highly spreadable malignancies. They might even exploit blood vessels to develop quickly. One such kind of tumor is glioblastoma multiform [29]. Ischemic stroke: Ischemic stroke is a severe brain illness that is a leading cause of mortality and disability worldwide.

Brain imaging techniques:

The four main techniques for analysing the structure of brain tumors—PET, CT, MRI and DWI—are frequently employed.

1. Tomography with Positron Emission:

A unique class of radioactive tracers is used in Positron Emission Tomography (PET). PET is used to examine characteristics of metabolic brain tumors, including blood flow, glucose metabolism, lipid synthesis, oxygen consumption, and amino acid metabolism. It still ranks among the most potent metabolic procedures and makes use of fluorodeoxyglucose (FDG), the best nuclear medicine [33].

One common PET tracer used in brain imaging is FDG. However, there are certain drawbacks to FDG-PET pictures, such as the incapacity to distinguish between radiation necrosis and a recurring high-grade (HG) tumor [34]. Furthermore, radioactive tracers can have negative impacts on during a PET scan.

2. The computed tomography:

Compared to images from standard X-rays, computed tomography (CT) images offer more detailed information. Since its premiere, the CT scan has been widely recommended and utilized. According to a study [36], there are 62 million CT scans performed annually in the USA alone, of which 4 million are performed on children. CT scans display the bones, blood arteries, and soft tissues of various body parts. Higher radiation is used than with standard X-rays. The risk of cancer may rise with repeated CT scans due to this radiation. CT radiation doses have been used to quantify the related cancer risks [37]. Because MRI offers a strong contrast between the soft tissues and can evaluate structures that a CT scan is unable to see, it can be useful in evaluating.

3. Imaging with Magnetic Resonance:

In addition to being utilized to fully examine various body areas, magnetic resonance imaging (MRI) can identify brain problems earlier than other imaging modalities [25]. Consequently, tumor segmentation is a difficult task due to complex brain architecture [1-5].

Pre-processing strategies, segmentation techniques [48, 49], feature extraction and reduction techniques, classification techniques, and deep learning approaches are all covered in this review. Lastly, performance metrics and benchmark datasets are shown.

Diffusion weighting imaging magnetic resonance imaging (MRI) sequences are employed to examine stroke lesions according to multiple criteria, including age, location, and extent regions [10]. A computerized approach may be used to accurately diagnose the rate at which the disease progresses during treatment [21]. The cognitive neuroscientists, who often carry out studies linking cerebral abnormalities to cognitive function,

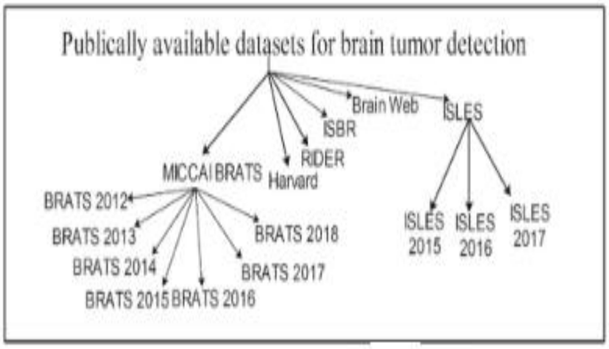


Figure 11.1: Datasets for the diagnosis of brain tumors

Figure 11.1 Datasets for the diagnosis of brain tumors that segment the stroke lesions are essential for analysing the entire affected brain region and aiding in the therapeutic process [12]. Segmenting the stroke lesions is a challenging task, though, because the look of the stroke changes as time passes.

Stroke lesions are detected using MRI sequences such as FLAIR and dif-fusion weighted imaging (DWI). The DWI sequence highlights the infection portion as a hyper intensity in the acute Stoke stage. The mapping magnitude of perfusion is represented by the under-perfusion zone [33]. One could classify the differences between two locations as penumbra tissue. Stroke lesions manifest in different forms and places. Lesions can emerge in a variety of sizes and shapes, are not oriented according to vascular patterns, and may occur at the same time as other lesions. The stroke lesions develop in a complete hemisphere and have radii of a few millimeters. The hemispheres' structures differ, and its intensity could considerably The DWI sequence highlights the infection portion as a hyper intensity in the acute Stoke stage. The mapping magnitude of perfusion is represented by the under-perfusion zone [23]. One could classify the differences between two locations as penumbra tissue. Stroke lesions manifest in different forms and places. Lesions can emerge in a variety of sizes and shapes, are not oriented according to vascular patterns, and may occur at the same time as other lesions. The stroke lesions develop in a complete hemisphere and have radii of a few millimetres. The hemispheres differ in structure, and the affected area may experience substantial variations in intensity. Furthermore, because of the comparable look of the automated stroke segmentation.

Assessment and confirmation: To confirm the robustness of algorithms, experimental findings are assessed on datasets that are accessible to the public in the literature currently under publication.

Datasets that are accessible to the public: The researchers employ a number of publicly accessible datasets to assess the suggested methodologies. This section discusses a few tough and significant datasets. The most difficult MRI datasets are BRATS [15–27]. The BRAT'S Challenge is released every year and features more challenges with a resolution of 1 mm by 3 voxels. The dataset details are provided in both Table 1. and Figure. 11.1

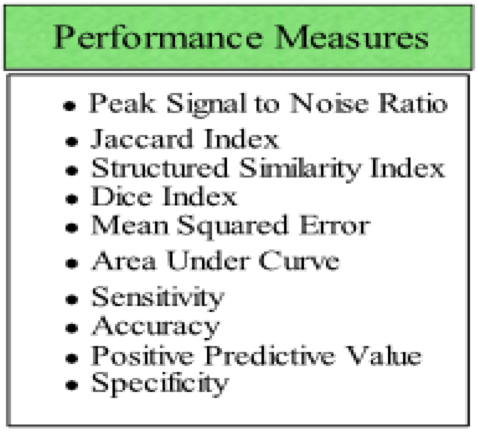


Figure. 11.2: List of performance measures for evaluation of brain tumour.

Metrics of performance:

The efficiency of the method is largely determined by the performance measures. Figure 11.2 provides a set of performance metrics.

Prior to processing:

Pre-processing is an essential step in order to extract the necessary region [11]. Non-brain tissue removal is achieved using the 2D brain extraction algorithm (BEA) [12], FMRIB software library [13], and BSE [14], as illustrated in Figure 11.3. Due to flaws in the radio frequency coil known as intensity inhomogeneity, a major issue with MRI is the bias field [15, 16]. It has been adjusted as Figure 11.4 [27] illustrates. Pre-processing techniques such as pixel-based, fixed, multi-scale, nonlinear [68], and linear are applied in different situations Direct image analysis is frequently hampered by the minute differences between normal and aberrant tissues caused by noise and artifacts. AFINITI is employed in the segmentation of brain tumors Segmentation is carried out by software, which does away with the necessity for manual human contact. There is widespread usage of fully and semi-automated methods lists the outcomes of the segmentation of brain tumors.

The following categories comprise the segmentation techniques.

- Conventional methods.
- Machine learning methods.
- Different in homogeneities related to MRI noise have shading artifacts and partial volume effects

Partial volume effect is the term used when various tissues use the same pixel. MRI-related random noise exhibits a Rician distribution [35]. Various filters, including wavelet, anisotropic diffusion, and adaptive, are described in the literature to improve edges [36].

Table 1 Summary of the publicly available datasets

Datasets	Description	Sequences	Number of slices (images)
BRATS series	BRATS 2012 challengeimage (Patients) dataset (05 LGG, 10 HGG) casesSyntheticdataset (04 LGG,11HGG) cases	T1weighted, T1C weighted, T2 weighted and Flair	240 x 240 x 155
	BRATS 2012 Trainingimage dataset (10 LGG, 20 HGG) casesSyntheticdataset (25 LGG, 25 HGG) cases		
	BRATS 2013 challenge30 Subjects		
	BRATS 2013 Leaderboard, 25 Subjects		
	2014 challenge, 400 Subjects		
	2015 challenge, 274 Subjects		
	2016 challenge, Training cases of BRATS 2015		
	BRATS 2017 challenge, 285 Subjects		
	BRATS 2018 challenge, 191 Subjects		
	BRATS 2019 challenge, 22,087 training and 22,087 testing slices		
	BRATS 2020 challenge, 25,962 training and 25,962 testing slices		
Harvard [58]	65 tumor and 35 non-tumor images	T2 weighted	256 x 256 (100 images)
RIDER[59]	126 Subjects	T1 weighted, T2 weighted, and Flair	256 x 256 126 cases
ISLES 2015	64 Subjects	SISS- ISLES DWI, T1 weighted, T2 weighted, Flair SPES-ISLES CBF, CBV, DWI, T1C weighted, T2 weighted, Tmax, TTP	SISS- ISLES 230 x 230 x 154 (154 slices in each case) SPES-ISLES 230 x 230 x 154 (154 slices in each case)
ISLES 2016	75 Subjects	MTT, rCBV, relative rCBF, Tmax, TTP	192 x 192 x 19 (19 slices in each case)
ISLES 2017[60]	57 Subjects	PWI, ADC, MTT, rCBV, rCBF, Tmax, TTP	192 x 192 x 19 (19 slices in each case)

Partial volume effect is the term used when various tissues [11] use the same pixel [12]. MRI-related random noise [10, 23, 34] exhibits a Rician distribution [25]. Various filters, including wavelet, anisotropic diffusion, and adaptive, are described in the literature to improve edges [16]. the margins [29]. An additional step in the preprocessing stage is to normalize the picture intensity [2, 10]. To do this, the modified curvature diffusion equation (MCDE) [22] is used. In medical imaging, the Wiener filter is employed to improve the local and spatial information [13]. Preprocessing techniques that are frequently used include picture smoothing with median filter [14], image registration [16], sharpening [17], anisotropic diffusion filter [15], and N4ITK [14] for bias field correction.

Traditional Techniques- Conventional Methods:

The following categories also apply to the standard approaches [36]:

- Methods of thresholding.
- Methods of growing regions.
- Methods involving watersheds

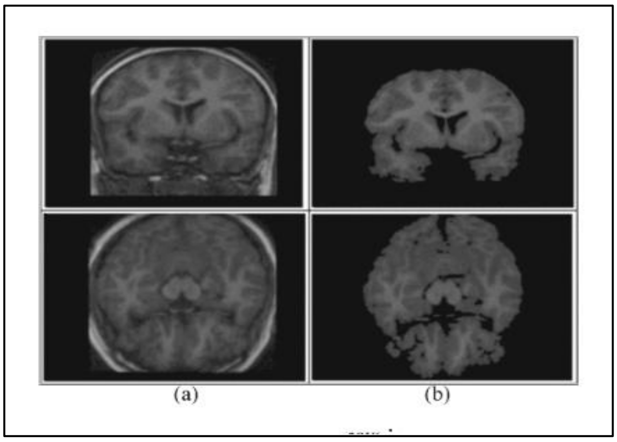


Figure 11.3: Skull removal an input, b skull removed [1]

Division:

Segmentation takes input photos and extracts the needed region. As such, accurately segmenting lesion sites is an increasingly important endeavor [9]. Because the manual segmentation technique is inaccurate [10], fully and semi-automated approaches are used [6]. When compared to manual segmentation, semi-automated tumor region segmentation produces satisfactory results [11, 12]. Three further categories of semi-automated procedures exist: initialization, assessment, and feedback response [13, 14].

Methods of Thresholding:

The thresholding technique is a straightforward yet effective way to divide up the necessary elements [18] and choosing the correct threshold in low-contrast photos can be challenging. Threshold settings are chosen depending on image intensity using histogram analysis [15].

There are two categories for thresholding techniques: local and global. In cases where there is a significant degree of homogenous contrast or intensity between the objects and background, the global thresholding method provides the most effective segmentation alternative. The Gaussian distribution approach [16] can be used to estimate the ideal threshold value. These techniques are applied when a single threshold value does not yield satisfactory segmentation results or when the threshold value cannot be determined from the entire image histogram [17]. Generally speaking, the thresholding approach as demonstrated in Fig. 5, numerous unique regions are segmented within the gray-level pictures in the first stage of segmentation.

Techniques for Region Growing (RG):

Using nearby pixels to analyze picture pixels that constitute discontinuous areas, RG methods combine homogeneousness features based on pre-established simil- itude criteria. The partial volume effect may prevent the region developing from offering greater precision [18, 19]. MRGM is recommended in order to counteract this impact [86, 120]. Additionally, the region expanding via BA techniques is introduced [37].

Watershed Techniques:

Watershed methods are used to assess the image intensity because MR pictures exhibit higher proteaceous fluid intensity [14, 12, 22]. The watershed approach causes over-segmentation [24] as a result of noise [23]. When statistical methods [26, 27] are combined with the watershed transform, accurate segmentation [25] results can be obtained.

Topological watershed [12], image foresting transform (IFT) watershed [29], and marker-based watershed [10] are a few examples of watershed algorithms. There is potential for improvement in brain tumor detection, according to the thorough literature study [31] [2]. Since brain tumors can have a wide range of sizes and shapes, further advancements in tumor segmentation techniques are necessary. The shortcomings of previous approaches have been addressed by segmentation [35–37] and improvement [13–14].

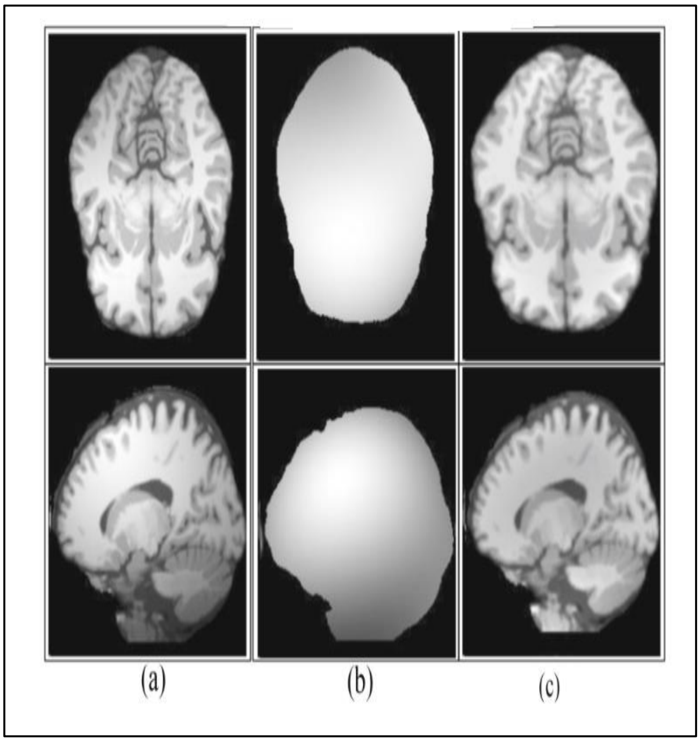


Figure 11.4: Bias field correction a input, b estimated, c corrected

Techniques For Extracting Features:

For classification, the following feature extraction techniques are used: GLCM [15, 141, 12], geometrical features (area, perimeter, and circularity) [15], first-order statistical (FOS), GWT [13, 14], Hu moment invariants (HMI) [15], multifrac-tal features [146], 3D Haralick features [17], LBP [148], GWT [11], HOG [14, 17], texture and shape [2, 14, 14, 15], co-occurrence matrix, gradient, run-length matrix [11], SFTA, curvature features [12, 15], Gabor like multi-scale texton features [14], Gabor wavelet, and statistical features [12, 13].

Techniques For Selecting Features or For Selecting and Reducing Features:

High-dimensional features optimize processing memory and system execution time in machine learning and computer vision applications. Therefore, several feature selection techniques are needed to reduce redundant information in order to distinguish between relevant and non-related features [168]. Optimal feature extraction remains a difficult task [7].

Repetitive features are eliminated using the following methods: single-point heuristic search method, ILS, genetic algorithm (GA) [169], GA+ fuzzy rough set [170], hybrid wrapper-filter [11], minimum redundancy maximum relevance (mRMR) [152], Kullback–Leibler divergence measure [173], iterative sparse representation [14], recursive feature elimination (RFE) [15], CSO–SIFT [16], entropy [11, 17, 18], PCA [19], and LDA [10].

Methods of classification:

Training and testing are carried out on known and unknown samples using the classification procedures, which divide input data into distinct classes [16, 24].

Tumor classification into suitable classes, such as tumor substructure (complete/non-enhanced/enhanced) [193], tumor and non-tumor [26], and benign and malignant tumor [15, 37, 13, 14, 19], is a common application of machine learning.

While FCM [19, 18], hidden Markova random field [19], self-organization map [11], and SSAE [20] are unsupervised techniques, KNN [16], SVM, closest subspace classifier, and representation classifier [13] are supervised techniques.

Current Developments in Medical Imaging to Identify Cancer:

Tumor localization and classification are commonly achieved by the application of deep learning and quantum machine learning techniques [21]. Automatic feature learning aids in the discrimination of complex patterns in these techniques [16, 22].

Techniques For Deep Learning:

A range of cutting-edge deep learning techniques, such as CNN [15, 16], Deep CNN, cascaded CNN [27], 3D-CNN [18], convolutional encoder network, LSTM, CRF [18], U-Net CNN [19], dual-force CNN [220], and WRN-PPNet [21], are employed to learn the data in the medical domain [14].

By using an LSTM model, the categorization issue for brain tumors has been resolved. Using the N4ITK and 5×5 Gaussian filter, MRI images are smoothed before being fed into the four LSTM model. Based on the four hidden units, which are 200, 225, 200, and 225, respectively, the LSTM model is constructed. This model's performance has been evaluated using the SISS-2015 and BRATS (2012–2015 and 2018) series.

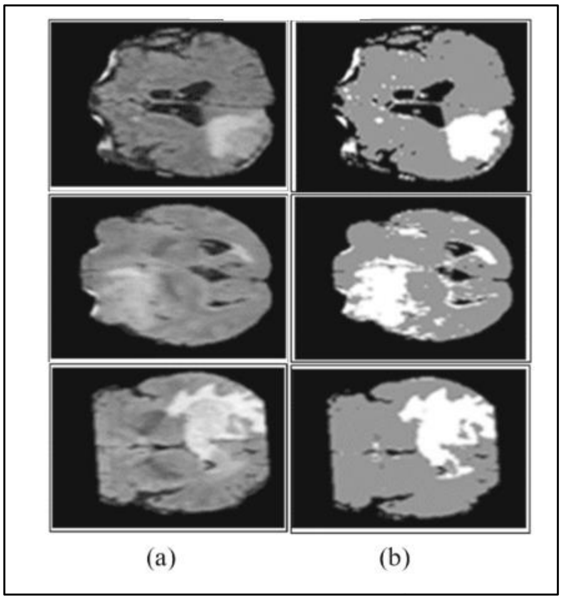


Figure. 11.5: Segmentation using Otsu thresholding a original images, b Otsu thresholding.

Limitations Of Current Deep Learning and Machine Learning Techniques:

There is still opportunity for improvement, according to this survey's analysis of recent research on brain tumor detection. MRI includes noise during image capture, and removing this noise is a difficult task [2, 8-10]. Given that brain tumors contain tentacles and dispersed features [30-34], accurately segmenting them is a challenging task [5]. For improved classification, it's also crucial to choose and extract the best features and the right quantity of training and testing samples. Deep learning models are becoming more popular because they automatically learn features, but they need a lot of processing power and memory. Consequently, the necessity to create a lightweight model that offers high in a shorter amount of time for computation. lists a few machine learning techniques now in use along with some of their drawbacks. The primary obstacles to brain tumor detection are listed below.There is not enough distinction between the stroke and glioma tumors. Its tentacles and dispersed features complicate the processes of classification and segmentation [27]. Since a little volume of tumor can be mistaken for a normal region, detection remains difficult Certain current techniques yield good results when applied to a single, fully tumored region, but not when applied to other regions (enhanced or non-enhanced) [26].

Results and discussion of the research:

A thorough analysis of the most advanced techniques, the following difficulties are discovered:

A brain tumour expands quickly in size. Tumour diagnosis at an early stage is therefore a critical duty. An MRI image caused by variations in the coil's magnetic field.Gliomas have hazy borders, which makes them infiltrative. As a result, segmenting them gets harder [34].

Because stroke lesions might have unclear borders, varying intensities, and complex shapes, segmenting them is a highly difficult task.Another challenging procedure that leads to an incorrect classification of brain tumours is the optimal and best feature extraction and selection.

Conclusion:

Because brain tumors vary in size, form, and structure, it is still exceedingly difficult to detect them accurately. Tumor segmentation techniques have demonstrated a great deal of promise in identifying and interpreting MR images; nevertheless, a great deal of work needs to be done before the tumor region can be precisely segmented and classified. The current body of work presents obstacles and limits when it comes to classifying images as healthy or unhealthy and identifying the substructures of the tumor region.

In summary, this study addresses all pertinent topics including the most recent research, along with its shortcomings and difficulties. The ability to do fresh study quickly and in the right direction would be beneficial to the researchers. Although deep learning techniques have made a substantial contribution, a general technique is still needed. These techniques When training and testing are conducted on similar acquisition features (intensity range and resolution), these approaches perform better. Future studies might be carried out to more precisely identify brain cancers by utilizing actual patient data from any source (such as various image acquisition methods or scanners). Fusing deep and handcrafted characteristics can enhance the classification outcomes. Likewise, low-weight techniques like quantum machine learning are important for increasing accuracy and efficacy, saving radiologists' time, and raising patient survival rates.

References:

1. Park JG, Lee C (2009) Skull stripping based on region growing for magnetic resonance brain images. Neuroimage 47:1394–1407
2. Khan MA, Lali IU, Rehman A, Ishaq M, Sharif M, Saba T et al (2019) Brain tumor detection and classification: A framework of marker-based watershed algorithm and multilevel priority features selection. Microsc Res Tech 82:909–922
3. Raza M, Sharif M, Yasmin M, Masood S, Mohsin S (2012) Brain image representation and rendering: a survey. Res J Appl Sci EngTechnol 4:3274–3282
4. Watson C, Kirkcaldie M, Paxinos G (2010) The brain: an intro-duction to functional neuroanatomy. Academic Press, New York
5. (2015). https://en.wikipedia.org/wiki/Brain_size. Accessed 19 Oct 2019
6. Dubin MW (2013) How the brain works. Wiley, New York
7. Koziol LF, Budding DE, Chidekel D (2012) From movement to thought: executive function, embodied cognition, and the cerebel- lum. Cerebellum 11:505–525
8. Knierim J (1997) Neuroscience Online Chapter 5: Cerebellum.The University of Texas Health Science Center, Houston
9. Nuñez MA, Miranda JCF, de Oliveira E, Rubino PA, Voscoboinik S, Recalde R et al (2019) Brain stem anatomy and surgi-cal approaches. Comprehensive overview of modern surgical approaches to intrinsic brain tumors. Elsevier, Amsterdam, pp 53–105
10. DeAngelis LM (2001) Brain tumors. N Engl J Med 344:114–123

11. Amin J, Sharif M, Raza M, Saba T, Sial R, Shad SA (2020) Braintumor detection: a long short-term memory (LSTM)-based learn-ing model. Neural Comput Appl 32:15965–15973
12. Sajjad S, Hanan Abdullah A, Sharif M, Mohsin S (2014) Psy-chotherapy through video game to target illness related problem-atic behaviors of children with brain tumor. Curr Med Imaging10:62–72
13. Yasmin M, Sharif M, Masood S, Raza M, Mohsin S (2012) Brain image reconstruction: a short survey. World Appl Sci J 19:52–62
14. Amin J, Sharif M, Raza M, Yasmin M (2018) Detection of brain tumor based on features fusion and machine learning. J Ambient Intell Human Comput:1–17
15. Amin J, Sharif M, Yasmin M, Fernandes SL (2020) A distinctive approach in brain tumor detection and classification using MRI.Pattern Recogn Lett 139:118–127
16. Saba T, Mohamed AS, El-Affendi M, Amin J, Sharif M (2020) Brain tumor detection using fusion of hand crafted and deep learn-ing features. Cogn Syst Res 59:221–230
17. Sharif M, Amin J, Nisar MW, Anjum MA, Muhammad N, ShadSA (2020) A unified patch-based method for brain tumor detectionusing features fusion. Cogn Syst Res 59:273–286
18. Sharif M, Tanvir U, Munir EU, Khan MA, Yasmin M (2018) Brain tumor segmentation and classification by improved bino-mial thresholding and multi-features selection. J Ambient IntellHuman Comput:1–20
19. Sharif MI, Li JP, Khan MA, Saleem MA (2020) Active deepneural network features selection for segmentation and recog-nition of brain tumors using MRI images. Pattern Recogn Lett129:181–189
20. Sharif MI, Li JP, Naz J, Rashid I (2020) A comprehensive reviewon multi-organs tumor detection based on machine learning. Pattern Recogn Lett 131:30–37
21. Ohgaki H, Kleihues P (2013) The definition of primary and sec-ondary glioblastoma. Clin Cancer Res 19:764–772
22. Cachia D, Kamiya-Matsuoka C, Mandel JJ, Olar A, CykowskiMD, Armstrong TS et al (2015) Primary and secondary gliosarco-mas: clinical, molecular and survival characteristics. J Neurooncol125:401–410
23. Amin J, Sharif M, Gul N, Yasmin M, Shad SA (2020) Brain tumorclassification based on DWT fusion of MRI sequences using con-volutional neural network. Pattern Recogn Lett 129:115–122
24. Sharif M, Amin J, Raza M, Yasmin M, Satapathy SC (2020) An integrated design of particle swarm optimization (PSO) withfusion of features for detection of brain tumor. Pattern RecognLett 129:150–157
25. Amin J, Sharif M, Anjum MA, Raza M, Bukhari SAC (2020) Con-volutional neural network with batchnormalization for glioma andstroke lesion detection using MRI. Cogn Syst Res 59:304–311
26. Amin J, Sharif M, Raza M, Saba T, Anjum MA (2019) Braintumor detection using statistical and machine learning method.Comput Methods Progr Biomed 177:69–79
27. Amin J, Sharif M, Gul N, Raza M, Anjum MA, Nisar MW et al(2020) Brain tumor detection by using stacked autoencoders indeep learning. J Med Syst 44:32
28. Johnson DR, Guerin JB, Giannini C, Morris JM, Eckel LJ, Kaufmann TJ (2017) 2016 updates to the WHO brain tumor classification system: what the radiologist needs to know. Radio-graphics 37:2164–2180
29. Wright E, Amankwah EK, Winesett SP, Tuite GF, Jallo G, CareyC et al (2019) Incidentally found brain tumors in the pediatricpopulation: a case series and proposed treatment algorithm. J Neu-rooncol 141:355–361

30. Pellegrino MP, Moreira F, Conforto AB (2021) Ischemic stroke. Neurocritical care for neurosurgeons. Springer, New York, pp517–534
31. Garrick R, Rotundo E, Chugh SS, Brevik TA (2021) Acute kidneyinjury in the elderly surgical patient. Emergency general surgeryin geriatrics. Springer, New York, pp 205–227
32. Lehmann ALCF, Alfieri DF, de Araújo MCM, Trevisani ER,Nagao MR, Pesente FS, Gelinski JR, de Freitas LB, Flauzino T, Lehmann MF, Lozovoy MAB (2021) Carotid intima media thickness measurements coupled with stroke severity strongly pre-dict short-term outcome in patients with acute ischemic stroke: amachine learning study. Metab Brain Dis 36:1747–176133. Scott AM (2005) PET imaging in oncology. In: Bailey DL,Townsend DW, Valk PE, Maisey MN (eds) Positron emission tomography. Springer, London, pp 311–325
33. Wong TZ, van der Westhuizen GJ, Coleman RE (2002) Positron emission tomography imaging of brain tumors. Neuroimaging Clin 12:615–626
34. Wong KP, Feng D, Meikle SR, Fulham MJ (2002) Segmentation ofdynamic PET images using cluster analysis. IEEE Trans NuclearSci 49:200–207
35. Brenner DJ, Hall EJ (2007) Computed tomography—an increasing source of radiation exposure. N Engl J Med357:2277–2284
36. Smith-Bindman R, Lipson J, Marcus R, Kim K-P, Mahesh M, Gould R et al (2009) Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. Arch Intern Med 169:2078–2086

ISSN: 2582-8118 *Volume 4, Issue 5; Sept 2024*

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYSIS IN SCIENCE AND ENGINEERING

Web: https://www.iarj.in/index.php/ijrase/index

76. Convolutional Neural Network and Fog Computing Approach for Smart Industry

Basavaraj G. Kumbar
Department of Computer Science and Engineering, BGMIT, Mudhol Research Scholar, BEC Bagalkote, Karnataka, India.

Lokesh B. Bhajantri
Department of Information Science and Engineering Basaveshwar Engineering College, Bagalkote, Karnataka, India.

Abstract:

The smart industry has made significance change in manufacturing field by integrating advanced technologies, rid of the traditional manufacturing process in the industry, which revolutionize the concept of smart industry. Ensuring product quality and operational efficiency is one of the important tasks in the smart industry. In this paper, we propose integration of Convolutional Neural Networks (CNN) in fog computing enables efficient automated inspection system. High accuracy of image recognition can be achieved from CNN and deploying in fog computing helps to reduce latency, enhance data security and efficient bandwidth usage.

Keywords:

CNN, Automated Inspection System, Fog computing

1. Introduction:

Internet of things (IOT) has taken swift in the recent days [1] which makes challenge for the existing network infrastructure, since lot of sensors and actuators are used, which produces huge amount of data contiguously. Industry 4.0 is mainly focusing on smart manufacturing and supply chain management this need to handle digital information efficiently and securely. Fog computing provides cloud computing services to the network edge, allowing distributed processing, memory for storage, and decision-making nearer to data sources. This offers significant advantages, including low latency, reduced network traffic, and improved responsiveness for real-time applications [2]. Fog computing is the bridge between cloud computing and physical devices aims to bring the computation services of the cloud nearer to the network edge, and presents its own set of challenges and issues [3].

A few key challenges and issues related to fog computing are, resource limitations, network connectivity, security and privacy, data management, heterogeneity, scalability, edge intelligence, reduced latency, bandwidth optimization, context aware, energy efficiency, and regulatory compliance

The arrival of Industry 4.0 scripts a significant swing in manufacturing paradigms, characterized by the fusion of cyber-physical systems, the Internet of Things (IoT), and advanced computational methods.

This transformation, often referred to as the Smart Industry, emphasizes automation, data exchange, and real-time processing to enhance productivity and quality across various industrial sectors. One of the most crucial components in this ecosystem is the inspection system, which plays a vital role in maintaining product quality and ensuring operational efficiency.

Traditional inspection systems, predominantly reliant on manual oversight or basic automated methods, often fall short in meeting the high-speed, high-accuracy demands of modern manufacturing processes. These limitations necessitate the adoption of more sophisticated technologies capable of handling complex tasks with precision and speed. Convolutional Neural Networks (CNNs), a class of deep learning algorithms, have emerged as a powerful tool for image recognition and classification tasks, making them an ideal candidate for advanced inspection systems. CNNs can accurately identify defects and anomalies in products, thereby enhancing the reliability and effectiveness of the inspection process.

However, the deployment of CNN-based inspection systems faces challenges, particularly concerning latency and data management. Traditional cloud computing solutions, while providing substantial computational power, often suffer from high latency and bandwidth limitations. These issues can impede real-time processing and decision-making, which are critical for maintaining the continuous flow of manufacturing operations.

The proposed work incorporates the integration of fog computing with CNN-based inspection systems to detect the faults. Computing power of cloud can be brought into Fog network that enables the processing of data closer to the source. This approach significantly reduces latency, enhances data security, and optimizes bandwidth usage. By combining the strengths of CNNs and fog computing, the proposed system aims to deliver a robust, efficient, and scalable solution for real-time inspection in smart industries.

Through this research work, we aim to contribute to the ongoing evolution of Smart Industry by providing a innovative solution that improves inspection processes, thereby supporting the wider goals of efficiency, quality, and smart monitoring in manufacturing.

The remaining work is well thought-out as follows: Section II presents the summery of the various works related to the proposed scheme. Section III delineates the proposed work for efficient inspection system using CNN and fog computing. In section IV, proposed work simulation is depicted, and finally, the conclusion of the proposed scheme is depicted in Section V.

II. Related Work:

A comprehensive summary of current research on defect detection, CNNs, and Fog Computing in the context of smart manufacturing. Focusing key contributions and identify gaps in the literature that the proposed system aims to address.

Wenjin Tao et al. [4] addressed a fog computing framework for various assembly operation recognition, which helps to bring computation power near to the terminals to achieve real time recognition. Author has framed assembly model which consists of visual cameras used for monitoring operational activity. An assembly operation task is formulated, and a dataset is established, which contains 10 sequential operations, Transfer learning is utilized and the developed model isevaluatedonthedatasetandachievesa95%recognitionaccuracy. Yuanbin Wang et al. [5] proposed R-CNN algorithm in the cloud edge computing environment. Author has proposed smart surface inspection system which identifies the defects in the complex images with the help of cloud computing.

Y.C. Liang et al. [6] shown the fog computing over the cloud computing is effectively classify and detect defects in the machining process. The integration of convolutional neural network and fog computing increases the overall efficiency and optimizes the machining process.

Author has shown that, 29% and 16% on average for energy and production enhancement respectively. Hyeon Jong Haet al. [7] discussed about edge and industry IOT systems based on CNN for identifying defect for injection modeling. Data augmentation technique is used to solve data shortage and imbalance problem of small and medium. The experimental result shows the more than 90% and can be applied real time.

Natesha B Vet al. [8] discussed fog based intelligent machine malfunctioning system for smart industry. Author has proposed using sound variation in the machine used to detect the faults which can be computed with the help of fog computing. Also talk about, how industrial control units and fog micro data centers can used effectively to classify sound as normal or abnormal for detecting faults in machines. An integration of machine learning offers better features in the cloud computing. Linear Prediction Coefficients and Mel Frequency Cepstral Coefficients used along with supervised machine learningto extract the machine sounds.

Shih-Yang Linet al. [9] presented smart manufacturing system using deep learning framework with fog. The deep learning approach enhances the detection of faults in the assembly lines. FC-HDLF to analyze the captured images of the sensors and to classify faults or defects in the assembly. Lijun Wuet al. [10] presented efficient fault detection using binary convolutional neural network. The proposed algorithm is tested by using public datasets such as ImageNet and CIFAR-10. This scheme improves the accuracy and classify the defected items.

Various papers have been referred and it is observed that very few papers addressed usage of convolution neural network using fog computing-based approaches for efficient inspection system in the smart industry. Hence, it is proposed context aware resource discovery in fog networks.

III. Proposed Work:

The proposed work is enlightened new automated efficient inspection system that uses convolutional neural networks and fog network which revolutionize smart industry. This can be accomplished by integrating convolutional neural network in the fog layer. The proposed scheme shows improved classification of defective items and avoids the manual operations.

Overview of proposed system architecture is shown in Figure 1, outlinesrevolution of smart industry provides efficient inspection system using convolutional neural networks and fog computing. The architecture mainly consists of three components. In lower layer, visual camera along surface sensor is fixed to monitor production status. Visual cameras are connected to interfaces for image uploading and command transfer.

The processed image data are sent to low layer fog nodes to extract some lower-level features and if possible, which outputs fast inference results for efficiency enhancement and traffic saving. If the fog nodes fail to give accurate judgments or not possible to compute by themselves, they will forward the semi formatted CNN model inference to the remote cloud servers for further processing. The cloud servers have the ability to complete the whole CNN inference,which further complete the classification and regression, and produces the end results.

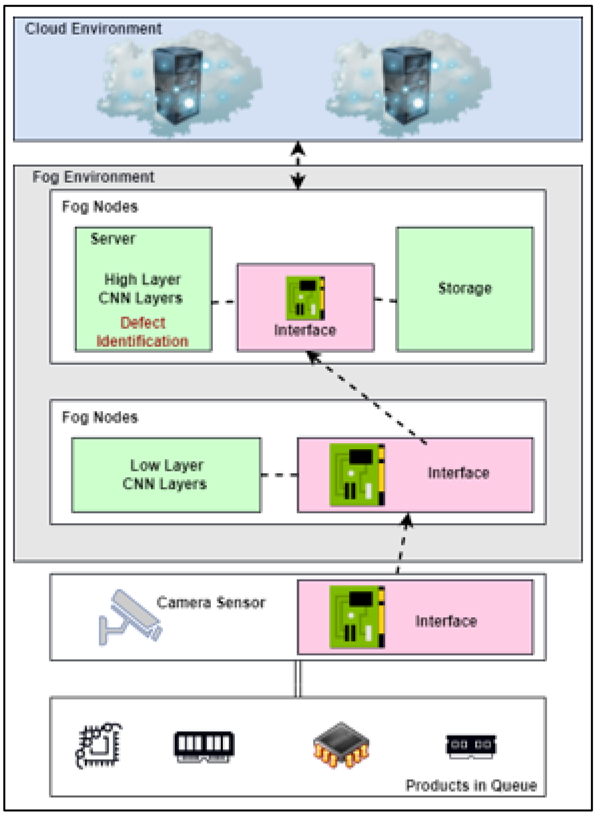


Figure 14.1: Proposed System Architecture

Fog nodes are distributed over fog layer in which CNN layers are embedded. Nodes are categorized as higher level and lower layer nodes. Initially images from the camera will be uploaded to lower layer CNN fog nodes where preprocessing will happen, further results will be proceeded to higher level CNN fog servers where image processing and analysis will be done. Figure 14.2 describes the CNN model used in our system. This contains sequence of convolutional layers and fully connected layers. This is most commonly used model in deep models, decision making part is the important which makes difference with another model. Classifier and regressor are the two parts used to judge faults in the assembly and regressor is used to infer the degree of approximated defects in the assembly production.

Then the problem comes to how to design a loss function to jointly train the classifier and the regressor

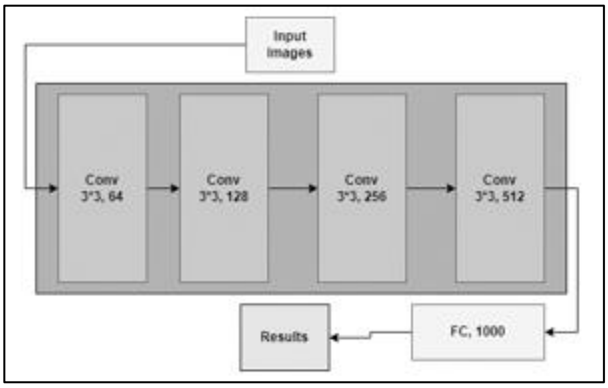


Figure 14.2: CNN Model Structure

A. Proposed scheme framework:

In the proposed framework there are four main components, such as production line, visual cameras, low layer CNN fog nodes and high layer CNN server nodes. The detailed proposed framework is descried as follows

Production line:

Production line is the main component which carry the products in the belt. This is integrated with control system. Control system will control the position of the parts in the line and which will work along the visual camera. Based on the complexity of the part, it controls the speed of the belt to get the clear images of the part with light sensor.

Visual Camera and surface image sensors:

Visual camera is fixed to watch the production parts arrives at the production line. It uses the light sources to ensure the brightness of the image is enough to identify the defects. Surface image sensors are used to capture the surface of the parts for defect detection. Further, captured images are fed to lower layer CNN fog nodes usually fog gateways for the data preprocessing

Lower layer CNN fog nodes:

These are kept near to the visual cameras since high-definition images need to be pre-processed before it reaches to the fog servers for analysis. In this layer, large sized images will be segmented into smaller size and smaller images are overlapped to ensure the edges of the images. Further segmented images will be fed to higher layer CNN servers for analysis.

Higher layer CNN fog servers:

The process of defect detect occurs at this level. Fast CNN uses the segmented images for analysis and identifies the defects in the images based on pre trained results. Identification of faults in the production part will be discussed in the next section along with results.

B. Proposed methodology:

Following steps are used to propose the scheme for identifying defects in the smart industry production

Step1: Set up data collection devices: Visual cameras and surface sensors are deployed near to production line

Step 2: Processing of captured image: Convert image to grayscale and normalize the pixels values and divide the image into smaller size

Step 3: Apply rotation and flipping using device controllers to get fixed position to improve the CNN model

Step 4:

a. Load a pre-trained CNN model for detecting defects
b. Correct the CNN model using Captured images. Split the data set into training, validation and testing

Step 5: Analyse change in the image by using CNN model to identify defects in the items

Step 6: Deploy fog nodes near the data source to ensure minimal latency

Step 7: For advanced analysis, send images to the cloud

IV. Results:

Comparative analysis of this work is done with cloud only solutions and traditional machine vision. Key performance parameters considered are accuracy, latency and bandwidth usage. Python language is used to generate the results and graphs.

A. Performance Parameters:

In this paper, we have evaluated the following parameters and compared the results with existing methods viz. cloud only solution and traditional machine vision:

1. Accuracy:Accuracy defines percentage of products classified successfully which is measured in %. CNN model with fog computing shows better accuracy compared with other two methods.
2. Latency: Time taken to process and identify defected images which is measured in milliseconds(ms). Since the fast CNN integrated in fog servers, reduces the latency compared with other two methods
3. Bandwidth Usage: Amount of data transmitted over the network. Which is measured in Mbps. CNN model is embedded in the fog network it near to the end device and uses less bandwidth for transmitting images.

B. Results and discussion:

Figure 14.3 shows the comparative analysis of CNN based fog computing method with other methods such as cloud only solution and traditional machine vision based with respect to accuracy.

It is observed that CNN based fog computing method achieves more accuracy. Using proposed method, it is achieved 98% accuracy.

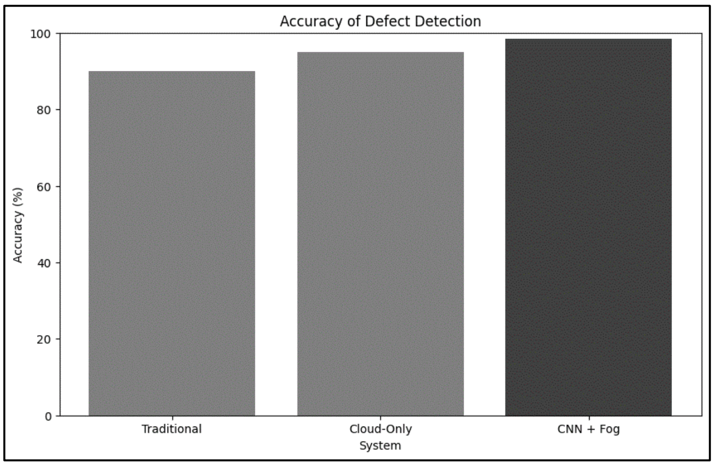


Figure. 14.3: Accuracy Vs Various methods

The comparative analysis of latency is described in the figure 4. It clearly demonstrates that the proposed scheme incurs less time delay compared to other two methods such as cloud only solution and traditional machine vision based with respect to latency.

Generally, numbers of images are increased, more latency is considered for completing the tasks efficiently in the fog computing. Hence, the proposed scheme achieves better efficiency compared with other methods.

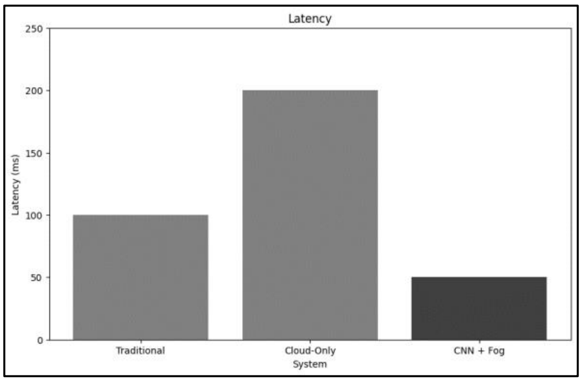


Figure 11.4: Latency Vs Various methods

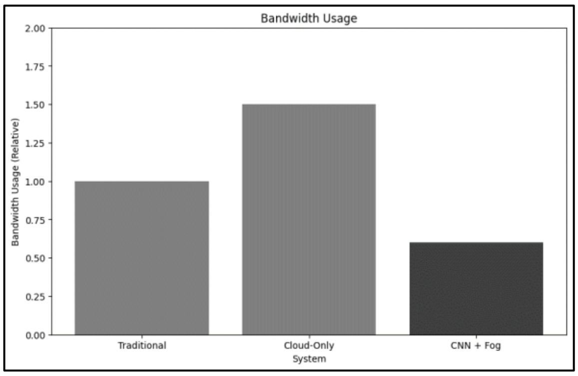


Figure 14.5: Bandwidth usage Vs Various methods

Figure 14.5 shows that, bandwidth usage for identifying defects, the proposed scheme uses less bandwidth compared to other two methods. As the number of images increases, gradually bandwidth usage also increases.

Conclusion:

The proposed approach for smart industry using CNN model and fog computing is efficient. The proposed scheme yields better-quality performance with respect to accuracy, latency, and bandwidth usage. The experimental results show that proposed method is efficient to identify the defects in the smart industry using CNN and Fog computing

References:

1. R. Hassan, F. Qamar, Hasan, M. K., Aman, A. H., M.., Ahmed, A. S., Ahmed,"Internet of Things and Its Applications: A Comprehensive Survey".*Symmetry*, vol.12, no.10, pp.1-29,https://doi.org/10.3390/sym12101674
2. R. Das, M. M., Inuwa, "A Review on Fog Computing: Issues, Characteristics, Challenges, and Potential Applications". *Telematics and Informatics Reports*, vo.*10*, pp.1-20, 2023.https://doi.org/10.1016/j.teler.2023.100049

3. H. Sabireen, V. Neelanarayanan, “A Review on Fog Computing: Architecture, Fog with IoT, Algorithms and Research Challenges”. *ICT Express*, vol.7, no.2, pp.162–176, 2021. https://doi.org/10.1016/j.icte.2021.05.004
4. Wenjin Tao, Md Al-Amin, Haodong chen, Ming C. Leu,Zhaozheng Yin, Ruwen Qin“Realtime assembly operation recognition with fog computing and transfer learning for human centered intelligent manufacturing”. *Procedia Manufacturing 48*, (2020) 926–931
5. Wang, Y., Liu, M., Zheng, P., Yang, H., and Zou, J. (2020). A smart surface inspection system using faster R-CNN in cloud-edge computing environment. *Advanced Engineering Informatics*, *43*, 101037. https://doi.org/10.1016/j.aei.2020.101037
6. Liang, Y., Li, W., Lu, X., and Wang, S. (2019). Fog computing and convolutional neural network enabled prognosis for machining process optimization. *Journal of Manufacturing Systems*, *52*, 32–42. https://doi.org/10.1016/j.jmsy.2019.05.003
7. Ha, H., and Jeong, J. (2021). CNN-Based Defect Inspection for Injection Molding Using Edge Computing and Industrial IoT Systems. *Applied Sciences*, *11*(14), 6378. https://doi.org/10.3390/app11146378
8. N. B. V and R. M. R. Guddeti, "Fog-Based Intelligent Machine Malfunction Monitoring System for Industry 4.0," in *IEEE Transactions on Industrial Informatics*, vol. 17, no. 12, pp. 7923-7932, Dec. 2021, doi: 10.1109/TII.2021.3056076.
9. Lin, S. Y., Du, Y., Ko, P. C., Wu, T. J., Ho, P. T., Sivakumar, V., and Subbareddy, R. (2020). Fog Computing Based Hybrid Deep Learning Framework in effective inspection system for smart manufacturing. *Computer Communications*, *160*, 636–642. https://doi.org/10.1016/j.comcom.2020.05.044
10. Wu, L., Lin, X., Chen, Z., Huang, J., Liu, H., and Yang, Y. (2021). An Efficient Binary Convolutional Neural Network With Numerous Skip Connections for Fog Computing. *IEEE Internet of Things Journal*, *8*(14), 11357–11367. https://doi.org/10.1109/jiot.2021.3052105

ISSN: 2582-8118 *Volume 4, Issue 5; Sept 2024*

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYSIS IN SCIENCE AND ENGINEERING

Web: https://www.iarj.in/index.php/ijrase/index

77. Detection of Phishing Cyber Attack Using ML Techniques

S. V. Saboji, Laxmi Kanthi

Department of Computer Science and Engineering

Basaveshwara Engineering College, Bagalkot, India.

Abstract:

One issue with internet security that targets human susceptibility instead of software flaws is phishing. It explained how to draw in internet users in order to obtain their private data, including usernames and passwords. We outline a sophisticated method for identifying phishing websites. The phishing website system is an add-on feature for an internet browser that alerts the user automatically when it comes across a phishing website. The foundation of the system is machine learning, namely supervised machine learning techniques and algorithms. Because of their strong categorization performance, we have chosen the Random Forest and Decision Tree techniques in this case. Phishing websites are cyber-attacks that aim to steal personal data from internet users, such as passwords, login credentials, and banking information. Attackers deceive users by making the masked webpage appear authentic or reliable in order to obtain their personal information. Machine Learning (ML) based strategies have been presented as a number of remedies to attacks on phishing websites.

Keywords:

Random Forest, Decision Tree, Machine Learning, Phishing Features, Phishing Website.

Introduction:

Today’s world has made technology an integral aspect of the twenty-first century. One piece of technology that is developing daily and has a significant impact on people’s lives is the internet. It has developed into a useful and appropriate method for online banking and other public transaction. Consequently, public information becomes a security thief and becomes a significant security issue. Social Engineering is a tactic used by fraudsters to attempt to obtain personal information about users by preying on human weakness as opposed to software flaws

Phishing is a fraudulent technique that uses social engineering as well as technical methods to get user passwords for financial accounts and personal identification. Phishing attacks are becoming more frequent every day, posing a threat to user privacy and security.

As per the findings of the Kaspersky Lab and Anti-phishing Working Group (APWG), the number of phishing attacks in 2016 grew by 47.48% Malicious software is installed on computers by technical approaches in order to directly steal information of use credentials and passwords.

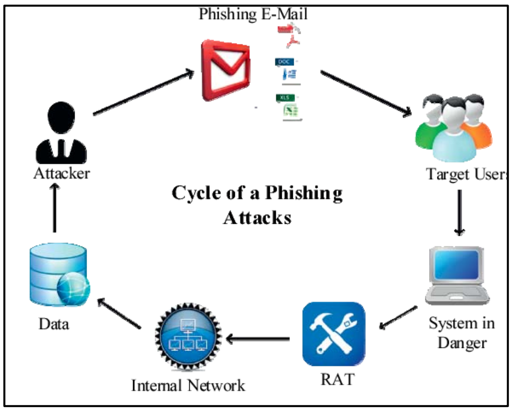


Figure 18.1: Cycle of Phishing attack

Moreover, the researchers have use ML techniques, it is also a Branch of AI. It performs task and capable of learning and acting in intelligent way. There are two types of Machine Learning Techniques Supervised and Unsupervised learning. Supervised Learning it states that, the labled dataset is used to train the model by giving set of features and data associated with target label releated data, once model is trained it can predict the outcomes. On the other hand, Unsupervised Learning states that, In the training process it can generate the new data without giving any target label to get accurate results Random Forest and Decision Tree algorithms are used Because, It has High Accuracy, a relatively robust and it has high performance. Phishing caused sever damages many of the Organizations and global economy.

There are 4 types of Phishing attacks:

1. E-mail phishing attack
2. Whaling attack
3. Spear attack
4. Vishing attack

1. E-mail phishing attack:

Phishing attack occurs when fraudsters send phony emails or texts to rouge websites. There is a chance that the website has malware on it that can damage organizations and systems. They could be made to click track users' private information, such passwords, or to move money. Both large and small-scale organizations are susceptible to phishing emails. It can be the start of a targeted attack against your business or a particular employee, or it might be a widespread campaign in which emails are blasted out of random to millions of inboxes. The attackers in these targeted efforts leverage personnel and company information to further authenticate their message. Most commonly referred as Spear Phishing.

There are numerous benefits in doing this:

- Emails from your organization are more likely to be received by recipients in their inboxes than to be marked as spam.
- From a reputational standpoint, no company wants fraud and scams to become associated with their brand.

2. Whaling attack:

Whaling is one of the targeted phishing attacks. Whaling is digitally enabled the fraud through social engineering trick that designed to encourage the users to sharpen the sensitive information to perform a secondary action such as transferring the funds Whaling does not require the technical knowledge it can deliver huge returns. Payment Services and financial institution are the most targeted.

3. Spear attack:

Spear-phishing is a deliberate attempt, frequently with malevolent intent, to get private data from users, including passwords and account information. This is accomplished by gathering personal information about the user, including friends, birthplace, place of employment, places they frequently visit, and recent internet purchases. With 91% of attacks, this is the most effective way to obtain information from the internet.

4. Vishing attack:

One kind of social engineering attack that takes place over the phone is called “vishing” callers are attempting to obtain your personal information in order to use ti to commit more crimes.

Their objective is to obtain the details of your bank account. Callers who engage in vishing attacks rely on societal normal and expectations to persuade the target to communicate with them.

They pose as representatives of your bank, the police, the tax department, or the government. You may feel obligated to answer when someone you trust in authority calls, or you may just be interested and uneasy. Scammers are aware of this and take advantage of your emotions to lure you in.

1.1 Features of Phishing E-mails:

Grammar and spelling Errors:

Poor grammar and spelling are typical indicators of a phishing email. For outgoing emails, the majority of company personnel have enabled the spell check function in their email software. Most web browser alow you to apply the autocorrect or highlight feature. Since the emails are coming from a professional source, we anticipate that they will be error-free in grammar.

Inconsistencies in Email Addresses, Links and Domain Names:

Looking for discrepancies in email addresses, URLs, and domain names is another technique to spot a phishing attack. For instance, it is worthwhile to verify that the emails match the prior email addresses. In the event that email appears to be from Paypal yet the link's domain does not contain "Paypal.com". Clicking the provided link is not advised if domain names do not match.

Threats or a Sense of Urgency:

Negative Consequences-threatening emails should always be viewed with mistrust. Another strategy is to incite the recipient to move quickly by creating a sense of urgency that encourages or even demands action right away. The con artist thinks that if the victim reads the email quickly, they won't have a chance to carefully review the content, which will allow additional irregularities related to a phishing operation to go unnoticed.

Request for Credentials, payment Information or Other personal details:

When an attacker creates a phony landing pages that recipients are sent to via a link in an email that appears legitimate, this is one of the more sophisticated forms of phishing emails. The fraudulent landing pages will ask for payment to settle an unresolved issue or have a login box. If the email came as a surprise, the receivers should go to the assumed website by typing in the URL rather than clicking on a link. This will prevent them from entering their login information on the fraudulent website or praying the attacker money.

Unusual Request:

Expanding on the previous point, one further clue that the email could be malicious as if it requests an action that is out of the ordinary. For the instance, if you receive an email posing as the IT department requesting that you install a program or click on link to patch your computer, but this kind of work is usually handled centrally, it's likely a phishing email and you shouldn't follow the instructions.

II. Literature Survey:

Literature survey plays a vital role in research work as it gives necessary background information in the area of Interest. And with the help of Literature survey will able to know about the various attacks in cyber security and detects the attacks based on the Machine Learning Techniques. In paper [1] , we have presented about the "Techniques of phishing attack", where the users lure about the cybercrime. Here, they choose the Random Forest algorithm to process the combination of Features. The System act as number of additional functionalities to the internet browser as extensional automatically it notifies the users when it detects the phishing website.

In paper [2], we have discussed about the "Malicious detection of Phishing using Machine Learning". They choose the random Forest and decision Tree algorithm; Random Forest was used for classification and Decision Tree was used for Classify the website.

A model to identify phishing attempts has been constructed as part of the proposed research project using machine learning techniques and algorithms such as Random Forest (RF) and Decision Tree (DT). ML processing helped create the standard dataset of phishing attacks can be downloaded from kaggal.com.

In paper [3], author have presented about the “Phishing attack” where, best way to prevent phishing attack is to educate the users about the different kinds of assaults because, the phishing attack is successful only when the user click on the chosen link. Although many methods and algorithms have been proposed to detect phishing websites, phisher’s have evolved their methods to escape from these detection methods.

In paper [4], author have discussed about the “Implemented and evaluated of classifier” on the phishing website it includes twelve classifier dataset that consists of 6156 legitimate websites and 4897 phishing websites. By classifying the algorithms Support vector machine, Random Forest, Decision Tree, KNN, Gradient Boosting get a good performance in classifying the dataset.

In paper [5], author have presented about the, “Phishing detection of Cyber criminals” phishing is a social Engineering cyber criminals used the common Social Engineering trick identify the phishing is to get the personal information from the user such as credit card information, user names and passwords.

In paper [6], author have discussed about the, “Phishing Website Detection” One of the common issues is difficult to handle phishing. It is necessary to establish several forms of authentication for email networks. The phishing attack attempt is successfully only when the user clicks on the given link. To prevent the phishing attack the best way to educate the users about the cyber-criminal scams.

In paper [7], author have presented about the, “Detection of Cyber Crimes” Cyber criminals who rely on their legitimate use of digital assets, particularly personal data, harm others. Identity theft, which is defined as using someone else identity to steal and use that person’s information for one’s own gain- not just to steal money but also to commit other crimes-is one of the most dangerous crimes that may happen to any internet user.

In paper [8], author have discussed about the,”Phishing detection using ML algorithms” (i)Determine how well traditional ML algorithms detect phishing as a result of excessive feature distribution and web development technology changes; (ii)Determine how well CL-based phishing detection performance on traditional ML algorithms and models, allowing for better learning outcomes with less training data and less time.

In paper [9], author have presented about the: (1) identify the mitigation strategies against the phishing attack; (2) identify the phishing vectors that are primarily taken into consideration in the development of the mitigation strategies; and (3) synthesize guidelines and recommendations of anti-phishing for the organization and end-users. Given that the manifestations and forms of phishing alone pose a serious challenge. (4) To recognize the problems and gaps in the art of state.

In paper [10], author discussed about, “Detection of Phishing” the detection of black and whitelist technology has the high-speed detection. The main drawback, though. Is that the detection rate is dependent on how many websites are on the whitelist and blacklist; also, it is vulnerable to zero-day assaults. Services provided by third parties, such as EHOIS data, network traffic detection, and website ranking, to address problems with blacklisting methods.

In paper [11], author have presented about the contributes of solving the phishing problem by using Google web browser. This functionality is being developed with the help of JavaScript PL. The use of blacklisting and semantic analysis techniques will assist us in recognizing and thwarting phishing attacks. In addition, a database of phishing sites is created, and patterns are identified by analyzing the text, links, photos, and other data present on the website.

In paper [12], author have discussed about the Phishing is kind of Cyber-attack, it is most dangerous to acquire the personal information, credit card details, account details, organizational details and transactions. Phishing websites seem to like the appropriate ones and it is difficult to differentiate among those websites.

In paper [13], author have proposed an approach of phishing detection system to detect the black and whitelist URL also referred as phishing websites. The individuals can alert while browsing or accessing the website. This process adopts the AUP (Agile unified process) for the development the process. This new process of black and whitelist based on detection verifies validity of the website.

In paper [14], author proposed the Phishing detection using ML algorithms, three distinct datasets were utilized to compare the outcomes after various algorithms were applied to the URL analysis. The suggested system is tested using some current datasets from the literature, and the outcomes are compared with the most recent publications in the field. The suggested model will have a high success rate. In paper [15], author discussed about the practice of using ML in cyber security is still in its infancy or experimental stages, demonstrating a substantial gap between practice and research. As a result, the use of ML algorithms and techniques in cyber security that is presented in a very disjointed manner, which makes it difficult to deploy in practice.

2.1 Issues Identified:

From the above Literature survey, the identified issues are:

- The issues with the phishing attack are that, the attacker continuously look for the new creative ways to fool the users into believing their action to involve in website or email.
- Social Engineering is a tactic used by cybercriminals to deceive a victim into disclosing personal information. It is based on the taking advantage of human weakness Rather than software flaws.
- Cyber criminals can get around security systems by hacking less-protected networks belonging to third parties.

- System protected by weak passwords are now accessed from unprotected home networks, sticky note passwords.

III. Proposed Methodology:

After reviewing the literature, the problem description is determined that machine learning is the study of algorithms used in mathematical modeling with probabilitic theories, wherein the decision to solve a problem is predicated on a certain quantity of historical data or scenario related to the problem. Building mathematical models and integrating complex equations such that the target variable's output value depends on the dependent variable is the process of machine learning. Through the examination of distinct characteristics seen on phishing and legitimate websites, machine learning algorithms are able to forecast the likelihood of a new, unidentified website being classified as either of the two. Based on known results, supervised learning is a predictive model.

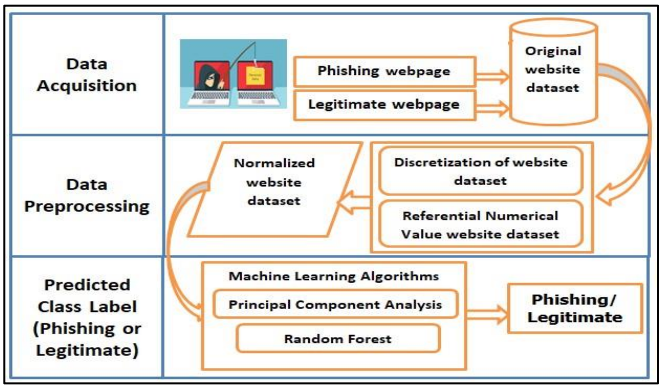


Figure 18.2: The Framework of phishing detection

1. Data Acquistion:

- **Features Extraction:** There are some patterns and traits that can be categorized as features of phishing assaults. We address every aspect of phishing websites that has been utilized in prior studies in this section. Additionally, as we examine the traits and patterns of phishing, we discover a few new traits that qualify as features. There are 36 phishing features in all, 3 of which are brand-new.

They fall under three primary groups, which are displayed below:

- The URL can be used to extract features.
- The content of pages can be used to extract features.
- Page ranks can be used to extract features.

Since the goal of phishing is to steal sensitive data from visitors, including passwords and emails, we are using the amount of entered emails and passwords as the new features for the website. We take into account the quantity of input with the password or email type as a feature for phishing websites.

Another new feature is the quantity of buttons. While researching phishing features, we found that many phishing websites utilize standard buttons rather than submit buttons, thus we classify this as a feature for phishing websites.

2. Data Pre-processing:

Data preprocessing is the essential technique in the Machine Learning application. It can be done from the formatted using data mining techniques and row data. A Noise-free and clear dataset is was needed for the analysis of the dataset. The raw dataset contained missed and incomplete values which are filled and completed by Machine Learning process.

Feature Explanation: The crucial step in the computation and analysis of the dataset is feature selection. There are 32 features in the table below. One can make assumptions about whether someone is malevolent or mistrustful based on the features provided. The features include classes, domains, prefixes, suffixes, IP addresses, indexes, and more.

Feature Name	***Feature Explanation***
Index	Used To Find The Search Engine.
Using IP	Used IP Addresses Instead Of DNS.
Long URL	Used For More Than Hundred Characters.
Short URL	Used Form Of URL. Bit. Ly Is An Example Of Short URL
Symbol@	Used As Special Character.
Redirecting//	Get Another Destination.
Prefixs Uffix	Present At The Beginning And End Of The Original Word.
Sub Do Mains	Present Before The Main Domain To Separate The Website Into A Section.,
HTTPS	Modern Browser Used For Handling Secure Website.
Domain Reg Len	Domain That HAS Been Registered. Paid Advance Amount To Start The Website.
Favicon	It Is Used To Save The Space Of The Website.
HTTPS Do Main URL	Used For Secured HTTP Widely With TLS/SSL Protocol.
Request URL	Used For Request The Resource From Server To Client Side.
Anchor URL	Clickable Text Used For Hyperlinks.
Links In Script Tags	Script Tag To Manipulate The Image, Form Validation And Dynamic Content
Server For M Handler	Item Identified In The Server For The Client.
Info Email	Used With Its Domain Or Business Website.
Ab Normal URL	Not Normal URL Unlikely To Occur
Website For Warding	Redirect Into The Single Web
Status Barcust	Message Displayed At The Bottom Of The Screen

FEATURE NAME	***FEATURE EXPLANATION***
DISABLERIGHTCLICK	STOP SAVING OF WEB PAGE
USING POPUP WINDOW	DISPLAY WITH A MENU ON THE SCREEN AND DISAPPEARS AFTER A CLICK
I FRAMEREDIRECTION	SURVEY WEBSITE AND THEN REDIRECT
AGE OF DO MAIN	TIME DURATION OF THE DOMAIN THAT HAS ALIVE
DNS RECORDING	PRESENT THE IMFORMATION ABOUT IP ADDRESS.
WEBSITE TRAFFIC	NUMBER OF USERS VISITED THE WEBSITE.
PAGERANK	RANK THE WEBPAGE
GOOGLE INDEX	USED TO LIST IN GOOGLE
LINKS POINT	RANK THE WEBPAGES
STATS REPORT	COLLECT INFORMATION ABOUT DOWNLOADED FILES.
CLASS	IT CONTAINS ATTRIBUTES AND BEHAVE

Methodology:

In the above features, we study all the features of phishing attack to indicate the strongest, weakest and to remove the irrelevant features; the study is based on examining all possible combination of 36 features. The size of all possible combination can be specify using this formula:

$$\sum_{k=1}^{36} = \frac{n!}{k!(n-k)!}$$

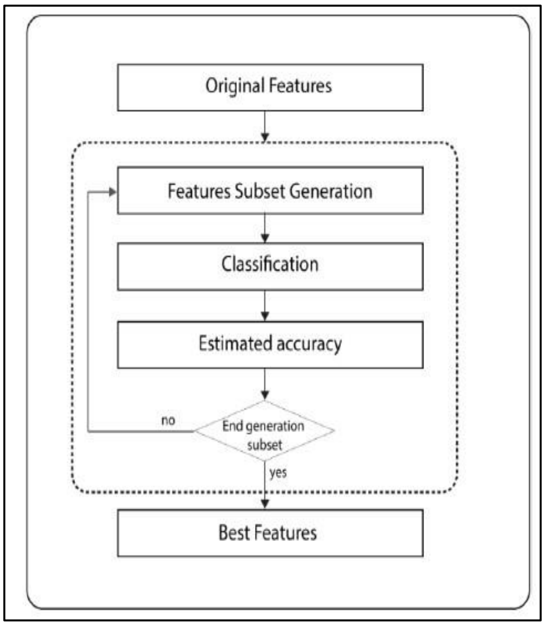


Figure 18.3: The process of feature selection

Here, k is the number of features taken that from 1 to 36. Where, n is the total number of features that is 36. Since the number of all possible combination are huge numbers. Study will be summarized in all the maximum and minimum result of all k combination. At the end of the result, the higher accuracy with the smallest number will be taken for better combination.

The main function of the system is to decide the state of the website if it is a phishing or legitimate website. This function can be performed using the algorithm as shown in Figure 18.3. This algorithm will be triggered whenever the user enters a new website, the role of the algorithm is to extract the features of the website using URL and Document Object Model (DOM) object. The URL used to extract the URL's and page rank's features

IV. Results and Discussion:

ML algorithm	*Accuracy*	*Recall*	*Precision*	*F1 Score*
Decision Tree	**0.9193**	**0.938**	**0.8804**	**0.9085**
Random Forest	**0.9695**	**0.421**	**0.9687**	**0.5875**

One of the greatest machine learning algorithms is Random Forest. It is a regression and classification algorithm that uses supervised learning. It combines an average model with a learning model using the bagging method, which improves overall prediction. A random forest, which combines numerous individual trees, can yield results with a high degree of accuracy. It started by choosing samples at random. Even with missing data, RF outperformed a single decision tree in terms of accuracy. It could solve the issue of over-fitting.

ML Algorithms	*Correlation coefficient*	*Mean absolute error*	*Root mean squared error*	*Relative absolute error*
RF	**0.6448**	**0.2201**	**0.2877**	**76.3956**
DT	**0.6388**	**0.2189**	**0.2905**	**76.5318**

The chosen dataset's feature and data analyses serve as the foundation for the detection of phishing attacks. The performance and accuracy in feature classification are displayed in the confusion matrix. For performance validation, which was determined using the confusion matrix, performance, accuracy, mean, and F1 score were employed.

The following equation can be used to determine the confusion matrix's table layout for the performance projection, as seen in the table above:

Precision = TP / (FP+TP)

Recall = TP / (FN+TP)

Accuracy = (TN+TP) / (TN+FP+TPS=FN)

F1 Score = 2 * ((precision*recall) / (precision + recall))

Conclusion:

From the above papers we concluded that, In this project we have investigate many strategies to identify the phishing detection. The Standard datasets of phishing attacks detection are downloaded from kaggle.com were we used as input and output for the ML algorithms. The two popular machine learning algorithms were used Decision tree and Random Forest algorithm, it helps to implement and analyze the selected dataset. Decision was used to classifying the website and Random Forest is used to classifying. Finally, Confusion matrix is used to validate the accurate result of the performance. The accuracy of Random Forest is 97%. From the literature survey papers we have learned different types of phishing attacks and methods, steps of collecting, preprocessing, can be done from the literature survey papers. Different Features can be explained and extracted from the dataset

References:

1. AO Kaspersky. (2019). The Dangers of Phishing: Help employees avoid the lure of cybercrime. [Online] Available: https://go.kaspersky.com/Dangers-Phishing-Landing-Page- Soc.html [Oct 30, 2019].
2. O.K. Sahingoz" Financial threats in 2019: Every Second Phishing Attack Aims to Steal Your Money" Internet: https://www.kaspersky.com/about/pressreleases/2019 financial-threats-in-2018. Feb 22, 2019[Oct 30, 2019].
3. AO Kaspersky. (2019). The Dangers of Phishing: Help employees avoid the lure of cybercrime. [Online] Available: https://go.kaspersky.com/Dangers-Phishing-Landing-Page- Soc.html [Oct 30, 2019].
4. M. Korkmaz, O. K. Sahingoz, and B. Diri, "Detection of Phishing Websites by Using Machine Learning-Based URL Analysis," in 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020: IEEE.
5. " Financial threats in 2020: Every Second Phishing Attack Aims to Steal Your Money" Internet: https://www.kaspersky.com/about/pressreleases/2020 financial-threats-in-2022, 2020[Oct 30, 2020].
6. Y. Zhang, J. I. Hong, and L. F. Cranor," Cantina: A Content-based Approach to Detecting Phishing Web Sites," New York, NY, USA, [2020]
7. M. Blasi," Techniques for detecting zero-day phishing websites." M.A. thesis, Iowa State University, USA, 2020
8. J. Ma, L. K. Saul, S. Savage, and G. M. Voelker," Learning to Detect Phishing URLs", ACM Transactions on Intelligent Systems and Technology, Article 30, Publication date: April 2020.
9. C. Ludl, S. McAllister, E. Kirda, and C. Kruegel, "On the effectiveness of techniques to detect phishing sites," in International Conference on *Detection of Intrusions and Malware, and Vulnerability Assessment*, Springer, 2021.
10. R. M. Mohammad, F. Thabtah, and L. McCluskey, "Phishing websites features," *School of Computing and Engineering, University of Huddersfield*, 2021.
11. L. A. T. Nguyen, B. L. To, H. K. Nguyen, and M. H. Nguyen," Detecting phishing web sites: A heuristic URL-based approach," in 2022 International Conference on Advanced Technologies for Communications (ATC 2013), 2022.
12. R. M. Mohammad, F. Thabtah, and L. McCluskey, "Phishing websites features," *School of Computing and Engineering, University of Huddersfield*, 2022.

13. R. S. Rao and S. T. Ali, "A computer vision technique to detect phishing attacks," in *2023 Fifth International Conference on Communication Systems and Network Technologies*, pp. 596–601, IEEE, 2023.
14. S. E. Raja and R. Ravi, "A performance analysis of Software Defined Network based prevention on phishing attack in cyberspace using a deep machine learning with CANTINA approach (DMLCA)," (in English), *Computer Communications*, Mar 2020.
15. R. M. Mohammad, F. Thabtah, and L. McCluskey, "Phishing websites features," *Computing school and Engineering, University of Huddersfield*, 2015.

ISSN: 2582-8118 ***Volume 4, Issue 5; Sept 2024***

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYSIS IN SCIENCE AND ENGINEERING

Web: https://www.iarj.in/index.php/ijrase/index

78. Glimpse into the Future - Holographic Projection Technology

Nisha Uday Araballi, Prateek Arunkumar Jahagirdar, Dr. Govindraj Chittapur

Department of Computer Science,
Basaveshwar Engineering College, Bagalkote.

Abstract:

Holographic projection technology, once relegated to science fiction, is rapidly approaching reality. This paper offers a glimpse into the future of holography, exploring its historical trajectory from theoretical concept to the cusp of widespread application. We---examine key milestones in holographic development, delve into current uses across various fields, and then unveil the transformative potential holography holds for the way we learn, communicate, and interact with the world around us.

I Introduction:

Holographic projection technology, the creation of three-dimensional images that appear to float in space, has captured our imaginations for decades, it has captivated imaginations for decades, promising a world where light dances to create three-dimensional illusions. From a theoretical concept in 1947 by Dennis Gabor, holography has undergone a remarkable journey. Early milestones included the creation of the first true hologram and the development of pulsed lasers, paving the way for more sophisticated techniques.

"As the name suggests, holographic communication depends on holography technology, which has made significant progress in the past decade. There are different stages in the development of holography technology. Optical holography generates holograms via recording and recreating optical wave front, and the corresponding holograms are recorded interference patterns (e.g., on photographic emulsions) of an "object wave" and a "reference wave." When the recorded interference pattern is illuminated by the reference wave, a 3D light field can be recreated using diffraction." [1]

Today, holographic projections are no longer confined to science fiction. We see them in museums, concert venues, and even our homes. They offer a unique blend of realism and interactivity, blurring the lines between the physical and digital worlds. This research paper delves into the fascinating world of holographic projection. We will explore the technology's history, highlighting pivotal moments that propelled it forward. We will then examine the current state of the art, discussing its applications and limitations.

Finally, we will cast an eye towards the future, exploring the potential breakthroughs that could revolutionize how we interact with information and each other. This exploration is particularly timely, as holography stands at a turning point. With advancements in lasers, computer processing, and material science, the technology is poised for significant leaps. Understanding its past and present will equip us to navigate its exciting future.

II Milestones of Holographic Projection:

The origins of holography can be dated back to the 1940s, attributed to the pioneering work of Hungarian-British physicist Dennis Gabor, who delved into the realms of electron microscopy and inadvertently introduced the concept of holography. Gabor's groundbreaking contributions to the field were duly recognized when he was awarded the prestigious Nobel Prize in Physics in 1971, solidifying his place in scientific history. Following this pivotal moment, a series of significant milestones emerged over the years:

In the year 1960, a monumental achievement was realized with the advent of the Laser, a transformative invention credited to Theodore Maiman at Hughes Research Laboratories. This revolutionary development heralded a new era for holography by providing the much-needed coherent light source essential for the creation of high-fidelity holograms, pushing the boundaries of the field to unprecedented heights.

Moving forward to 1962, a pivotal moment in the history of holography unfolded with the creation of the First Laser Hologram by Emmett Leith and Juris Upatnieks at the esteemed University of Michigan. This historic event marked a significant leap forward as the duo successfully crafted a three-dimensional image utilizing laser light, showcasing the immense practical potential inherent in holographic technology.

Fast forward to 1972, a momentous occasion took place with the inception of the First Holographic Art Exhibition at the Cranbrook Academy of Art in Michigan. This groundbreaking exhibition served as a platform to unveil the artistic possibilities of holography, signaling the dawn of holography's profound influence in the realm of art and creativity.

Transitioning into the 1980s, holography witnessed a surge in Commercial Applications, particularly in the domains of security and authentication. The integration of holograms on credit cards, banknotes, and various consumer products emerged as a pivotal strategy to combat counterfeiting and ensure the integrity of these items.

In the year 2008, a significant breakthrough was achieved with the development of Real-Time Holographic Video by a team of researchers at the esteemed University of Arizona. This innovative technology enabled the rendering of three-dimensional images in real time, unlocking novel avenues for communication and entertainment, thereby reshaping the landscape of holographic displays.

By 2012, a paradigm shift occurred in the realm of holography with the introduction of Holographic Telepresence by tech giant Microsoft. This cutting-edge technology revolutionized remote communication by enabling individuals to interact with lifelike holographic representations of others in real time, paving the way for a new era of immersive and dynamic virtual interactions. Advancing into 2018, a groundbreaking development emerged with the creation of Interactive Holograms by a team of researchers at RMIT University in Australia. These interactive holograms allowed for tactile interaction and manipulation, enhancing user engagement and fostering a deeper connection with virtual objects, thereby revolutionizing the user experience in the realm of holography.

As we progress into the 2020s, holographic projections have witnessed Mainstream Adoption across various industries, propelled by advancements in augmented reality (AR) and virtual reality (VR) technologies. This widespread integration of holographic displays has permeated diverse sectors ranging from entertainment and gaming to education and healthcare, underscoring the growing significance and ubiquitous presence of holography in contemporary society.

III Literature Survey:

Recent progressions in the field of holographic projection technology have ushered in a plethora of intriguing possibilities within the realm of entertainment applications. The incorporation of color liquid crystal gratings within time-sequenced holography systems has been instrumental in tackling pertinent issues such as limited viewing angles and chromatic aberrations, thereby significantly augmenting the overall performance of three-dimensional displays [2].

Moreover, the emergence of dynamic holography methodologies, including but not limited to the innovative 3D scattering-assisted dynamic holography, has facilitated unparalleled precision in depth modulation, facilitated high-density multi-plane projection capabilities, and enabled dynamic vectorial projections, all of which collectively introduce novel vantage points for the realm of lifelike three-dimensional video displays [3]. Additionally, the ongoing advancements in computer-generated holography techniques, particularly those leveraging deep learning methodologies and the establishment of top-tier hologram datasets, hold the promise of delivering photorealistic three-dimensional holographic projections characterized by meticulously corrected vision aberrations, thus significantly enriching the visual landscape for entertainment-centric purposes [4]. These groundbreaking developments in holographic technology not only serve to elevate the standards of image quality and depth regulation but also pave the way for the realization of immersive and interactive entertainment encounters across diverse sectors encompassing art, science, and education [5]. Recent progress in holographic projection technology provides an immersive 3D viewing experience without the need for headgear, thereby decreasing eye strain. This technology has various applications, such as family entertainment systems that offer visual immersion and help mitigate the risk of eye damage from prolonged use of digital devices. [6]

IV Digital Holography:

Digital holography is an advanced imaging technique that captures and reconstructs three-dimensional images using digital technology. Unlike traditional methods, which rely on photographic films or plates, digital holography employs electronic sensors and computational algorithms to record and process holographic information.

Digital holography captures the interference pattern created by the interaction of an object beam and a reference beam using electronic sensors. The recorded data is processed digitally to reconstruct the three-dimensional image.

Generation/Recording: In digital holography, a laser beam is divided into two parts: the object beam and the reference beam. The object beam illuminates the subject, and the light reflected from the subject interferes with the reference beam. This interference pattern is recorded by an electronic sensor, creating a digital representation of the hologram.

Reconstruction: The captured interference pattern is converted into digital data. Computational algorithms, such as Fourier transforms, are then used to process this data and reconstruct the original three-dimensional image. This digital processing enables precise and detailed visualization of the object.

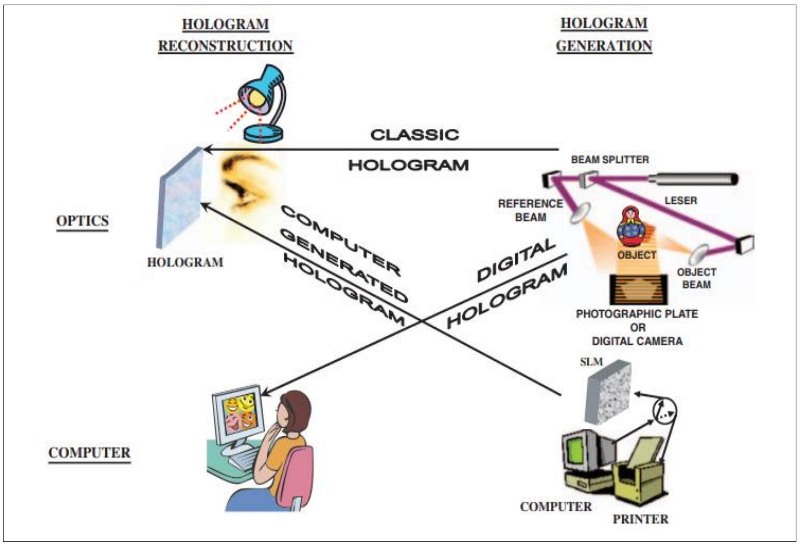


Figure 1: Schematic of the mutual relations in holography discipline between digital computers and optics. SLM, spatial light modulator. [3]

Equipment Used in Digital Holography:

CCD (Charge-Coupled Device) Sensors: CCD sensors are commonly used in digital holography due to their high sensitivity and ability to produce high-quality images. CCD sensors convert the interference pattern into electronic charges, which are then digitized. Each pixel in the CCD sensor captures a portion of the interference pattern, allowing for detailed image capture.

CMOS (Complementary Metal-Oxide-Semiconductor) Sensors: CMOS sensors are another type of electronic sensor used in digital holography, known for their speed and efficiency. CMOS sensors capture the interference pattern and convert it to digital data directly at each pixel. This process enables faster readout and lower power consumption compared to CCD sensors.

Computational Hardware and Software: Processing holographic data requires advanced computational resources and specialized software.

Hardware: High-performance computers and GPUs are used to handle the extensive data and complex computations necessary for image reconstruction.

Computer-generated holography (CGH) relies on computational techniques to create holograms instead of traditional optical recording methods. Here is the essential hardware needed:

- ***High-Performance Computer:*** A powerful computer with a robust CPU and ample RAM is necessary to handle the intensive computations involved in generating holograms.
- ***Graphics Processing Unit (GPU):*** GPUs are crucial for speeding up the holographic computation process through parallel processing. Modern GPUs with many cores and substantial memory are ideal.
- ***SLMs (Spatial Light Modulator):*** Devices like liquid crystal on silicon (LCoS) or digital micro mirror devices (DMDs) modulate light based on the computed holographic data. The resolution and refresh rate of the SLM play a significant role in the quality and responsiveness of the holographic display.
- ***Coherent Laser:*** A coherent light source is necessary to illuminate the SLM and produce the holographic image. The laser's wavelength must match the SLM's specifications.
- ***Lenses and Mirrors:*** These components are used to focus, direct, and shape the laser beam onto the SLM.
- ***Beam Splitters and Combiners:*** Used in setups requiring multiple beams.

Network Speed Required to Transfer Data for CGH: The following graph illustrates the relationship between different internet bandwidth capacities and the types of holographic applications that can be effectively implemented at each level.

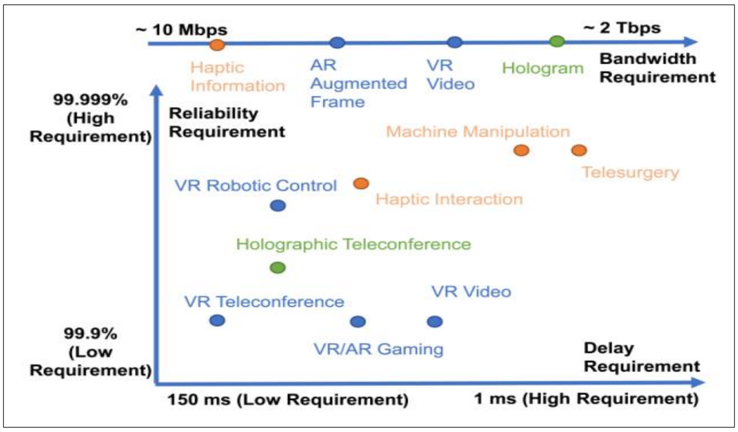


Figure 2: Requirements of representative immersive communications use cases: an illustration. [1]

As bandwidth capacity increases, the complexity, resolution, and interactivity of holographic applications also rise. Simple 3D visualizations and basic communication can be achieved with lower bandwidths, while high-quality, real-time, and immersive holographic experiences require significantly higher bandwidths.

This progression underscores the importance of robust network infrastructure to support the growing demand for advanced holographic applications across various fields.

Software:

Software tools apply algorithms such as Fourier transforms and phase unwrapping to process the captured data and generate three-dimensional images. These tools are crucial for accurate and detailed holographic reconstruction.

1. Fourier Transform Algorithms [7]:

These mathematical operations convert spatial data into frequency data and vice versa. In CGH, Fourier transforms are employed to analyze and synthesize holographic patterns. The Fourier Transform is a crucial tool in computer-generated holography (CGH), transforming spatial data into the frequency domain to facilitate the creation and analysis of holographic patterns. Here's an overview of how the Fourier Transform algorithm is applied in CGH:

The Fourier Transform is a crucial tool in computer-generated holography (CGH), transforming spatial data into the frequency domain to facilitate the creation and analysis of holographic patterns. Here's an overview of how the Fourier Transform algorithm is applied in CGH:

Object Representation: Define Object: Represent the object to be holographically imaged as a two-dimensional complex function O(x, y). This function includes both the amplitude and phase information of the object.

Fourier Transform of the Object: Apply Fourier Transform: Compute the Fourier Transform of the object function O(x, y) to get its frequency representation O(u, v) :

$$O(u,v) = \mathcal{F}\{O(x,y)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} O(x,y)e^{-j2\pi(ux+vy)}dxdy$$

Frequency Domain: This result, O (u, v) , represents the object in the frequency domain, essential for creating the hologram.

Reference Wave: Define Reference Wave: Choose a reference wave R(x, y), usually a plane wave or spherical wave. For simplicity, a plane wave

$$R(x,y) = e^{j2\pi(wx+vy)}$$

.

Interference Pattern Calculation:

Compute Interference: The hologram is formed by the interference pattern between the object wave and the reference wave. The interference pattern H(x, y) is given by:

$$H(x,y) = |O(x,y) + R(x,y)|^2$$

Expand and Simplify: This expands to:

$$H(x,y) = O(x,y)O^*(x,y) + R(x,y)R^*(x,y) + O(x,y)R^*(x,y) + O^*(x,y).$$

Here, O*(x, y) and R*(x, y) are the complex conjugates of the object and reference waves, respectively.

Hologram Formation:

Hologram Matrix: The resulting interference pattern H(x, y) is recorded as the hologram. This hologram contains the information needed to reconstruct the original object wave.

Reconstruction:

Apply Inverse Fourier Transform: To reconstruct the holographic image, apply the inverse Fourier Transform to the recorded hologram H(u, v):

$$O'(x,y) = \mathcal{F}^{-1}\{H(u,v)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(u,v)e^{j2\pi(ux+vy)}dudv$$

Extract Object Wave: The reconstructed image O'(x, y) represents an approximation of the original object wave, allowing visualization of the holographic image.

The Fourier Transform algorithm is integral to CGH, enabling the conversion of spatial information into the frequency domain, which is crucial for creating and reconstructing holographic images. By utilizing the Fourier Transform and its inverse, along with careful handling of interference patterns and computational techniques, high-quality holographic images can be generated and visualized.

2. Phase Unwrapping:

- ***Phase Unwrapping:*** This process resolves ambiguities in phase data caused by the periodic nature of phase measurements, converting wrapped phase data into continuous phase data.
- ***Application:*** Essential for precise 3D reconstruction, phase unwrapping ensures that phase information accurately represents the original object's shape and surface details.

3. Image Processing Tools:

- ***Noise Reduction:*** Algorithms designed to reduce noise and enhance the quality of captured holographic data.
- ***Filtering:*** Spatial and frequency domain filters refine holographic images by removing artifacts and improving clarity.
- ***Resolution Enhancement*****:** Techniques such as super-resolution increases the apparent resolution of holographic images.

4. ***Reconstruction Software:***

- ***Wave front Reconstruction:*** Software that reconstructs original light wave fronts from processed data to create 3D holographic images.
- ***Visualization:*** These tools include modules for viewing and interacting with reconstructed holographic images in real-time.

V Applications:

Holographic projection has transcended science fiction and found its way into various aspects of our lives today. Here are some applications:

Entertainment**:**

Live Events: Imagine attending a concert where a long-gone artist performs alongside the live band, thanks to holographic technology. This has been used to resurrect deceased musicians or create life-sized projections of characters. Museums and Exhibitions: Static displays are becoming a thing of the past. Holographic displays can bring historical figures or artifacts to life, allowing visitors to interact with them in a more engaging way.

Education and Training:

- 3D Visualization: Complex scientific concepts like molecules or historical events can be presented in a captivating way with holographic models, aiding in better understanding and retention.
- Remote Collaboration: Students from different locations can interact with the same 3D holographic representation during a virtual lesson, fostering a more immersive learning experience.

Medical Field:

- Surgical Planning: Surgeons can use holographic representations of a patient's organs to visualize complex procedures beforehand, leading to improved planning and potentially better outcomes.
- Medical Education: Learning anatomy can be revolutionized by using 3D holographic models of the human body. Students can explore these models from any angle, gaining a deeper understanding of the body's structures.

Marketing and Advertising:

- Product Launches: Imagine a captivating product launch where a holographic model showcases the product's features and functionalities in a dynamic way, grabbing the audience's attention.
- Trade Shows and Events: Eye-catching holographic displays can be used to promote brands and products with a unique and interactive experience, leaving a lasting impression on potential customers.

The future applications of holographic projections are brimming with endless possibilities.

VI Machine Learning and Deep Learning in Holographic Projections:

Machine learning algorithms play a crucial role in enhancing the accuracy of holographic projections in computer-generated environments by improving the quality of generated holograms and reducing artifacts. Various approaches have been proposed to achieve this, such as diffraction model-driven neural networks like Res-Holo [8], phase dual-resolution networks (PDRNet) for generating high-fidelity holograms with fixed computational complexity [9], and the combination of stochastic gradient descent (SGD), neural networks, and double-sampling Fresnel diffraction (DSFD) to accelerate the generation of large field-of-view holograms [10]. Additionally, techniques like replacing binary pixel values with continuous variables and using surrogate gradients have been employed to optimize binary holograms and train deep learning-based models effectively [11].

By learning holographic light transport and complex-valued convolution kernels, machine learning algorithms can significantly improve simulation accuracy and image quality in holographic displays [12].

By leveraging algorithms that enable systems to learn from data, machine learning enhances the capabilities of holography, improving both the creation and utilization of holographic projections. Machine learning can help in enhancing image quality, noise reduction, real-time processing, fast algorithms, adaptive systems, predictive modeling, simulation and training, customized content, interactive interfaces which are used by the holographic projections.

VII Conclusion:

Holographic projection technology is transitioning from the realm of science fiction into practical and impactful applications across various sectors.

This paper has traced the evolution of holography from its conceptual inception by Dennis Gabor in 1947 to its current state, highlighting key milestones such as the advent of laser technology and the development of digital holography. Holography today is no longer a novelty confined to research labs; it is utilized in diverse fields such as entertainment, education, healthcare, and marketing. From resurrecting iconic performances in concerts to enhancing surgical planning and creating immersive educational experiences, holographic projections are transforming the way we interact with and perceive our world.

Today, holographic projections are making tangible impacts across entertainment, education, medical fields, and marketing. They offer unprecedented opportunities for interactive and immersive experiences, enhancing how we visualize, learn, and communicate. The integration of digital holography and advancements in computational techniques, including machine learning and deep learning, has further expanded the potential applications and improved the quality of holographic displays. These technologies promise to refine the accuracy and realism of holographic projections, paving the way for even more sophisticated and engaging applications in the future.

As we look to the future, the potential of holography is boundless. It promises to revolutionize communication and collaboration, offering more natural and immersive remote interactions. The integration of augmented reality (AR) and virtual reality (VR) with holographic technology is set to blur the lines between physical and digital worlds, providing unprecedented interactive experiences.

In conclusion, the future of holography is bright, with ongoing advancements poised to overcome current limitations and unlock new possibilities. Understanding its historical trajectory and current applications equips us to navigate and harness the transformative potential of holographic projection technology.

References:

1. Xuemin, Shen., Jie, Gao., Mushu, Li., Conghao, Zhou., Shisheng, Hu., Mingcheng, He., Weihua, Zhuang. "Toward immersive communications in 6G." Frontiers in computer science, 4 (2023). doi: 10.3389/fcomp.2022.1068478
2. Lou, Fei., Li, Wenshuo., Fan, Guojiang. (2019). Holographic image stage system.
3. Journal of Holography and Speckle Vol.5, 1–17, 2009
4. Yulong, Liu., Shan, Wu., Qi, Xu., Hubin, Liu. "Holographic Projection Technology in the Field of Digital Media Art." Wireless Communications and Mobile Computing, undefined (2021). doi: 10.1155/2021/9997037
5. Francesco, Reda., Marcella, Salvatore., Marco, Astarita., Fabio, Borbone., Stefano, Luigi, Oscurato. (2022). Reprogrammable Holograms from Maskless Surface Photo morphing. Advanced Optical Materials, doi: 10.1002/adom.202300823.
6. Anzhuo, Li., Fan, Li., Dayong, Ma. "Research and Experimental Application of Family Holographic Projection Entertainment Equipment." undefined (2018). doi: 10.1051/MATECCONF/201817601006
7. Zhang, J., Zheng, L., & Liu, Q. (2013). Computer-Generated Holography for 3D Display. Optics Express, 21(4), 3110-3119.
8. Huadong, Zheng., Zhen, Wang., Xinghua, Shui., Xinxing, Xia. "Diffraction model-driven neural network trained using hybrid domain loss for real-time and high-quality computer-generated holography." Optics Express, 31 (2023).:19931-19931. doi: 10.1364/oe.492129
9. Ting, Yu., Shijie, Zhang., Wei, Chen., Juan, Liu., Xiangyang, Zhang., Zijian, Tian. "Phase dual-resolution networks for a computer-generated hologram." Optics Express, 30 2 (2022).:2378-2389. doi: 10.1364/oe.448996
10. Chentianfei, Shen., Tong, Shen., Qi, Chen., Qinghan, Zhang., Jihong, Zheng. "Machine-learning-based high-speed lensless large-field holographic projection using

double-sampling Fresnel diffraction method." Chinese Optics Letters, 20 (2022).:050502-050502. doi: 10.3788/col202220.050502
11. M., Hossein, Eybposh., Aram, Moossavi., Vincent, R., Curtis., Nicolas, C., Pégard. "Optimization of time-multiplexed computer-generated holograms with surrogate gradients." null (2022). doi: 10.1117/12.2607781
12. Koray, Kavaklı., Hakan, Urey., Kaan, Akşit. "Learned holographic light transport: invited." Applied Optics, 61 (2022). doi: 10.1364/AO.439401

ISSN: 2582-8118 *Volume 4, Issue 5; Sept 2024*

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYSIS IN SCIENCE AND ENGINEERING

Web: https://www.iarj.in/index.php/ijrase/index

79. Integrating YOLO and GAN for Robust Signature Detection and Similarity Assessment

Thejaswini B. L., Dr. H. S. NagendraSwamy, Rakshitha P., Manjuchethan N.

Department of Studies in Computer Science, University of Mysore, Manasagangotri, Mysore, India.

Abstract:

The proposed approach in this study uses the sequence flow of YOLO models and Cycle GAN to detect and extract signatures from scanned documents in conjunction with a template matching strategy. The created dataset was first trained by the model using the v8, v9, and v10 versions of the YOLO model. Owing to its superior signature-detecting skills, YOLO v8 was selected as the top model in a comparison analysis. This YOLO version 8 achieved 98% detection accuracy for signature detection from the scanned documents. Secondly, the model finds and extracts signatures that are noisy. In order to get clear, noiseless signatures, the retrieved signatures were also subjected to a Cycle GAN model-based noise reduction procedure. Thirdly, it assesses the similarity methods used to the original signature templates and the cleaned signature images produced by the GAN in order to verify the efficacy of the noise reduction methodology. The results demonstrate with 92% of average similarities and perform how well the signatures, which makes it valuable for applications requiring precise signature processing. This illustrates how well the suggested approach increases the accuracy of signature extraction, detection, and identification phase, making it a valuable tool for a variety of applications. The results verify that our method has the potential to significantly impact the domains of signature analysis and authentication.

Keywords:

Dataset, YOLO v8, detection, CycleGAN, extraction, similarity.

I Introduction:

In the digital age, the verification and processing of handwritten signatures remain crucial for various applications, including banking, legal documentation, and identity verification. Traditional methods of signature verification, which rely on manual inspection, are time-consuming and prone to human error. Hence, there is a growing demand for automated systems that can efficiently and accurately detect and process signatures from documents. This project addresses this demand by proposing a novel method for signature detection and enhancement using state-of-the-art deep learning models. Deep learning approaches in signature detection utilize advanced neural network architectures to effectively recognize and verify signatures. Convolutional Neural Networks (CNNs) are predominantly used due to their strong performance in image processing tasks. These networks learn to identify unique features and patterns within signature images through training on extensive datasets. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are also applied in dynamic signature verification, capturing the temporal sequence of signature strokes. Techniques such as transfer learning and data augmentation further enhance model performance by leveraging pre- trained models and increasing dataset diversity. In the context of signature detection, YOLO (You Only Look Once) is a real-time object detection algorithm that can be adapted to detect signatures within documents. YOLO divides an image into a grid and predicts bounding boxes and probabilities for each grid cell, enabling the model to detect multiple signatures in a single pass efficiently. Its speed and accuracy make it suitable for applications requiring real-time processing, such as automated document verification systems. By training YOLO on labelled signature datasets, it can learn to precisely locate and identify signatures within various document types, improving the efficiency and reliability of signature detection tasks. The model employs the You Only Look Once (YOLO) object detection framework, specifically versions 8, 9, and 10, to accurately locate signatures within scanned documents. After rigorous training and evaluation, YOLO v8 was selected due to its superior performance in terms of detection accuracy and speed.

Once the signatures are detected, they often contain noise and distortions due to the quality of the scanned documents. To address this, the proposed model utilizes a Cycle- Consistent Generative Adversarial Network (Cycle GAN) to clean and enhance the detected signatures. The Cycle GAN model is trained to remove noise and produce a clean version of the signature, closely resembling the original. To validate the effectiveness of our approach, similarity measure between the GAN-generated signatures and the original signature templates has been mapped and evaluated. This measure ensures that the cleaned signatures maintain their authenticity and can be reliably used for verification purposes. Our proposed method not only automates the detection and enhancement of signatures but also improves the overall accuracy and reliability of signature verification systems. This project represents a significant step towards more efficient and error-free processing of handwritten signatures in various professional and legal contexts.

II Related Works:

Nabin Sharma et al. (2018) [1] proposed Signature and Logo Detection using Deep CNN for Document Image Retrieval. This paper deals with signature and logo detection from a repository of scanned documents, and in this paper, Faster R-CNN and YOLO v2 were examined for automatic detection of signatures and logos from scanned administrative documents. And experiments were conducted on Tobacco-800 and COCO dataset. They got 89.2% accuracy for Faster R-CNN with VGG16.

Rabi Sharma et al. (2021) [2] proposed Detecting Signatures in Scanned Document Images". They have used YOLOV5 and Retina net for signature detection. And experiments were conducted on Tobacco-800 and COCO dataset. They got 87% accuracy for 3000 epochs using YOLOv5.

Kaihong Yan et al. (2022) [3] proposed Signature Detection, Restoration, and Verification: A Novel Chinese Document Signature Forgery Detection Benchmark. They have used Chinese-based signature datasets and experiment done using YOLO v2, YOLO v5 and Faster R-CNN. They got 84% accuracy for YOLOv5 Venkata Rami Reddy Ch et al. (2024)[4] proposed A Review on YOLO v8 and Its Advancements. The paper utilizes various datasets for evaluating the performance of the YOLO v8 model, including the COCO val 2017 dataset. The paper focuses on the analysis of the YOLO v8 model, a real-time object detection system, by presenting its architecture, advancements, and performance evaluation on various datasets in comparison with previous YOLO models.

Chien-Yao Wang et al. (2024) [5] proposed YOLO v9: Learning What You Want to Learn Using Programmable Gradient Information. The research paper utilizes the MS COCO dataset for object detection experiments, Programmable Gradient Information (PGI) methods is used to address information loss in deep networks and enable reliable gradient information for weight updates.

Ao Wang et al. (2024) [6] proposed YOLOv10: Real-Time End-to-End Object Detection". In this paper they have used COCO dataset for object detection. The paper introduces consistent dual assignments for training YOLOs without relying on non-maximum suppression (NMS) during post- processing, enhancing both performance and inference latency.

Deniz Engin et al. (2020) [7] proposed "Offline Signature Verification on Real-World Documents: The research paper utilizes a dataset consisting of 109 individuals' signatures, proposes a method for writer-independent offline signature verification in real-world scenarios, where occluded signatures are compared against clean reference signatures. The main components of the proposed method include a stamp cleaning process based on Cycle GAN and signature representation using Convolutional Neural Networks (CNNs).

Yongbing Zhang et al. (2020) [8] Multiple Cycle-in-Cycle Generative Adversarial Networks for Unsupervised Image Super-Resolution": The dataset used in this paper is the DIV2K dataset provided by Timofte et al., which consists of 1000 images.

The proposed method combines multiple "CycleGANs" to map LR images to HR images without the need for paired training data, addressing the limitations of traditional supervised methods. The results highlight the success of MCinCGAN in handling practical super-resolution challenges and advancing image super-resolution research using unsupervised learning strategies

Shashidhar Sanda et al. (2017) [9] proposed Online Handwritten Signature Verification System using Gaussian Mixture Model and Longest Common Sub-Sequences The research paper utilizes the MCYT-100 dataset. Mainly focuses on online signature verification using the GMM classifier model and LCSS comparison technique, aiming to authenticate signatures reliably and achieve better performance compared to the DTW method. The results indicated that the proposed method using GMM and LCSS achieved effective online signature verification by accurately distinguishing between genuine and forged signatures based on unique physical and behavioural features.

Piotr Porwik et al. (2009) [10] proposed A new signature similarity measure the data used in this paper consists of 50 handwritten signatures, which are represented as a set of features including coordinates of signature points, pen pressure, and speed of writing. By calculating the distance between windows and considering mean differences, the method offers a comprehensive framework for evaluating signature similarities. The results demonstrate the potential of the Czekanowski's measure and the windowing technique in enhancing the accuracy of signature recognition systems.

D Gunawan et al. (2018) [11] proposed The Implementation of Cosine Similarity to Calculate Text Relevance between Two Documents: The research paper utilizes textual data from web pages or documents to implement text relevance calculation for finding topic-specific documents efficiently, Punctuation removal, converting text to lowercase, stop word removal, and extracting root words using the Porter Stemming algorithm are performed to prepare the documents for analysis.

III Proposed Methodology:

An overview of the proposed methodology is given in Figure 1. As can be seen, input was processed based on the annotated documents using three different versions of the YOLO (You Only Look Once) model, namely YOLOv8, YOLOv9, and YOLOv10. Each model is used to detect signatures within the documents. The dataset is split into three subsets for the purposes of training, testing, and validation, following the standard protocol of allocating 70% of the data for training, 20% for testing, and 10% for validation. This methodical approach guarantees a thorough assessment of each model's performance in terms of accuracy and robustness in signature detection.

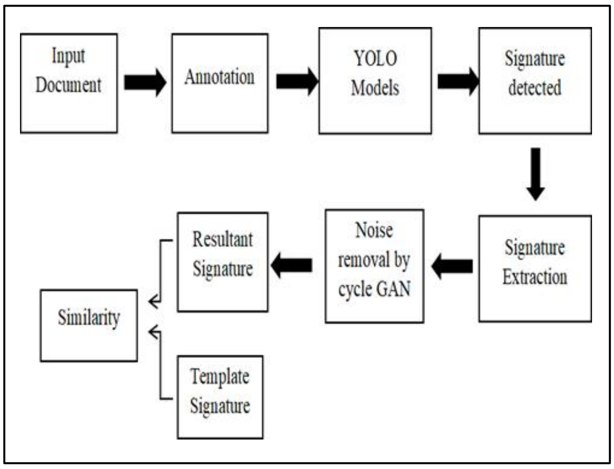


Figure 1: Architecture of The Proposed Model

The proposed model begins with an input document containing signatures. The document is first annotated to highlight the areas where signatures need to be detected. These annotated regions are then processed by YOLO models to detect the signatures within the document. Once the signatures are detected, they are extracted for further processing. The extracted signatures are then subjected to noise removal using Cycle GAN to enhance their quality. The result is a cleaned, noise-free signature, referred to as the resultant signature. This resultant signature is then compared with a pre-provided template signature to determine their similarity, completing the process.

After evaluating three models for signature detection, YOLO v8 emerged as the top performer, delivering the most accurate results. Consequently, YOLO v8's outputs are chosen for further processing. The signatures detected within the bounding boxes are manually extracted for refinement. These extracted signatures often contain unwanted elements like seals or dates, which can obscure the actual signature. To address this issue, the extracted signatures undergo processing with a Cycle GAN model. This specialized neural network is employed to effectively remove noise and enhance the clarity of the signatures. The cleaned signatures produced by the Cycle GAN are then subjected to similarity measures, such as comparing them against a template signature. This step ensures a reliable assessment of how closely the cleaned signature matches the expected standard, thereby validating the signature accurately for verification purposes.

The suggested approach to signature detection that makes use of the YOLO (You Only Look Once) architecture and concentrates on versions 8, 9, and 10. The suggested method entails training three separate models—each representing a distinct YOLO version—to assess and contrast how well they can extract signatures from documents and detect them. You Only Look Once, or YOLO, was initially popularized in 2015. It has since undergone several revisions, each of which improved upon the initial model. The most recent version, YOLOv8, includes improvements in real-time object identification capabilities.

A. YOLOV8:

A cutting-edge deep learning model called YOLOv8 [4] is intended for computer vision applications that require real- time object recognition. YOLOv8's architecture is an evolution of previous YOLO models, utilizing a convolutional neural network divided into two main parts: the backbone and the head. The backbone is based on a modified version of the CSPDarknet53 architecture, consisting of 53 convolutional layers enhanced with cross- stage partial connections. The head comprises multiple convolutional layers followed by fully connected layers responsible for predicting bounding boxes, confidence scores, and class probabilities. Notably, YOLOv8 integrates a self-attention mechanism in the head of the network and a feature pyramid network for multi-scaled object detection, enabling it to focus on various parts of an image and detect objects of different sizes and scales. The model's speed is a key factor in environments requiring quick responses, such as autonomous vehicles and surveillance systems. Its support for various backbones like EfficientNet and ResNet offers adaptability, beneficial in robotics for different operational needs. This model is versatile, supporting a range of vision AI tasks including detection, segmentation, pose estimation, tracking, and classification.

B. YOLOV9:

YOLOv9's [5] real-time object detection capabilities continue the legacy of the YOLO series by quickly processing images or video streams to accurately identify objects without sacrificing speed. It integrates the idea of Programmable Gradient Information (PGI), which ensures that the deep functions maintain the function necessary for the target tasks and avoid information loss during the input process to the deep neural networks, using a reversible extra branch for reliable creation gradients. YOLOv9 optimizes parameters, computational complexity, accuracy and inference speed using the Generalized ELAN (GELAN) architecture.

This design allows users to choose the right computing blocks for different terminal devices, increasing flexibility and efficiency. Experimental results show that YOLOv9 outperforms existing real-time object detection algorithms in terms of accuracy, speed and overall performance on benchmark datasets such as MS COCO. Designed to be customizable, YOLOv9 can be easily integrated into different systems and environments, making it suitable for a wide range of applications such as surveillance, autonomous vehicles, robotics and more**.**

C. YOLOV10:

In May 2024, the YOLOv10[6] real-time object detection model made its debut, marking a major development in the YOLO (You Only Look Once) series. YOLOv10, which was created with the goal of improving efficiency and accuracy, offers several novel features and techniques.

The removal of the need for Non-Maximum Suppression (NMS) during inference is one of its main enhancements. YOLOv10 decreases inference latency (Home) (OpenCV) and improves alignment between the training and inference stages by using a dual assignment method during training (one-to-many and one-to-one assignments).

YOLOv10 uses a comprehensive approach to model design that optimizes accuracy and efficiency. To reduce computational expense, it uses depth-wise separable convolutions for its classification heads. The model also uses a technique called spatial channel decoupled down sampling.

D. Cycle GAN:

A One kind of neural network utilized for unpaired image- to-image translation is called a CycleGAN. It doesn't require paired training examples like conventional GANs do. In order to map images from one domain to another, CycleGAN does not require a direct match between the source and target images.

This is accomplished by ensuring that, in the event that an image is translated from domain A to domain B and back again, it retains its original characteristics. This is known as cycle consistency. Style transfer, photo editing, and other applications are available.

The process begins with the collection and annotation of a diverse dataset containing various types of documents with signatures. This dataset is then used to train the YOLO models, leveraging the architecture's capability to perform real-time object detection with high accuracy. Each version of YOLO v8, v9, and v10—introduces enhancements in terms of model structure, feature extraction, and detection capabilities, which are meticulously evaluated.

E. Similarity Measure:

Once the signature is recovered, the model applies the Cosine Similarity to measure the degree to which the discovered signature resembles the reference signature template.

$$Cosine\ Similarity = \frac{A \cdot B}{\| A \| \| B \|}$$

Where A is the template signature vector and B is the extracted GAN-generated signature vector.

IV Experimental Results:

In order to study the performance of the proposed methodology, experiments were conducted on the dataset created as a part of this research work. The following paragraph provides the details of the dataset created and experimental results.

A. Dataset: A dataset of signature documents that was initially collected manually from users, totaling 150 documents. These documents were then annotated. To increase the sample size, data augmentation was performed, expanding the dataset to 353 documents (Figure 2). To ensure a robust evaluation, the dataset has been divided into training, validation, and test sets, allocating proportions of 80%, 10%, and 10%, severally. This led to a distribution of 285 images for training, 34 for validation, and 34 for testing. It's crucial to emphasize that the dataset was randomly divided, enhancing the diversity and reliability of our experimentation.

Figure 2: Samples of Dataset

B. Results:

- ***YOLOV8:***

After thoroughly training the dataset using the YOLOv8 model with a training, validation and testing split of 80%,10% and 10%. Model was trained on the epochs of 25, ensuring a robust evaluation of the model's performance. YOLOv8 (Figure 3), known for its high accuracy and efficiency in object detection tasks, has proven to be a powerful tool in this scenario. The results from the testing phase were highly promising, with the model demonstrating a remarkable ability to accurately detect signatures in various document formats. This success highlights the YOLOv8 model's sophisticated architecture and its capacity to generalize well from the training data to unseen examples.

The model's precise detection capabilities are crucial for applications requiring reliable signature verification and extraction. Overall, these results underscore the effectiveness of the YOLOv8 model in handling complex signature detection tasks, reaffirming its suitability for deployment in practical, real-world scenarios.

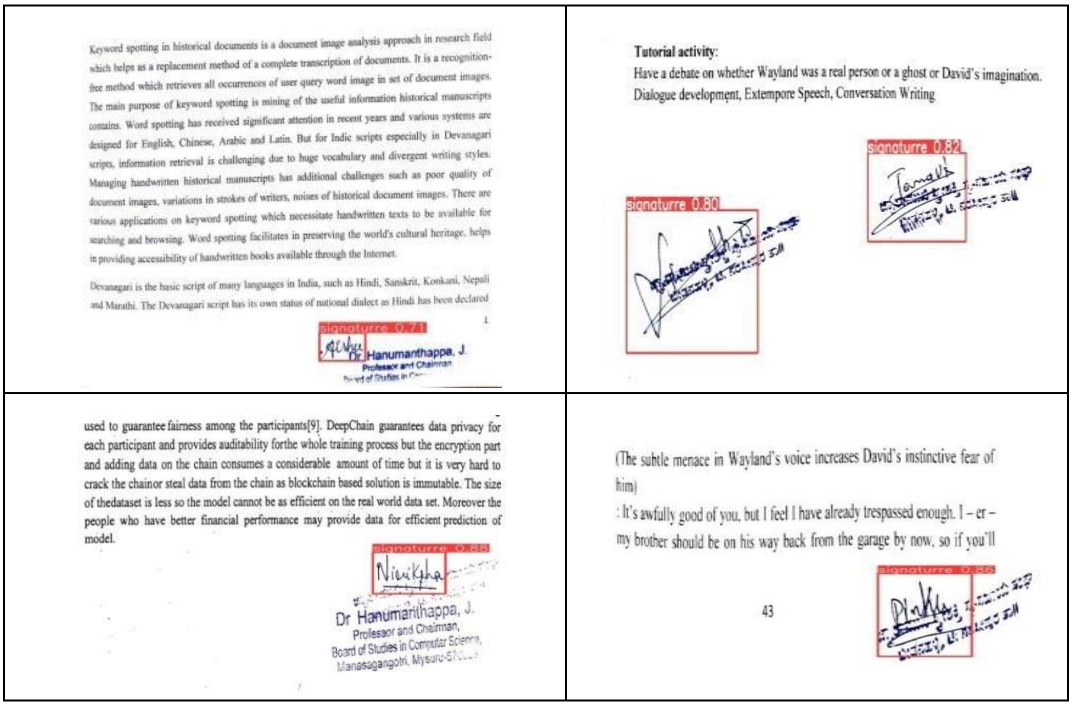


Figure 3: Sample results of signature detection by YOLOv8

- ***YOLOV9:***

The model was trained on 20 epochs after the dataset was properly trained using the YOLOv9 model with the same proportion as used for YOLOv8. Even though YOLOv9 (Figure 4) is renowned for its great efficiency and accuracy in object identification tasks, the testing phase's results showed a mixed performance.

Although some signatures were detected with less than 50% accuracy, the model showed some capacity to detect signatures in a variety of document types. This result emphasizes the difficulties in reaching the same degree of accuracy as YOLOv5. Even with these problems, YOLOv9 still has potential for applications that need to detect and extract signatures; nevertheless, additional optimization could be required to enhance its functionality in realistic, practical situations.

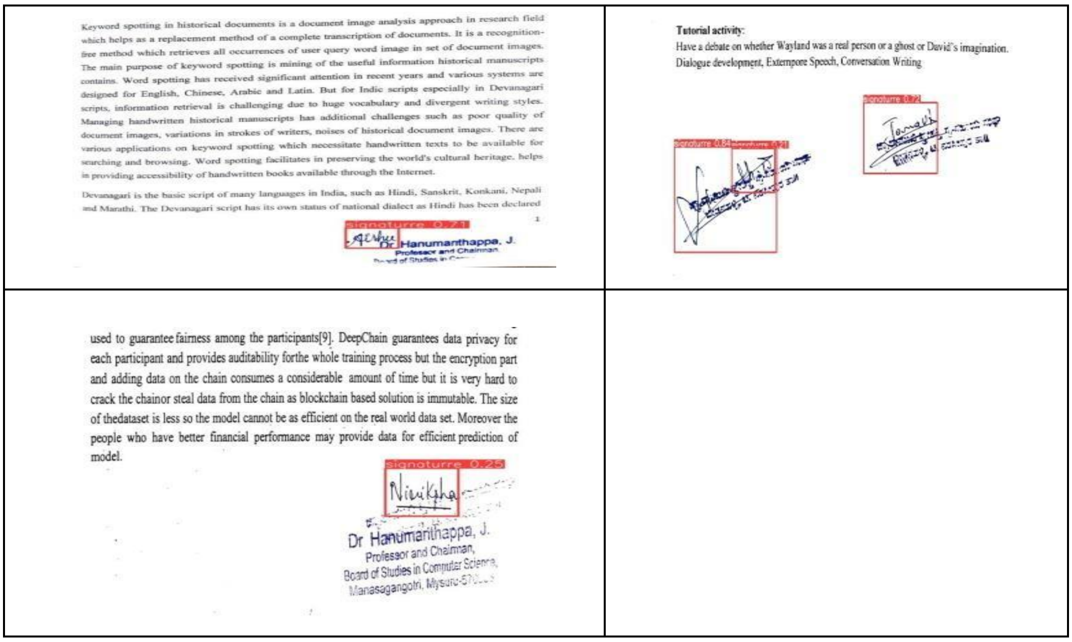


Figure 4: Sample results of signature detection by YOLOv9

- ***YOLOV10:***

Using the same dataset proportion as YOLOv8 and v9, the YOLOv10 model was trained for 50 epochs. The results from the testing phase indicated that signatures were detected very accurately from the documents, YOLOv10 (Figure 5) demonstrated a significant improvement in precision, making it a promising choice for applications requiring signature detection and extraction.

Among all the versions, YOLOv8 performs the best. It accurately detects signatures in documents, including multiple signatures, with high precision.

In contrast, YOLOv9, while detecting all signatures, lacks accuracy, with some documents showing 0.0 accuracy and multiple bounding boxes for a single signature.

YOLOv10 also detects signatures accurately but similarly suffers from multiple bounding boxes for some documents. Therefore, YOLOv8 is the best-performing version. However, among all the models, YOLOv8 performed the most accurately, surpassing the performance of YOLOv9 and YOLOv10 in signature detection and extraction tasks.

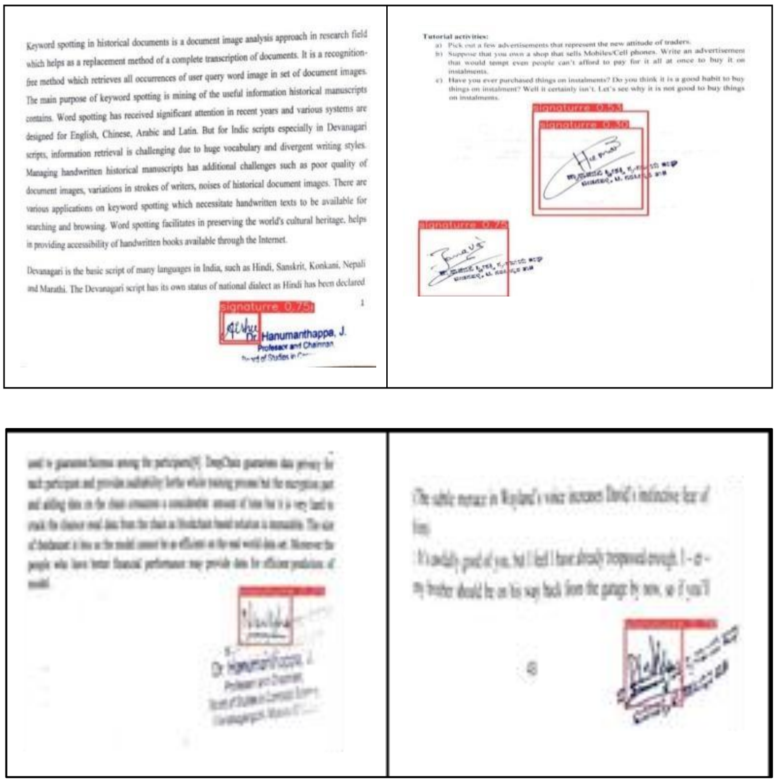
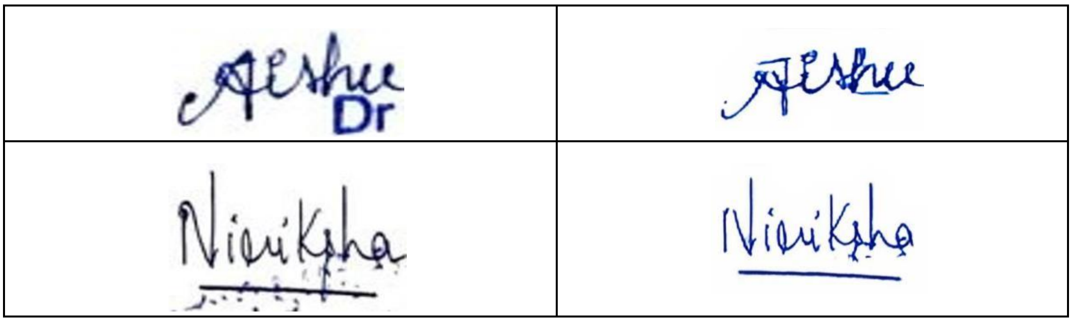


Figure 5: Sample results of signature detection by YOLOv10

A performance indicator called Intersection over Union (IOU) is used to assess how accurate object identification, segmentation, and annotation algorithms are. It measures the amount of overlap between a dataset's ground truth bounding box or annotated region and the anticipated bounding box or segmented region. By measuring the degree to which a projected item matches the actual object annotation IoU is widely used to evaluate the accuracy of object detection models, with common thresholds like 0.5 or 0.75 to determine true positives. Its effectiveness lies in providing a clear, quantifiable measure of how well the predicted objects match the actual objects. Based on IOU values, YOLOv8 was chosen for the next flow. Among all versions, YOLOv8 provided the best IOU values compared to YOLOv9 and YOLOv10.

After detecting the signature, signature region is extracted from the bounding box. Next, detected signature regions are cropped from the documents using the bounding box coordinates. The extracted signatures contain some noise such as seal and date. A noise cleaning method based on CycleGAN will be performed on the extracted signatures to generate noise free signatures. For noise removal and enhancement, a CycleGAN model was trained. The dataset contained noisy signatures as the source domain and corresponding clean signatures as the target domain. By organizing the data appropriately, it was used to train the CycleGAN model. During training, the CycleGAN learned to transform noisy signatures into clean signatures through its generative and discriminative networks. Once the CycleGAN model was trained, it was applied to the cropped signature images to remove noise and enhance their quality (Figure 6).



(a) Signature with noise (b) Noise removed signature

Figure 6: CycleGAN generated signature

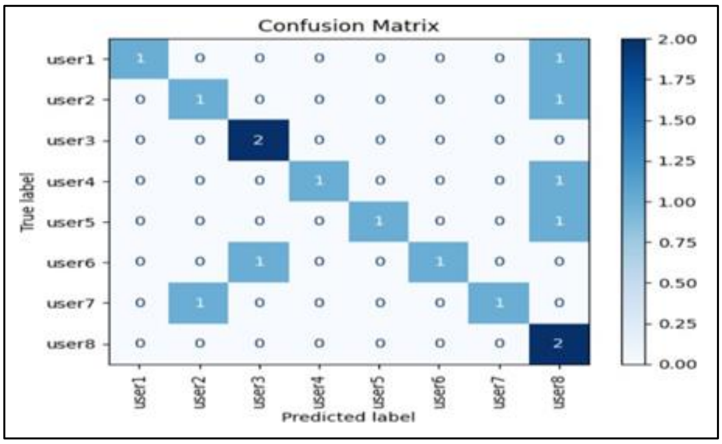


Figure 7: Confusion matrix for GAN generated signatures

After successfully removing noise from the documents, similarity is checked between the cleaned signatures and the corresponding template signatures (Figure 8). Our analysis revealed an impressive overall accuracy rate of 90% for the similarity measure. This high level of accuracy indicates that the cleaned signatures closely match the template signatures, validating the effectiveness of our noise removal and similarity measurement processes. This outcome underscores the reliability of our methods in maintaining the integrity and consistency of signature verification.

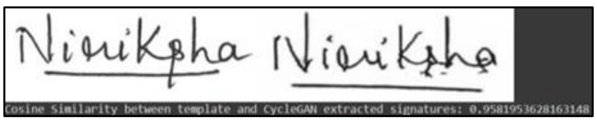


Figure 8: Result of Cosine Similarity

This experiment was also done on tobacco800 dataset and spods dataset even for this dataset yolov8 was giving a promising result on detecting the signatures accurately.

V Conclusions:

In summary this paper introduces an improving signature detection and verification using YOLO models. The proposed method and experimental results show that YOLOv8 is the best model for signature recognition among the three tested versions (YOLOv8, YOLOv9 and YOLOv10). YOLOv8 has demonstrated the highest accuracy and robustness, consistently outperforming other models in terms of accuracy and reliability. This was evident from the Intersection over Union (IoU) values, where YOLOv8 achieved excellent performance and accurately identified multiple signatures with minimal false positives. CycleGAN's integration with de-noising was very effective in cleaning the extracted signatures of unwanted elements such as the stamp and date. A CycleGAN model trained to clean noisy signatures significantly improved the clarity and quality of signatures, which is crucial for accurate authentication. This step ensures that only real signatures are considered compared to model signatures, which improves the reliability of the verification process. The successful application of YOLOv8 in signature detection highlights its potential in real applications that require high accuracy and efficiency. Methods and results confirm the effectiveness of YOLOv8 and CycleGAN in complex signature detection and cleaning tasks, making them very suitable for practical use in signature verification systems.

References:

1. Nabin Sharma, Ranju Mandal and Rabi Sharma "Signature and Logo Detection using Deep CNN for Document Image Retrieval", 2018. (DOI:10.1109/ICFHR2018.2018.00079)
2. Rabi Sharma "Detecting Signatures in Scanned Document Images", 2021
3. Kaihong Yan, Ying Zhang, Haoran Tang, Chengkai Ren, Jian Zhang "Signature Detection, Restoration, and Verification: A Novel Chinese Document Signature Forgery Detection Benchmark", 2022. (DOI:10.1109/CVPRW56347.2022.00564)
4. Venkata Rami Reddy Ch "A Review on YOLOv8 and Its Advancements", 2024. (DOI:10.1007/978-981-99-7962-2_39)
5. Chien-Yao Wang, I-Hau Yeh, Hongpeng Liao "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information", 2024. (DOI: 10.48550/arXiv.2402.13616)
6. Ao Wang, Hui Chen, Lihao Liu, Kai Chen, Zijia Lin, Jungong Han, Guiguang Ding" YOLOv10: Real-Time End-to-End Object Detection,2024. (DOI:10.48550/arXiv.2405.14458)
7. Deniz Engin, Alperen Kantarci and Secil Arslan "Offline Signature Verification on Real-World Documents", 2020. (DOI:10.1109/CVPRW50498.2020.00412)
8. Yongbing Zhang, Siyuan Liu, Chao Dong, Yuan Yuan "Multiple Cycle-in-Cycle Generative Adversarial Networks for Unsupervised Image Super-Resolution", 2020. (DOI: 10.1109/TIP.2019.2938347)
9. Shashidhar Sanda, Sravya Amirisetti "Online Handwritten Signature Verification System using Gaussian Mixture Model and Longest Common Sub-Sequences",2017.
10. Piotr Porwik, Rafal Doroz and Krzysztof Wrobel "A new signature similarity measure", 2009. (DOI:10.1109/NABIC.2009.5393858)
11. Dani Gunawan, C A Sembiring, M A Budiman "The Implementation of Cosine Similarity to Calculate Text Relevance between Two Documents",2018. (DOI:10.1088/1742-6596/978/1/012120).

ISSN: 2582-8118 ***Volume 4, Issue 5; Sept 2024***

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYSIS IN SCIENCE AND ENGINEERING

Web: https://www.iarj.in/index.php/ijrase/index

80. Intrusion Detection Analysis using Deep Learning Models in IoT

Vinayak A. Telsang
Dept. of Computer Science and Engineering
Biluru Gurubasava Mahaswamiji Institute of Technology
Mudhol, India.

Mahabaleshwar S. Kakkasagerim
Dept. of Electronics and Communication Engineering
Basaveshwar Engineering College Bagalkote, India.

Anil D. Devangavi
Dept. of Artificial Intelligence and Machine Learning
Basaveshwar Engineering College, Bagalkote, India.

Abstract:

In the rapidly evolving landscape of the Internet of Things (IoT), maintaining strong security is crucial. IoT devices are integrated into many industries which becomes potential points for security breaches. This security breaches may lead to data loss, impact security of the system, and even disruption of operations causing financial loss. IoT environment connected with these devices faces many security attacks namely: data breaches, man-in-middle (MIM) attack, session key attack, denial of service (DoS) attacks. The machine learning (ML) models can be incorporated into these IoT environment so that it becomes possible to detect these security attacks and provide resilience to IoT systems. Our paper provides the implementation of anomaly detection system through deep learning models such as: Convolutional neural networks (CNN), Recurrent neural networks (RNN), and Long Short-Term Memory (LSTM). Also, by comparing the simulation parameters, we justify LSTM model can be effective in detecting the security attacks.

Keywords:

IoT, security threats, machine learning.

I Introduction:

The Internet of Things (IoT) that connects billions of devices globally and enables seamless communication and data exchange. As IoT continues to proliferate across sectors including healthcare, smart cities and industrial automation, the importance of security in these environments is crucial. IoT devices with their restricted computing capabilities pose security risks that must be addressed to ensure data integrity, confidentiality and availability [1]. Security becomes very important in case of all large electronic industries with any sensitive data lost could be make them worse hence the security policies with partners, suppliers are needed for safe environments across the internet environment [2]. The security threats in IoT environment can have severe consequences leading to potential breaches of privacy and sensitive information. The most common type of attack is Denial of Service (DoS), where devices and networks are flooded with high volumes of traffic, causing service interruptions [3]. The intrusion attack where hidden methods of bypassing normal authentication mechanisms to gain unauthorized access to a system. The malware attack dismantles the IoT devices by executing malicious code and undermines the functionality of the network [4].

The vulnerabilities in communication protocols are exploited by spoofing and eavesdropping attacks to collect sensitive information. Dealing with these security attacks require adoption of sophisticated detection methods in which machine learning (ML) plays a key role. The deep learning models have emerged as the popular models which can used to analyze large amount of data. The supervised algorithms such decision trees and random forests can categorize network activity as either safe or harmful. The unsupervised algorithms such as K-Means can detect abnormalities in behavior, making them crucial for uncovering new attacks [5].

Several methods have been developed to mitigate the attacks based on traditional ML and deep learning algorithms to identify the malicious behavior in the IoT environment. But deep learning methods have been considered for detecting the patterns for detecting attacks in the system [6]. The deep learning models such as CNN, RNN, and LSTM can uncover the temporal data sequences allowing identifying the complex sequences and patterns making them invaluable for detecting previously unseen attacks [7].

There are several algorithms such genetic and decision trees were used to enhance the capability to detect the existing intrusion attacks. The integration of neural networks, RNN's, were suggested by [8] [9] for anomaly detection. CNNs are known for their strong performance in spatial data analysis because they use convolutional layers that apply filters to the input data for pattern recognition. Pools are connected to these layers, which improves the efficiency of the network and preserves important properties. RNN’s maintain a hidden state that stores information from previous inputs, allowing them to consider the context of the data.

Analyzing sequences of network flows can be useful when using RNN technology, but learning long-term patterns can be problematic due to vanishing gradients [10]. In contrast, LSTM’s use a complex representation of memory cells. And gates that allow them to store important data in large sets, making them ideal for tasks involving long-term dependencies, such as detecting trends in IoT security. LSTM can utilize the temporal relationships and store memories for long periods of time to analyze network traffic data and detect potential threats [11] [12]. The rest of the paper is presented as: section II provides related works. Detection of security attacks through ML models is provided in section III. Our proposed deep learning algorithms discussed in section IV. Section V discusses the simulation parameters and results. Section VI presented with conclusion of our paper.

II Related Works:

The network security technique to reduce the network traffic by leveraging neural network by using metrics such as: conciseness, information loss, interestingness, and intelligibility. Finally, the Modified Synergetic Neural Network (MSNN) is proposed on summarized datasets for detecting the real-time anomaly-behaved nodes in network [13]. A decision tree is a classification technique is used in extracting the knowledge from a database suggested as a hybridization strategy. The number of input variables has reduced the computational cost and improved the performance but it suffers from overfitting and local judgments [14].

The open-source elastic stack is proposed for clustering, visualization, and computer-assisted network security analysis. The network anomalies were detected by enabling data clustering and visual mapping techniques to filter the data. A case study is also provided for efficacy of the proposed solution for detecting security attacks [15]. The paper leverages security attacks stored in a database and applies support vector machine algorithm for classifying based on user input. The GUI based feature is provided to determine the phishing attack on the website by utilizing the scalable database technology making it suitable prediction model [16].

The paper presented [17] provided the comparison of different ML algorithms to protect the malicious data injection by using anomaly detection system. The addition of ML techniques helps in improving the intrusion detection in the internet connected things. The network traffic with specific network-layer attacks were implemented by SVM models to detect malicious behavior. Two SVM approaches, the C-SVM and the OC-SVM, were proposed to detect normal, abnormal activity and the latter only for normal behavior activity [18].

III Machine Learning Algorithms for Attack Detection:

Within the field of artificial intelligence (AI), machine learning (ML) is the process of creating statistical models and algorithms that allow computers to carry out tasks without explicit instructions. Rather, these systems use data to learn and make decisions or predictions. ML can be classified as supervised learning where the algorithm is trained with a labeled dataset, and paired with an output label.

The underlying structure of the data is learnt by the algorithm in-case of unsupervised learning. The semi-supervised learning takes the advantages of both supervised and unsupervised learning by using small amount of labeled data and a large amount of unlabeled data for training.

This approach can significantly improve the accuracy. The action-reward based learning happens in reinforcement ML, where an agent makes a decision by performing certain actions and receiving rewards or penalties and learns itself. The main objective is to maximize the cumulative reward over time.

The supervised ML algorithms such as decision tress, random forest, and support vector machines by detecting learned rules, traffic patterns, and by detecting complex patterns from large dataset. The K-means clustering, principal component analysis, and auto encoders of unsupervised learning algorithms group similar patterns from network behavior, can reduce dimensionality by highlighting anomalies, and error construction to indicate anomalies respectively. The reinforcement learning can dynamically handle attacks in cybersecurity environments.

Although traditional ML algorithms are powerful but the deep learning models such as CNN, RNN, and LSTM are well suited for handling different kinds of data and recognizing complex patterns. The CNN model can utilize the convolutional layers to focus on local patterns and detect malicious activities. The RNN model can be used to analyze time-series data such as logs, network usage time, user activities and identify the patterns to indicate attack on system. LSTM is a type of RNN but retain information over long periods. It uses the memory cells to store information, making them highly effective for tasks which needs to store extended sequences of information for long duration.

IV Proposed Machine Learning Algorithms:

We have proposed three deep learning models: CNN, RNN, LSTM for detecting the security attacks in the IoT environment.

The CNN model uses kernels that creates feature maps by sliding over input. Then the pooling layers are applied on feature maps to reduce spatial dimension which reduces computational load. After series of convolutional and pooling layers, it is connected to dense layers for final predictions. We have used Rectified Linear Unit (ReLU) activation function to make network to learn complex patters by introducing non-linearity. The algorithm 1. Provides detail of working of CNN model.

The RNN model works with sequential data where current output of data not depend on current input but also on previous input. RNN uses hidden state and weights are added across all steps making them to train long sequences. The softmax activation function converts the raw output scores of the RNN into a probability distribution over all possible classes, which helps in making decisions based on the highest probability. Algorithm 2. provides the working of RNN model.

LSTM model uses cell state as memory allowing network to maintain long-term dependencies. LSTM uses input gate to get new information of the attack and add it the cell state. The forget gate include the information to be discarded. The output gate based on cell state and current input. The fuzzers generate random inputs to detect the unusual sequence of vulnerabilities by analyzing the sequence of network request over time. The sequence of steps of LSTM model is provided in algorithm 3.

Algorithm 1: CNN Algorithm

BEGIN:

Step 1: Data Preparation

a) Initialize the sequential model
b) Preprocess the data and extract the features suitable for X_{train}, X_{test}.

Step 2: Model Preparation

c) Add CNN model layers with 64 filters and ReLu activation
d) Add Max pooling with pool size=2
e) Optimize the model with 'adam optimzer'

Step 3: Model Testing

f) Train and test the model with epoch =10 and with batch size=64

END:

Algorithm 2: RNN Algorithm

BEGIN:

Step 1: Data Preparation

a) Initialize the sequential model
b) Preprocess the data and extract the features suitable for X_{train}, X_{test}.

Step 2: Model Preparation

c) Add RNN model layers with softmax activation
d) Each layer adder with 100 units

Step 3: Model Testing

e) Train and test the model with epoch =10 and with batch size=64

END:

V Simulation Setting:

The proposed models were simulated using python 3.8, tenser flow and keras for training and testing. We have used NSL-KDD and UNSW-NB15 dataset, which are effective bench marking for detection of attacks on system.

Algorithm 3: LSTM Algorithm

BEGIN:

Step 1: Data Preparation

a) Initialize the sequential model
b) Preprocess the data suitable for LSTM and use as X_{train}, X_{test}.

Step 2: Model Preparation

c) Add LSTM model layers with softmax activation
d) Optimize the model with 'adam optimzer' and cross entropy loss

Step 3: Model Testing

e) Train and test the model with epoch =10 and with batch size=64

END:

A. Simulation Parameters:

We have considered the accuracy, f1-score, precision and recall as our simulation parameters for the proposed machine learning models. The simulation results are provided in table I.

1. ***Accuracy:*** It defines the number correctly predicted attacks in total number of attacks as given in (1).

accuracy= number of correct predictions /total no. of predictions (1)

2. ***Precision:*** It is the ratio of true positive security attack detections to the total number of detections as given in (2).

precision= Correct classified attacks / total no. of predictions (2)

3. ***Recall:*** It is defined as the ratio of actual security attack detected to the total number of security attacks as given in (3). High recall indicates model detects all security attacks.

recall= positively classified attacks / no. of attacks (3)

4. ***F1-Score:*** It measures the balance between precision and recall of a model in detecting malicious activities in the network as given in (4).

A high f1-score indicates that the security system is good at detecting actual attacks.

f1-score= 2 x precision x recall /precision + recall (4)

B. Result Discussion:

LSTM shows higher accuracy, than CNN and RNN which indicates that it is better in classifying the attacks. The RNN has lower precision value indicate more false detection of attacks. CNN has better recall than RNN but lower than LSTM, indicates it detects less actual attack instances. LSTM achieves the highest f1-score, indicating that a good balance between precision and recall. The Receiver Operating Characteristic Curve (ROC) for CNN model is shows slight variation in detecting the attacks, RNN has also shown the detecting actual attacks as compared with CNN. Compared to CNN and RNN, LSTM ROC curve shows less variation in detecting the attacks. As we can also observer the accuracy is also higher for LSTM model as compared with CNN and RNN models in both the dataset used in our simulation.

Table I: Simulation Results of ml Models

Dataset Used	Model	Accuracy	Precision	Recall	F1-Score
NSL-KDD	CNN	92%	91%	90%	90.5%
	RNN	88%	87%	86%	86.5%
	LSTM	94%	93%	92%	92.5%
UNSW-NB15	CNN	90%	89%	88%	88.5%
	RNN	85%	84%	83%	83.5%
	LSTM	92%	91%	90%	90.5%

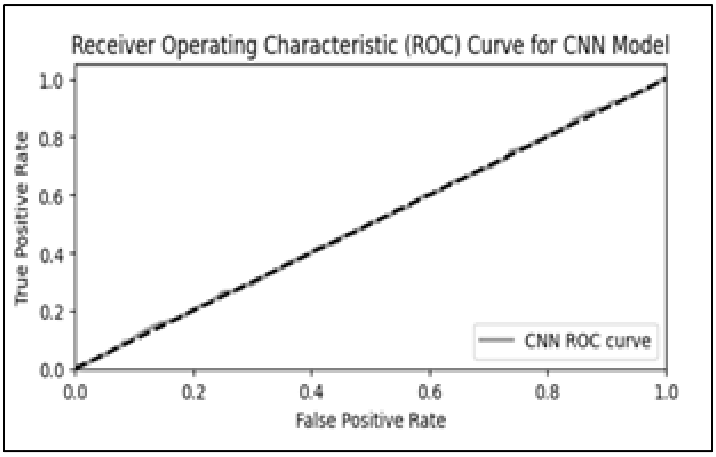


Figure 1: ROC curve of CNN model

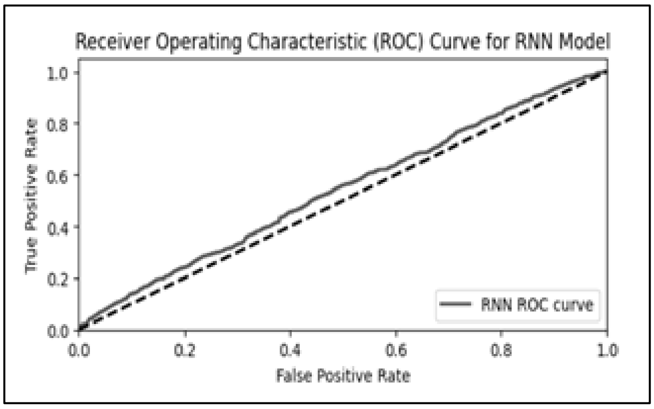


Figure 2: ROC curve of RNN model

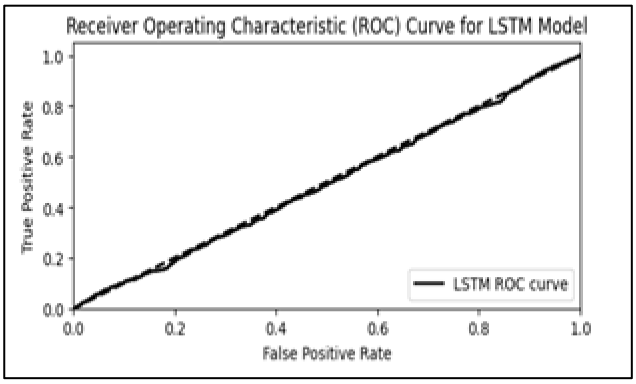


Figure 3: ROC curve for LSTM model

VI Conclusion:

We have provided the detail regarding the necessary of detecting the security attacks in IoT environment. We have also presented details how the traditional ML algorithms are powerful in detecting the attacks but the need of deep learning algorithms is inevitable for handling real-world capabilities. We have proposed the three deep learning models and discussed the results and their performance using two datasets. In result discussion we have observed that the LSTM model performs better and can be used in detecting security attacks. It is also an research challenge to adapt different deep learning models which can help not only detect intrusion attacks but also improve the performance of IoT environment.

References:

1. Khan, N. A., Awang, A., and Karim, S. A. A. "Security in Internet of Things: A review," IEEE access, vol. 10, pp. 104649-104670, 2022
2. Dijesh, P., Babu, S., and Vijayalakshmi, Y., "Enhancement of e-commerce security through asymmetric key algorithm," Computer Communications, 153, pp. 125-134, 2020
3. Saba, T., Rehman, A., Sadad, T., Kolivand, H., and Bahaj, S. A. "Anomaly-based intrusion detection system for IoT networks through deep learning model," Computers and Electrical Engineering, vol. 99, pp. 107810, 2022
4. Cerdà-Alabern, L., Iuhasz, G., and Gemmi, G. "Anomaly detection for fault detection in wireless community networks using machine learning," Computer Communications, vol.202, pp. 191-203, 2023
5. M. Ozkan-Okay et al., "A Comprehensive Survey: Evaluating the Efficiency of Artificial Intelligence and Machine Learning Techniques on Cyber Security Solutions," in IEEE Access, vol. 12, pp. 12229-12256, 2024
6. Saba, T., Rehman, A., Sadad, T., Kolivand, H., and Bahaj, S. A., "Anomaly-based intrusion detection system for IoT networks through deep learning model," Computers and Electrical Engineering, vol. 99, pp. 107810, 2022
7. P. L. S. Jayalaxmi, R. Saha, G. Kumar, M. Conti and T. H. Kim, "Machine and Deep Learning Solutions for Intrusion Detection and Prevention in IoTs: A Survey," in IEEE Access, vol. 10, pp. 121173-121192, 2022

8. H. Debar, M. Becker and D. Siboni., "A neural network component for an intrusion detection system," In Proceedings of the IEEE computer society symposium on research in security and privacy, pp. 240250. IEEE Computer Society, 1992
9. J. Cannady, "Artificial neural networks for misuse detection," In Proceedings of the 1998 national information systems security conference (NISSC), pp. 443456. Citeseer, 1998
10. Surendra Yadav, Hina Hashmi, Daxa Vekariya, Zafar Ali Khan N, and Vijay Fidelis J, "Mitigation of attacks via improved network security in IOT network environment using RNN," Measurement: Sensors, vol. 32, pp. 101046, 2024
11. Imeri and O. Rysavy, "Deep learning for predictive alerting and cyber-attack mitigation," IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), pp. 0476-0481, 2023
12. Adefemi Alimi KO, Ouahada K, Abu-Mahfouz AM, Rimer S, and Alimi OA. "Refined LSTM Based Intrusion Detection for Denial-of-Service Attack in Internet of Things," Journal of Sensor and Actuator Networks, vol. 11, pp. 32, 2022
13. Lalotra, G.S., Kumar, V., Bhatt, A., Chen, T., and Mahmud, M., "iReTADS: An Intelligent Real-Time Anomaly Detection System for Cloud Communications Using Temporal Data Summarization and Neural Network," Secure Communication Networks, pp. 9149164, 2022
14. Alweshah, M., Rababa, L., Ryalat, M.H., Al Momani, A., and Ababneh, M.F., "African Buffalo algorithm: Training the probabilistic neural network to solve classification problems," Journal King Saud University Computing Information Science, vol. 34, pp. 1808–1818, 2020
15. V. Ciric, M. Milosevic, L. Mladenovic and I. Milentijevic, "Clustering and Visualization of Network Security-Related Data using Elastic Stack," 10th International Conference on Electrical, Electronic and Computing Engineering (IcETRAN), pp. 1-5, 2023
16. Shombot, E. S., Dusserre, G., Bestak, R., and Ahmed, N. B., "An application for predicting phishing attacks: A case of implementing a support vector machine learning model," Cyber Security and Applications, vol. 2, pp. 100036, 2024
17. L. Yang, C. Ding, M. Wu, and K. Wang, "Robust detection of false data injection attacks for data aggregation in an Internet of Things-based environmental surveillance," Computer Network, vol. 129, pp. 410–428, 2017.
18. Ioannou C, Vassiliou V., "Network Attack Classification in IoT Using Support Vector Machines," Journal of Sensor and Actuator Networks, vol. 10, pp. 58, 2021

ISSN: 2582-8118 *Volume 4, Issue 5; Sept 2024*

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYSIS IN SCIENCE AND ENGINEERING

Web: https://www.iarj.in/index.php/ijrase/index

81. Precise Forgery Region Identification in Digital Images: A Novel Approach Integrating SIFT, DBSCAN, and K-Means

Rajashekara M., Chethan H. K.
Department of Computer Science and Engineering,
Maharaja Research Foundation,
Maharaja Institute of Technology, Mysuru, India.

Somanna M., Thejashwini B. L.
Department of Studies in Computer Science,
University of Mysore. Manasagangotri, Mysuru. India.

ABSTRACT:

In the present era, copy-move forgery detection in digital images has become a challenging problem for researchers. With software like Adobe Photoshop and GISM, etc., manipulating picture information has become easier.

One common form of manipulation is known as copy-move forgery, where a portion of an image is copied and pasted onto another part of the same image to conceal or alter information.

The challenge here lies in verifying the authenticity of the image after such forgery without access to the original image. In this paper, we propose an approach that identifies interesting key points using the SIFT operator. We select prominent key points, such as core points and border points, by adjusting epsilon values in density-based spatial clustering with noise (DBSCAN) techniques.

Furthermore, to cluster salient key points, we have adapted k-means clustering to group similar points and eventually identify the region where copy-move tampering is being done. This proposed approach can be used in real-time image copy-move forgery authentication.

KEYWORDS:

Copy-move forgery detection, key point based, DBSCAN, K-means clustering, locating forged region.

I Introduction:

In the realm of digital image processing, the act of introducing irregular patterns into original images, known as digital image forgery. This practice has seen a significant rise due to the wide availability and accessibility of advanced image editing tools. Digital image forgery often results in a diverse set of variations in image properties, rendering the distinction between genuine and forged images an increasingly challenging task. The widespread use of user-friendly image editing software has placed the power to manipulate images into the hands of even amateur users. As a consequence, the internet has witnessed a surge in the dissemination of forged images, which can be employed for various deceptive or malicious purposes.

The proliferation of these manipulated images has given rise to a pressing need for effective techniques to detect image forgery. In response to this need, researchers have developed a variety of algorithms and methods aimed at detecting and mitigating image forgery. These approaches generally involve the extraction of unique characteristics and features from digital images, enabling the identification of tampered areas. To better understand these methods, it is crucial to categorize forgery detection techniques into two main types: active and passive. Active forgery detection methods encompass strategies that involve embedding information within the original image. This embedded information can later be used to verify the authenticity of the image, even after it has undergone alterations.

Active methods, while robust, have their own set of requirements and challenges. On the other hand, passive forgery detection methods offer a different perspective. These techniques do not require any prior knowledge of the original image. They rely on the analysis of image features, patterns, and inconsistencies to detect manipulations. Passive methods have gained prominence due to their ability to address image forgeries without relying on embedded information. Within the realm of passive forgery detection, there exist distinct subcategories, each tailored to address specific forms of image manipulation. These subcategories include copy-move forgery detection, splicing detection, retouching detection, and resampling detection. The choice of technique within these categories depends on the specific nature of the image manipulation under investigation.

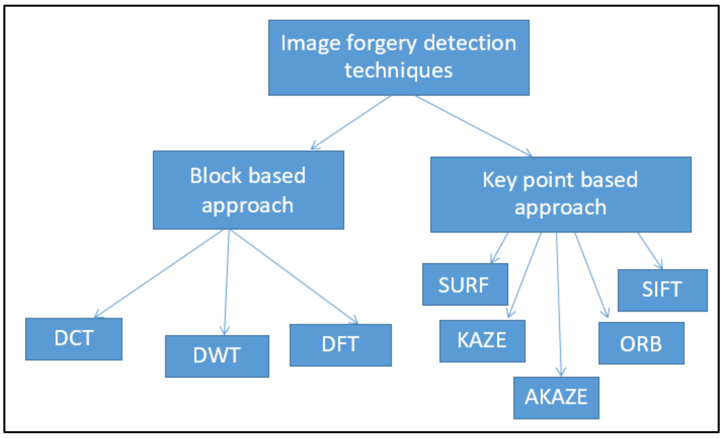


Figure 1: Types of Image Forgery

In this paper, we delve into the intricate landscape of digital image forgery detection, offering an in-depth exploration of passive forgery detection techniques. By examining and categorizing these techniques, we aim to provide valuable insights into the state of the art and the challenges faced in detecting various forms of image manipulation. Furthermore, we present our contributions in this field, highlighting the significance and novel aspects of our research.

II Related Works:

This the detection and mitigation of localizing copy-move forgery pose ongoing challenges in digital images. This review synthesizes recent literature to explore effective strategies and advancements in addressing copy-move forgery detection based on key points, providing a comprehensive overview of existing approaches.

P. Niu et al. (2021) proposed a fast and effective copy-move forgery detection algorithm based on complex-valued moment invariants. The algorithm extracts dense and uniform key points from the entire image, representing them using robust and discriminative moment invariants.

It then proposes a fast magnitude-phase hierarchical matching strategy to match a massive number of key points while maintaining accuracy. A reliable post-processing algorithm is developed to reduce false negative and false positive rates. Extensive experimental results show the superior performance of the proposed scheme compared to existing algorithms, with an average pixel-level F-measure of 94.54% and an average CPU-time of 36.25 son FAU, GRIP, MICC-F600 datasets.

Nitish Kumar et al. (2022) proposed an approach for detecting copy-move forgery in digital images. They extracted Scale-Invariant Feature Transform and KAZE key point features from the input image and selected salient key points to enhance robustness. They also introduced selective search-based region proposals to improve detection accuracy. These key points detect duplicate regions with bounding boxes. The model was tested on MICC-220 and CoMoFoD datasets. The method shows superior precision with 97.90% and faster in processing time.

Guangyu Yue et al. (2022) proposed SMDAF, it is a key point-based method for copy-move forgery detection. This approach combines second-key point matching and double adaptive filtering (SMDAF) to match key points extracted from copy-move forgery images. Experiments shows that this SMDAF method outperforms well with MICC-F220, CASIA, and CoMoFoD datasets.

YıIdız Aydin (2023) proposed a key point-based copy-move forgery detection approach using HSV color space. This approach extracts scale-invariant feature transform (SIFT) and KAZE features from the color channels in the HSV color space instead of greyscale images. The combined features are used in the matching step to detect similar regions. The method was tested on the Image Manipulation Dataset (IMD) and GRIP dataset, showing high performance with F1 score.

Anjali Dinawan et al. (2023) proposed an approach for detecting copy-move forgery in digital images using the Super Point key point detector. This method is capable of detecting various techniques in forged images.

The author combines key point detection with super-point architecture to detect copy-move forgery in digital images. These methods are tested on CMFD, GRIP, MICC-F2000, MICC-F220, CASIA V2.0, CoMoFoD, and COVERAGE datasets. It outperforms well in detection accuracy, with an average of 98.51% F1-score value.

Xiang-yang Wang et al. (2023) proposed an accurate and robust image copy-move forgery detection method using adaptive key points and FQGPCET-GLCM feature. This work addresses the challenges such as unsatisfactory key point distribution, poor in robustness to geometric attacks, and ineffective filtering of false matched pairs in the localization task.

To overcome from these issues, a novel key point-based CMFD method is proposed. This method involves segmenting the input image into non-overlapping blocks using simple linear iterative clustering (SLIC) and K-multiple-means (KMM), extracting dense and uniform key points, combining transform domain features with texture features to obtain robust hybrid features, and performing feature matching using double-bit quantized locally sensitive hash (DBQ-LSH). The experimental results show notably good superior performance on FAU, GRIP, and MICC-F600 datasets with various challenging conditions.

Guiwei Fu et al. (2023) proposed a copy-move forgery detection algorithm based on fused features and density clustering. The algorithm combines speeded up robust features (SURF) and accelerated KAZE (A-KAZE) to extract descriptive features and reduce false positives. The density-based spatial clustering of applications with noise (DBSCAN) algorithm is used to remove mismatched pairs and reduce false positives. The method was tested on Ardizzone and CoMoFoD datasets and demonstrated good detection performance.

Manaf Mohammed Ali Alhaidery et al. (2023) proposed detection and localization technique for copy-move forgery in digital images. This work combines block-based and key point-based techniques to detect and localize similar regions.

The proposed scheme uses SURF-HOG detector and descriptor to detect initial duplicated regions, SLIC segmentation to localize primary matched regions, and Zernike Moments to extract sufficient key points from active regions. The duplicated regions are then classified into authentic or forged regions. This model was tested on GRIP and CoMoFoD datasets with 93.75% F1-score.

Soumya Mukherjee et al. (2024) proposed a copy-move forgery detection technique using DBSCAN-based key point similarity matching. The authors developed an AKAZE-driven key point-based forgery detection technique, which extracts translation invariant features from the LL sub-band of a SWT-transformed images. This algorithm effectively locates the forged region. The method was tested on GRIP, MICC-220, and CoMoFoD datasets with post-processing techniques. This technique is crucial for detecting and localizing forged areas, which are difficult to detect and localize when subjected to post-processing and geometrical attacks.

Mayank Verma et al. (2024) conducted a survey on copy-move forgery detection, highlighting the importance of preserving image integrity in the digital era. They reviewed various copy-move forgery detection (CMFD) schemes, their pros and cons, performance evaluation criteria, and image datasets used.

The survey also discussed the limitations of block-based CMFD methods, which are not robust to geometric transformations, and the increasing popularity of deep learning techniques in the field. The study also highlighted existing research gaps in the field and aimed to provide researchers with a broad perspective on advancements in CMFD techniques. The survey aims to help researchers improve detection with new opinions and challenges with all benchmark datasets.

J. S. Sujin et al. (2024) propose a high-performance image forgery detection algorithm using adaptive Scale-Invariant Feature Transform (ADSIFT) feature extraction. This method is particularly effective in low-contrast, small, or smooth copy-move region images. By adjusting the contrast threshold and scaling the image, the authors worked on adequate number of key points, even in low-contrast or small regions.

The proposed method also uses an iterative localization scheme to reduce false alarm rates and locate copied areas. The experiments were conducted on MICC-F220, MICC-F600, and CMH-GRIP and COVERAGE datasets with good localization results.

Cong Lin et al. (2024) proposed a copy-move forgery detection scheme based on Regional Density Center (RDC) clustering and Refined Length Homogeneity Filtering (RLHF) policy.

The method employs scale normalization, gray value grouping, RLHF policy, RDC clustering algorithm, and correlation coefficients to identify duplicated regions in digital images. This proposed work was experimented on FAU, GRIP, and MICC-F600 datasets with good localization.

Recent advancements in copy-move forgery detection have been made through novel algorithms and rigorous validations across different datasets. Techniques like key point-based methods like SIFT and KAZE, DBSCAN-based similarity matching, and adaptive feature extraction have improved detection accuracy and reliability.

Integrating features, density clustering, and hierarchical matching strategies have reduced false positives and improved overall performance metrics. Moving forward, continued research efforts are poised to refine existing methodologies and explore new frontiers in image forensics, ensuring robust protection against digital image tampering in diverse applications.

III Proposed Workflow:

The workflow of the proposed model is shown in Figure 1, which includes detection of key points, and clustering of prominent key points followed by k- means clustering to identify forged regions in digital images.

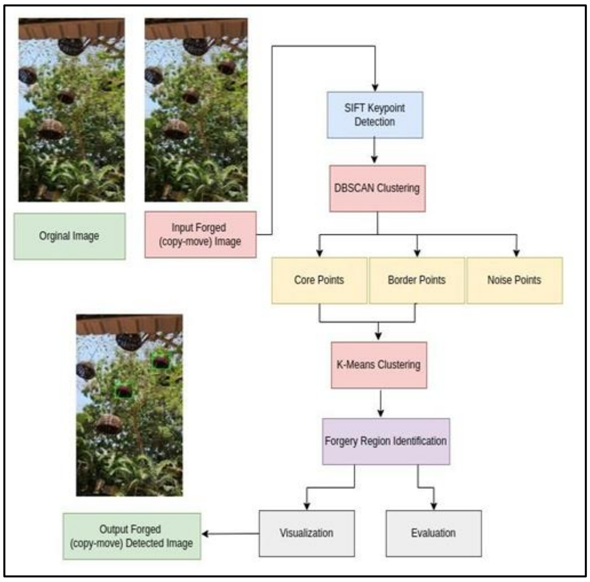


Figure 1: Proposed Workflow

1. Loading Images:

The first step is to convert the input image to color mode. It is important to preserve the original color information, helping to clearly and visually identify fakes.

2. Defining a SIFT Key Point:

We use the Invariant Shape Transform (SIFT) algorithm to detect points in the input image and calculate features. SIFT was chosen because it is robust to size and variability, making it very effective in detecting various features in images.

Scale-space peak selection:

$L(x,y,s)=G(x,y,s)*I(x,y)$L(x,y,s)=G(x,y,s)∗I(x,y)

Key point Localization:

$(x)=D+\partial D\partial xx+12xT\partial 2D\partial x2x$D(x)=D+∂x∂Dx+21xT∂x2∂2Dx

Orientation Assignment:

$\theta(x)=\text{argmax}(L(x)*G(x,\sigma))$θ(x)=argmax(L(x)∗G(x,σ))

Key point Descriptor:

$D(x)=\{d1,d2,...,dn\}$D(x)={d1,d2,...,dn} where didi is a high-dimensional vector.

3. DBSCAN Cluster:

We use the density-based spatial clustering of noise (DBSCAN) algorithm to cluster the SIFT-derived details. DBSCAN is useful because of its ability to find any cluster and effectively control noise. We check different values of neighborhood radius and minimum number of clusters to generate. We evaluate different distance measures such as "Euclidean" and "Manhattan" to determine the best cluster structure.

Core point: If any point Xi has at least K points in its neighborhood, where K is the total number of samples within the radius Ɛ (epsilon), K is automatically set by the algorithm according to the complexity of the images.

Border point: If any point Xi is not satisfied $Xi \in K$ and it contains at least one core point within the radius Ɛ(epsilon). This point is considered as border point.

Noise point: If point Xi is not satisfied by the conditions of the core point and border point, such points are referred to as noise points.

4. Identifying The Area of Counterfeiting:

After clustering is complete, we analyze the results to identify potential false positives. This includes distinguishing between cores (points within boundary regions) and endpoints (points at the edge of a region), as well as identifying noise sources (points that do not belong to any cluster).

5. K-Means Summary and Points and Limits:

To improve the fault localization, we apply K-Means clustering method in core and extreme regions detected by DBSCAN. We test different values of "k" (number of clusters) to determine the number of clusters that best represent the mock region.

6. Monitoring and Evaluation:

6.1 Fake Area Monitoring: The final step involves visualizing the identified false regions. Draw a rectangle around the area determined based on the K-Means clustering results. This visual cue will help you easily identify where the characters in the image were created.

6.2 Performance Indicators: We use metrics such as precision, recall, and F1 score to evaluate the performance of the proposed method. These indicators help report the accuracy and reliability of polygraph systems.

6.3 Comparative Analysis:

We compare the effectiveness of our method with other state-of-the-art lie detection methods. This evaluation provides insight into the strengths and limitations of our approach, as well as areas for improvement.

In summary, our approach consists of a multi-step process, first using SIFT for feature detection, then using DBSCAN for feature extraction, post-transformation using K-means clustering, and finally analytical evaluation of identified false positives.

This comprehensive approach allows reliable and accurate detection of false positives in images.

IV Methodology:

1. SIFT Key Point Detection and Description:

We start with an input image. The Scale-Invariant Feature Transform (SIFT) algorithm is used to detect key points and generate their corresponding descriptors.

Key points K: $K=\{k_1, k_2, \ldots, k_n\}$, where each key point k_i is defined by its location (x,y), scale σ, and orientation θ.

Descriptors D: $D=\{d_1, d_2, \ldots, d_n\}$, where each descriptor di is a 128-dimensional vector. These key points and descriptors capture distinctive features of the image, making them useful for identifying potential forgeries.

2. DBSCAN Clustering:

Next, we apply the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm to the set of descriptors D. DBSCAN groups the descriptors into clusters based on two parameters: the neighborhood radius ε and the minimum number of points to form a cluster minPts.

Clustering C:C=DBSCAN(D,ε,minPts), where C is the set of clusters $\{C_1, C_2, \ldots, C_m\}$, and each cluster C_i is a subset of descriptors D.

DBSCAN effectively groups similar key points while identifying noise points, which helps in locating areas with potential forgery.

3. K-Means Clustering:

After identifying clusters with DBSCAN, we focus on the core and border points of these clusters for further refinement using the K-Means algorithm.

Core and Border Points: Let C_{core} be the set of core points and C_{border} be the set of border points identified by DBSCAN.

Refined Clustering R: R=K-Means ($C_{core} \cup C_{border}$,k), where R is the set of refined clusters $\{R_1, R_2, \ldots, R_k\}$, and k is the number of clusters. K-Means further groups the core and border points into more precise clusters, enhancing the localization of forgery regions.

4. Forgery Region Identification:

For each refined cluster R_i, we compute a bounding box that encloses the key points in that cluster. This bounding box represents a potential forgery region.

Bounding Box Bi: Bi =(xmin ,ymin ,xmax ,ymax), where (xmin ,ymin) is the top-left corner and (xmax ,ymax) is the bottom-right corner of the bounding box.These bounding boxes visually highlight the areas in the image that are likely to be tampered with, making it easier to detect and analyze forgeries.

V Experimental Result:

The outcome is the ultimate measurement of any method or experiment. In the copy-move forgery detection technique, obtaining proper outcomes requires flawless algorithms as well as benchmark datasets which ensure the effectiveness of the designed model. Most of the work on copy-move forgery detection in the literature has focused on benchmark datasets such as CASIA v1, CASIA v2, and MICC.

While these approaches perform well on benchmark datasets, they often struggle with real-world images. To address this limitation, we aimed to develop a generalized approach that would be effective across a wide variety of images, regardless of their source or content.

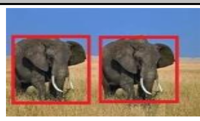
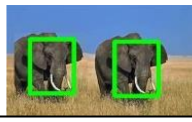
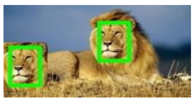
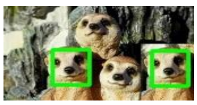
Original Image	Copy-move Forged Image	Ground Truth Region	Detected Forged Region
			
			
			
			
			
			

Table 1: Results of Proposed Methodology

For our experiments, we manually performed copy-move forgery on images captured by cameras and sourced from the internet. Our goal was to create a more versatile forgery detection algorithm capable of identifying forged regions in any image, not just those from benchmark datasets. We successfully achieved this objective and obtained satisfactory results with our generalized approach. We have experimented on 1000 image samples on our proposed model. The results are discussed in Table 1.

Performance Evaluation:

Explanation of the Intersection over Union (IoU) Evaluation Metric:

Definition: IoU is a metric used to evaluate the accuracy of an object detection or segmentation model. It measures the overlap between the predicted bounding box (or segmentation area) and the ground truth bounding box (or area).

Mathematical Model:

Let A be the area of the predicted bounding box.

Let B be the area of the ground truth bounding box.

The intersection of A and B is the area where both boxes overlap. The union of A and B is the total area covered by both boxes, minus the area of the intersection to avoid double-counting.

IoU = Area of Union / Area of Intersection

If Area of Intersection=0, IoU is 0.

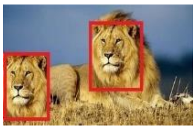
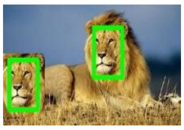
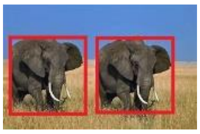
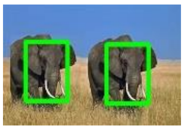
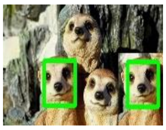
If the predicted and ground truth bounding boxes perfectly overlap, IoU is 1.

To calculate the IoU for the bounding boxes provided, you'll need to follow these steps:

Convert the bounding box coordinates from relative to absolute (if necessary): Since the coordinates are given as relative values (normalized), you can calculate the IoU directly.

Table 2: IoU Performance Evaluation comparison for Original and Copy- Move Images

Sl. No.	Original	Copy- Move	IoU Original with Original	IoU Copy- Move with Copy- Move
1			0.2034	0.1886

Sl. No.	Original	Copy- Move	IoU Original with Original	IoU Copy- Move with Copy- Move
2			0.2520	0.2578
3			0.5036	0.6914
4			0.6209	0.5602
5			0.5409	0.5796
6			0.2677	0.2709

To calculate IoU, following steps have been adopted. The format for the bounding box is x_{center}, y_{center}, width, height.

Step 1: Convert to Corner Coordinates

Original Bounding Box (original):

x_{center}=0.159375, y_{center}=0.391667, width=0.250000, height=0.683333

Predicted Bounding Box (predicted):

x_{center}=0.195312, y_{center}=0.444444, width=0.096875, height=0.444444

Convert to corner coordinates using:

Top-left corner:

$(x_{min}, y_{min}) = (X_{center} - \frac{width}{2}, Y_{center} - \frac{height}{2})$

Bottom-right corner

$(x_{max}, y_{max}) = \left(X_{center} + \frac{width}{2}, Y_{center} + \frac{height}{2}\right)$

Original Bounding Box:

(x_{min}, y_{min}) = (0.034375, 0.050000)

(x_{max}, y_{max}) = (0.284375, 0.733334)

Predicted Bounding Box:

(x_{min}, y_{min}) = (0.146875, 0.222222)

(x_{max}, y_{max}) = (0.243750, 0.666667)

Step 2: Calculate the Area of Intersection

The area of intersection is the overlap between the two bounding boxes.

It is calculated as:

Intersection Width = max (0, min ($x_{max,\ original}$, $x_{max,\ predicted}$) − max ($x_{min,\ original}$, $x_{min,\ predicted}$)) = 0.096875

Intersection Height = max (0, min ($y_{max,\ original}$, $y_{max,\ predicted}$) − max ($y_{min,\ original}$, $y_{min,\ predicted}$)) = 444444

Intersection Area = Intersection Width × Intersection Height = 0.0431 (approx.)

Step 3: Calculate the Area of Each Bounding Box

The area of each bounding box is calculated as:

Area=Width × Height

Original Bounding Box Area:

$A_{original}$=0.250000×0.683333=0.1708 (approx.)

Predicted Bounding Box Area:

$A_{predicted}$=0.096875×0.444444=0.0431 (approx.)

Step 4: Calculate the Union Area

The Union Area is the combined area covered by both bounding boxes, without double-counting the intersection:

Union Area=$A_{original}+A_{predicted}-A_{intersection}$=0.1708

Step 5: Calculate the IoU (Intersection over Union)

Finally, the IoU is calculated as:

$$IoU=\frac{A_{intersection}}{Union\ Area}=\frac{0.0431}{0.1708}\approx 0.252$$

The IoU between the original and predicted bounding boxes is approximately **0.252**.

The above-mentioned steps are used to evaluate the performance of the various images.

VI Conclusion:

In this paper, we have focused on key-point based copy-move forgery detection techniques. To demonstrate our experiments, we have combined techniques such as SIFT, DBSCAN, and K-means clustering to detect the forged areas precisely. To conduct our experiments, we used images from internet sources. The experimental results show that the proposed method gives promising results for all real-time probe images. While our proposed method is robust and efficient, it has some limitations. Although the method successfully detects most copy-move forged regions, it struggles with regions that have scaling variants. Additionally, since our work relies on unsupervised learning principles, a deep learning-based approach remains unexplored. Future research can focus on this area to further advance the intricacies of image forgery detection.

References:

1. Anjali Diwan, Kumar D, Mahadeva R, Perera HCS, Alawatugoda J (2023) Unveiling copy-move forgeries: Enhancing detection with Super Point key point architecture. IEEE Access, 11:86132-86148. https://doi.org/10.1109/ACCESS.2023.3304728.
2. Cong Lin, Wu Y, Huang K, Yang H, Deng Y, Wen Y (2024) Copy-move forgery detection using Regional Density Center clustering. Journal of Visual Communication and Image Representation, 103:104221. https://doi.org/10.1016/j.jvcir.2024.104221.
3. Guangyu Yue, Duan Q, Liu R, Peng W, Liao Y, Liu J (2022) Smdaf: A novel key point-based method for copy-move forgery detection. IET Image Processing, 16(13):3589-3602. https://doi.org/10.1049/ipr2.12578.
4. Guiwei Fu, Zhang Y, Wang Y (2023) Image copy-move forgery detection based on fused features and density clustering. Applied Sciences, 13:7528. https://doi.org/10.3390/app13137528.

5. J. S. Sujin, Sophia S (2024) High-performance image forgery detection via adaptive SIFT feature extraction for low-contrast or small or smooth copy–move region images. Soft Comput 28, 437–445. https://doi.org/10.1007/s00500-023-08209-6.
6. Manaf Mohammed Ali Alhaidery, Amir Hossein Taherinia, Haider Ismael Shahadi (2023) A robust detection and localization technique for copy-move forgery in digital images, Journal of King Saud University - Computer and Information Sciences, Volume 35, Issue 1, Pages 449-461, ISSN 1319-1578. https://doi.org/10.1016/j.jksuci.2022.12.014.
7. Mayank Verma, Singh D (2024) Survey on image copy-move forgery detection. Multimedia Tools and Applications, 83:23761–23797. https://doi.org/10.1007/s11042-023-16455-x.
8. Nitish Kumar, Meenpal T (2022) Salient key point-based copy–move image forgery detection. Australian Journal of Forensic Sciences, 55(3):331–354. https://doi.org/10.1080/00450618.2021.2016964.
9. P. Niu, Wang C, Chen W, Yang H, Wang X (2021) Fast and effective Key point-based image copy-move forgery detection using complex-valued moment invariants. Journal of Visual Communication and Image Representation, 77:103068. https://doi.org/10.1016/j.jvcir.2021.103068.
10. Rajashekara M, Chethan H K, Thejashwini B L, Somanna M, Manjuchethan N (2023) Forgery Detection in Digital Images using Transfer Learning Approaches. ICRRACS-2023.
11. Soumya Mukherjee, Pal AK, Maji S (2024) A copy-move forgery detection technique using DBSCAN-based key point similarity matching. International Journal of Machine Learning and Cybernetics. https://doi.org/10.1007/s13042-024-02268-3.
12. Xiang-yang Wang, Wang X, Niu P, Yang H (2023) Accurate and robust image copy-move forgery detection using adaptive key points and FQGPCET-GLCM feature. Multimedia Tools and Applications. https://doi.org/10.1007/s11042-023-15499-3.
13. YıIdız Aydin (2023) A comparative analysis of skin cancer detection applications using histogram-based local descriptors. Diagnostics, 13:3142. https://doi.org/10.3390/diagnostics13193142.

ISSN: 2582-8118 *Volume 4, Issue 5; Sept 2024*

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYSIS IN SCIENCE AND ENGINEERING

Web: https://www.iarj.in/index.php/ijrase/index

82. Smartsight: Android-Based Retinal Disease Detection Using Retinal Fundus Images with Xai

Dr. B. M. Reshmi, Prof. Nagaraj Kalligud,
Krutika Badiger, Manisha Bagewadi

Department of Artificial Intelligence and Machine Learning
Basaveshwar Engineering College Bagalkot, India.

Abstract:

Diabetic retinopathy (DR) is a serious eye condition affecting diabetic patients, stemming from microvascular damage to retinal blood vessels. If not diagnosed and treated early, DR can lead to blindness. Traditional diagnostic methods involve manual examination of retinal images by ophthalmologists, which is labor-intensive and time-consuming. This highlights the need for automated, efficient, and accurate diagnostic tools. This study proposes a deep learning-based framework using ResNet50 architecture enhanced with Gradient-weighted Class Activation Mapping (Grad-CAM) for the automatic detection and classification of diabetic retinopathy from retinal images obtained via a smartphone-based fundus camera. The ResNet50 model, known for its deep residual learning capabilities, was trained on a dataset of retinal images to differentiate between healthy and DR-affected eyes. Incorporating Grad-CAM provides visual explanations by highlighting the retinal image regions most indicative of DR, thereby improving model interpretability. The proposed model demonstrated high performance, achieving 98% accuracy on the training dataset, 91% on the validation dataset, and 96% on the test dataset. These results highlight the model's robustness and potential for facilitating early DR detection, especially in resource-constrained environments with limited access to specialized ophthalmic care.

KEYWORDS:

Diabetic Retinopathy, ResNet50, XAI, Grad CAM.

I. Introduction:

Diabetic Retinopathy (DR) arises from the bleeding of small blood vessels in the retina. If untreated, this condition can lead to varying degrees of vision impairment and potentially result in blindness. In the United States, more than 4.4 million individuals aged 40 and older experience DR at different stages. Due to its often-asymptomatic progression, DR might cause no noticeable symptoms or only minor vision issues initially. Early detection is crucial, as it significantly improves the effectiveness of treatments to prevent blindness. Therefore, annual eye examinations are recommended for diabetic patients. During retinal examinations, doctors utilize fundus cameras to identify eye diseases through digital imaging. However, evaluating these images manually is time-consuming, with reports potentially taking several days to complete. Additionally, fundus cameras are bulky, expensive, and not easily transportable, limiting their availability in many health clinics

To address these challenges, our project integrates ResNet50 and Explainable AI (XAI) into a smartphone-based system, aiming to make DR screening more accessible and efficient, especially in resource-limited areas.

This innovative approach supports early detection and treatment, providing healthcare providers with a reliable tool that enhances diagnostic accuracy and patient outcomes. By leveraging advanced AI and smartphone technology, our initiative seeks to bridge the gap in diabetic eye care, ultimately reducing the incidence of blindness caused by DR.

Figures 1 and 2 offer visual insights into the progression of diabetic retinopathy (DR) and its impact on the retina. Figure 1(a) depicts a healthy retina, characterized by clearly defined blood vessels and an intact optic disc, indicative of normal retinal function. In contrast, Figure 1(b) presents a retina affected by diabetic retinopathy, where pathological changes such as micro aneurysms, hemorrhages, and lipid exudates are evident, signaling the onset and progression of the disease.

Figure 2 provides a more detailed representation of the stages of diabetic retinopathy. It illustrates the transformation from a normal retina to one experiencing proliferative diabetic retinopathy (PDR), underscoring the escalating damage and potential for vision loss. Specifically, Figure 2(a) shows a normal retina, free of any diabetic changes. Figure 2(b) represents mild non-proliferative diabetic retinopathy (NPDR), marked by the presence of small aneurysms.

Figure 2(c) shows moderate NPDR, where hemorrhages and micro aneurysms become more apparent. Figure 2(d) depicts severe NPDR, with more significant vascular abnormalities and retinal damage. Finally, Figure 2(e) illustrates proliferative diabetic retinopathy, the most advanced stage, characterized by the proliferation of new blood vessels on the retinal surface, leading to substantial vision impairment if left untreated.

These figures collectively emphasize the critical importance of early detection and intervention in managing diabetic retinopathy to prevent severe visual outcomes.

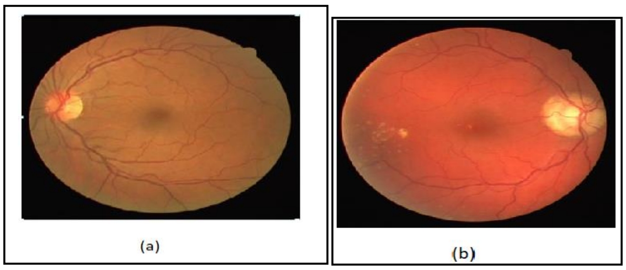


Figure 1: (a) Healthy Retinal Image, (b) Retinal Image Affected by DR

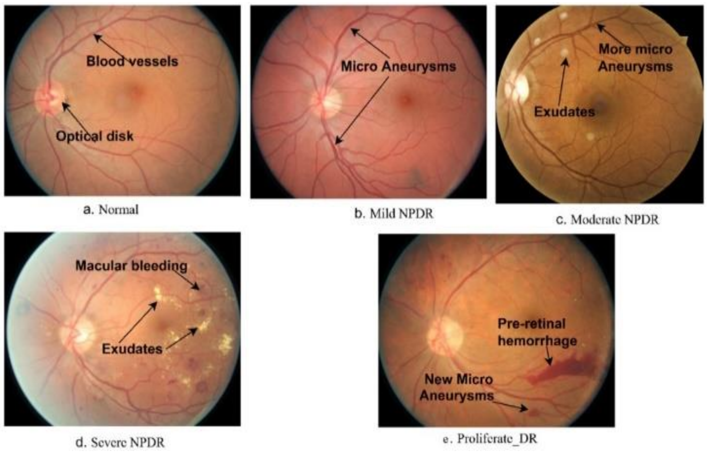


Figure 2: Stages of Diabetic Retinopathy(a) Normal, (b) Mild NPDR, (c) Moderate NPDR, (d) Severe NPDR, (e) Proliferative DR.

II. Literature Review:

In the field of Diabetic Retinopathy detection, several research papers have contributed to the development of different methodologies and techniques. The following literature survey summarizes the key findings and approaches discussed in selected papers.

In 2023, Recep Emre Hacisoftaoglu [1] proposed the effectiveness of deep learning techniques in detecting diabetic retinopathy (DR) using smartphone-based retinal imaging systems. It compares various smartphone ophthalmoscope systems like iExaminer, D-Eye, Peek Retina, and iNview in terms of field of view (FoV) and image quality.

The study develops DR detection algorithms using AlexNet and GoogLeNet, and evaluates their performance with synthetic images mimicking smartphone-captured images. Key findings reveal that the iNview system provides the best image quality and largest FoV, leading to higher DR detection accuracy. Challenges faced include reduced detection performance with smaller FoVs and the need for clinical validation of smartphone-based systems to ensure reliability in real-world applications. Despite these issues, the study supports the potential of these portable systems for DR screening, particularly in resource-limited settings.

In 2022, Mohamed A. Berbar [2] addresses the challenge of detecting and grading diabetic retinopathy (DR), a complication of diabetes that can lead to blindness if not detected early. DR detection is difficult due to the poor quality of fundus images, which often suffer from under saturation and noise.

The authors propose a new image enhancement process and two convolutional neural network (CNN) architectures to tackle this. The first architecture is a binary classifier that distinguishes between normal and abnormal images, while the second classifies the severity of DR into five grades. The study compares the performance of newly defined CNN models with pre-trained models using datasets from Messidor1, Messidor2, and Kaggle EyePACS.

The proposed methods demonstrated high reliability and efficiency, with significant improvements in accuracy and area under the curve (AUC) after preprocessing. However, the issues faced include image quality variability and the need for extensive preprocessing to achieve high accuracy.

In 2022, Chiranthana M, Suhas Gummalam, and Tharun Sekaar [3] proposed Detection of Retinal Diseases using Smartphone Fundoscopy and Deep Learning. The paper addresses the increasing prevalence of diabetic retinopathy and glaucoma, leading causes of blindness, particularly among individuals with diabetes. Traditional retinal examinations require regular visits to specialists, which is impractical for many. The authors propose a low-cost, accessible solution involving a smartphone-based fundoscopy device and a deep learning model for retinal disease diagnosis. Their system allows users to capture retinal images using a smartphone, which are then uploaded to a web application for analysis by a deep learning model. This model diagnoses conditions such as diabetic retinopathy and glaucoma with high accuracy, providing a timely alternative to in-person examinations. The main challenges highlighted include the limited accessibility to traditional ophthalmoscopes and the need for frequent, regular eye exams for early detection and treatment of retinal diseases. The proposed system aims to overcome these barriers by offering a portable, affordable, and efficient diagnostic tool.

In 2023, Vikas Shenoy Pete [4] published paper titled "XAI-Driven CNN for Diabetic Retinopathy Detection". This research focuses on developing a Convolutional Neural Network (CNN) to aid in the early detection of diabetic retinopathy through eye images. The project integrates explainable AI (XAI) techniques, specifically LIME (Local Interpretable Model-agnostic Explanations), to enhance the interpretability and transparency of the model's decision-making process. The CNN model achieved a high accuracy rate of 91%, demonstrating its potential for clinical application and improving healthcare diagnostics by providing clear insights into its predictions.

In 2020, Revathy R, Nithya B S, Reshma J J, Ragendhu S S, and Sumithra M D [5] discussed the challenges of diagnosing diabetic retinopathy, a severe complication of diabetes that can lead to complete blindness if not treated promptly.

The main issues faced in detecting this condition include the need for early and accurate diagnosis to prevent severe visual impairment and the time-consuming nature of manual detection by ophthalmologists. The paper proposes an automated system using a hybrid machine learning model that combines various classifiers such as SVM, KNN, Random Forest, Logistic Regression, and Multilayer Perceptron Network to improve detection efficiency. However, achieving high accuracy and precision in identifying features like exudates, hemorrhages, and micro aneurysms from retinal images remains a challenge. The highest accuracy achieved by their system is 82%, indicating that there is still room for improvement in the detection methods to enhance reliability and reduce the risk of misdiagnosis.

By exploring these papers, researchers and practitioners can gain insights into the methodologies and techniques employed for Diabetic Retinopathy detection, providing a foundation for further advancements in this field.

III. Dataset and Methodology:

A. Dataset:

The MESSIDOR-2 dataset, available on Kaggle for Diabetic Retinopathy Detection, was utilized in this study. This dataset, intended for research purposes, aids in developing and accessing algorithms for the automatic detection of diabetic retinopathy. MESSIDOR-2 comprises retinal fundus images obtained via fundus photography, showcasing a range of diabetic retinopathy stages from normal to severe. The dataset contains a total of 5750 images, divided into 2000 training images, 1750 validation images, and 2000 testing images. Table 1 provides number of training, testing and validation images. Figure 3 shows the distribution of images across dataset splits.

Classes	Test images	Train images	Validation images
Normal images	400	400	360
Mild images	400	400	360
Modarate images	400	400	360
Severe images	400	400	360
Poliferative images	400	400	360

Table 1: No. Of Test, Train and Validation Images.

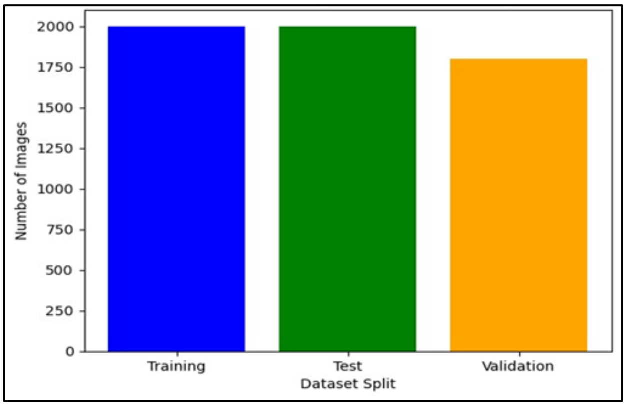


Figure 3: Distribution of Training, Testing and Validation Images.

B. Methodology:

The proposed methodology involves preprocessing fundus images to detect diabetic retinopathy, implementing ResNet50 with transfer learning, and utilizing Grad-CAM for visualization. The training process uses Adam optimization and sparse categorical cross-entropy loss for evaluation. The model is deployed via Gradio for interactive web-based access. Figure 4 shows the working process of proposed methodology.

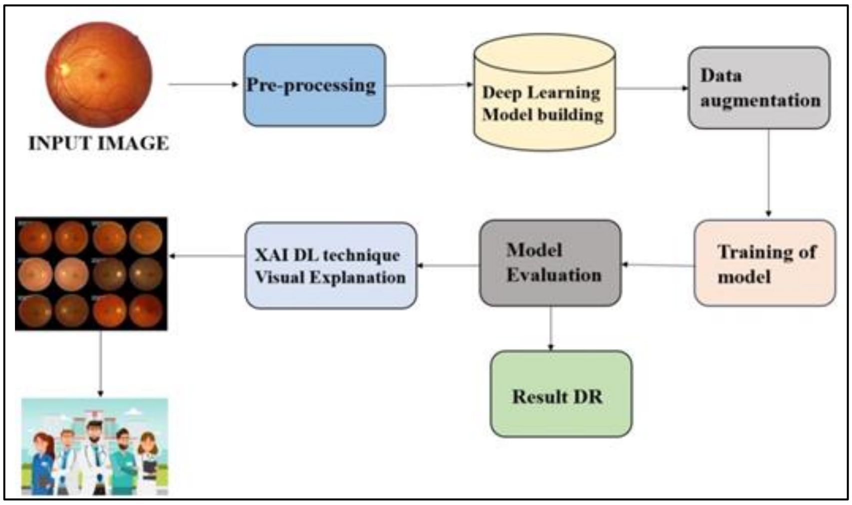


Figure 4: Working Process of the Proposed Methodology

The steps of the proposed methodology are as follows:

Step 1: Dataset Preparation:

Data Source: The Messidor dataset is organized into three directories: train, validation, and test.

Pre-processing: Images are loaded, resized to 224x224, and converted to RGB format.

Step 2: Model Architecture:

Base Model: Utilizes ResNet50 with transfer learning, pre-trained on ImageNet, with the top layer removed.

Fine-tuning: The last few layers of ResNet50 are unfrozen for fine-tuning.

Additional Layers: A flattened layer followed by dense layers with ReLU activation and dropout for regularization is added.

Output Layer: A dense layer with SoftMax activation for multi-class classification is used.

Table 2: Provides No. of Resnet50 Layers Used to Train the Model

Base Model	Total Parameters	Trainable Parameters	Non-Trainable Parameters
ResNet50	75511317	75458197	53120

Table 2: ResNet50 Parameters.

Step 3: Training

Optimization: Adam optimizer with a learning rate of 1e-5 is used.

Loss Function: Sparse categorical cross-entropy loss for multi-class classification.

Metrics: Accuracy is used as the evaluation metric.

Data Augmentation: Keras Image Data Generator is used for training data augmentation.

Step 4: Evaluation Test Set Evaluation: Model performance is evaluated on the test set after training.

Step 5: XAI Visualization Techniques Grad-CAM is implemented for visual explanations. Grad-CAM highlights important regions in the input image relevant to the predictions, enhancing the understanding of the model's focus areas. Figure 5 shows the output of Grad-CAM predicted by the model.

Working of Grad-CAM:

1. Forward Pass: Input the image into ResNet50 and perform a forward pass to get the predicted class scores.
2. Identify the Target Class: Select the class for generating the Grad-CAM visualization, usually the one with the highest predicted score.
3. Backward Pass: Compute the gradients of the predicted class score concerning the feature maps of the last convolutional layer.
4. Compute Weights: Average the gradients obtained in the backward pass across the spatial dimensions to calculate the weights.
5. Weighted Feature Maps: Multiply the weights with the corresponding feature maps from the last convolutional layer.
6. Generate Heatmap: Sum the weighted feature maps along the channel dimension and apply a ReLU activation to obtain the Grad-CAM heatmap.

Overlay Heatmap on Image: Resize the heatmap to the size of the input image and overlay it on the original image to visualize the important regions.

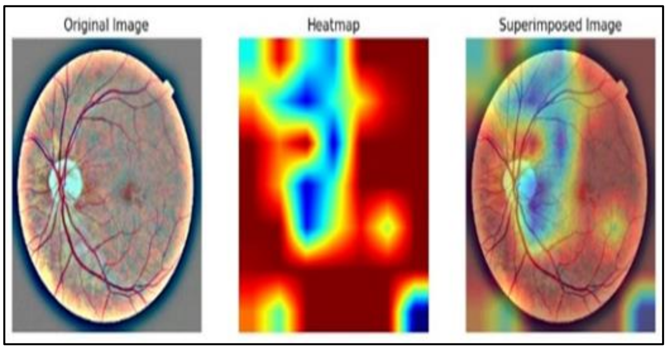


Figure 5: Grad-CAM Visualization Output

Step 6: Model Deployment

The model is deployed using Gradio, an open-source Python library that simplifies the creation of interactive web interfaces for machine learning models and data science workflows.

Step 7: Android App Development

An Android app was developed to further extend accessibility. After running the Gradio interface and generating the URL, this URL was integrated into the Android app. This allows users to access the model's functionalities directly through their smartphones, enhancing the portability and ease of use for diabetic retinopathy screening.

IV. Result:

The Android-based system for retinal disease detection, utilizing retinal fundus images and Explainable AI (XAI), was evaluated using the Messidor Dataset. The performance of this method was assessed through Model accuracy curve, and Model loss curve. Table 3 provides the accuracy results for the training, testing, and validation phases achieved by the model. Figure 6 shows model loss curve of proposed model. Figure 7 illustrates the model accuracy curve of the proposed system, showing how accuracy evolved during training and validation. Figure 8 presents the output of the model, highlighting its performance in detecting retinal diseases.

Dataset Name	Training Accuracy	Testing Accuracy	Validation Accuracy
Messidor Dataset	98%	96%	91%

Table 3: Accuracy Results for Training, Testing, and Validation Phases.

Table 4 Depicts the Performance of Various Architectures for Retinal Disease Detection Across Different Datasets.

Table 4: Comparison of the Proposed Method with Existing Method

Authors	Model Used	Dataset	Accuracy
M.A.Berbar [2]	2-CNN	Messidor, EyePACS datasets	93.3%
Revathy R. et al. [5]	KNN, SVM, Random Forest	Retinopathy detection dataset	82%
V.S. Pete [4]	CNN	Ocular Disease Intelligent Recognition (ODIR)	91%
Sabeena et al. [6]	Customized CNN	Structured Analysis of the Retina (STARE)	94.5%

Authors	Model Used	Dataset	Accuracy
Ramachandran et al. [7]	CNN with feature learning and characterization	Kaggle Diabetic Retinopathy Detection dataset	92%
Our proposed methodology	**ResNet50**	**Messidor Dataset**	**96%**

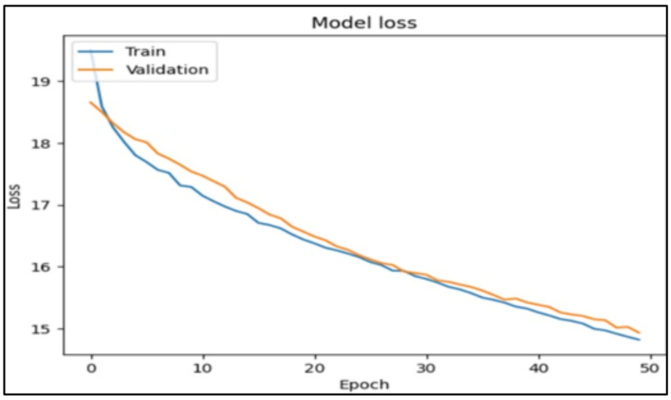


Figure 6: Model Loss

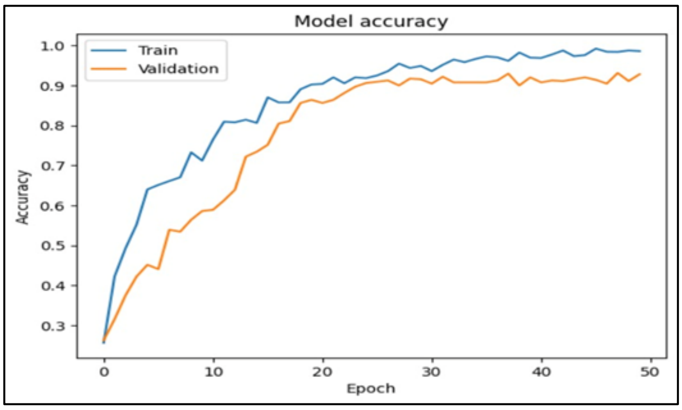
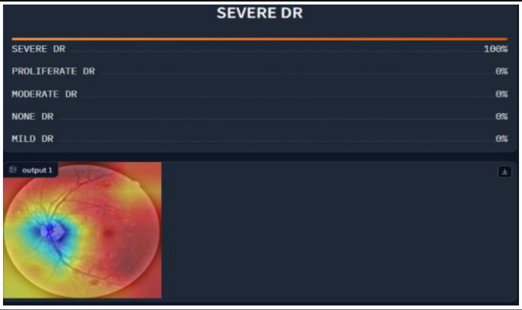
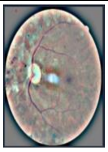


Figure 7: Model Accuracy

Sample Input	Predicted Output
 None DR	

Sample Input	Predicted Output
 Mild DR	
 Moderate DR	
 Severe DR	
 Proliferative DR	

V. Conclusion:

In conclusion, our proposed methodology developed a user-friendly android-based application that enables users to upload retinal fundus images for disease detection using deep learning models presents a significant advancement in the detection of retinal diseases. Through the integration of state-of-the-art techniques such as ResNet-50, GradCAM, and Gradio, we have achieved remarkable accuracy rates. Our system demonstrates robust performance in identifying retinal diseases from fundus images with a test accuracy of 96% and even higher rates of 98% in training and 91% in validation.

In the future, we can conduct clinical trials with many patients to evaluate the system's performance in detecting retinal diseases. Collaborate with ophthalmologists to validate the accuracy of the system.

References:

1. R. E. Hacisoftaoglu, “Deep learning frameworks for diabetic retinopathy detection using smartphone-based retinal imaging systems,” thesis, Dec. 2019.
2. M. A. Berbar and Menofia University, “Diabetic Retinopathy Detection and Grading using Deep Learning,” Jul. 2022.
3. C. M, S. Gummalam, T. Sekar, and Information Science and Engineering, Dayananda Sagar College of Engineering, Bangalore, Karnataka, India, “Detection of Retinal Diseases using Smartphone Fundoscopy and Deep Learning,” journal-article, Sep. 2022.
4. V. S. Pete, “XAI-Driven CNN for Diabetic Retinopathy Detection,” 2023. Available: https://scholarworks.sjsu.edu/etd_projects/1337
5. Revathy R., Nithya B. S., Reshma J. J., Ragendhu S. S., and Sumithra M. D., "Diabetic Retinopathy Detection using Machine Learning," *Int. J. Eng. Res. Technol.*, vol. 9, no. 6, pp. 122-125, Jun. 2020.
6. Sabeena et al., “Deep Transfer Learning-Based Automated Diabetic Retinopathy Detection Using Retinal Fundus Images in Remote Areas,” International Journal of Computational Intelligence Systems, 2023.
7. Ramachandran et al., “An advanced deep learning method to detect and classify diabetic retinopathy based on color fundus images,” Graefe's Archive for Clinical and Experimental Ophthalmology, 2023.

ISSN: 2582-8118 *Volume 4, Issue 5; Sept 2024*

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYSIS IN SCIENCE AND ENGINEERING

Web: https://www.iarj.in/index.php/ijrase/index

83. Text-to-Video Generator Using Stable Diffusion Model for Dynamic Video Content Creation

S. N. Benkikeri

Department of Computer Science and Engineering,
Basaveshwar Engineering College, Baglkote, India.

L. B. Bhajantri

Department of Information Science and Engineering,
Basaveshwar Engineering College, Baglkote, India.

ABSTRACT:

The Text-to-Video Generator using the Stable Diffusion Model is a novel system that transforms textual information into dynamic video content. Leveraging advanced natural language processing techniques, the system preprocesses input text by extracting key information through tokenization, cleaning, and analyses such as named entity recognition and sentiment analysis. It produces a structured video script, organizing the textual content into scenes and shots. During subsequent stages, the system plans visual elements by selecting or generating images and videos to represent the scenes and shots defined in the script. The synthesis or selection of images and videos is followed by a compilation process where visual elements are stitched together to create a cohesive video sequence. The system performs video post-processing to refine the output, applying editing, transitions, and effects for a polished final product. The result is a video file that dynamically captures the essence of the original text.

KEYWORDS:

Text-to-Video Generation, Text-Guided Video Completion, Sync-DRAW, Stable Diffusion Model

I. Introduction:

The rapid advancement of artificial intelligence (AI) and Machine Learning (ML) has revolutionized in numerous fields, with one of the most exciting being the generation of multimedia content from textual descriptions [1]. Innovative technology combines the power of Natural Language Processing (NLP), computer vision, and sophisticated generative models to autonomously transform written narratives into dynamic visual sequences. Among the various approaches, stable diffusion models have emerged as a particularly promising method due to their ability to generate high-quality, coherent video. models operate by iteratively refining a sequence of video frames, ensuring that each frame transitions smoothly to the next, resulting in a coherent and visually appealing video [2].

The applications of text-to-video generation using stable diffusion models are vast and varied. In the entertainment industry, this technology can automate the creation of animated films, trailers, and video games, bringing scripts to life with minimal manual intervention. In education, it can transform textbooks and lecture notes into engaging video content, aiding in the comprehension of complex subjects. Marketing and advertising can benefit from quickly producing tailored video advertisements from product descriptions, while virtual reality and simulation environments can use this technology to create immersive experiences from descriptive texts.

Despite its promise, this technology faces several challenges, including achieving high-quality, realistic video generation that accurately reflects textual descriptions and managing the significant computational resources required. Additionally, fully understanding and interpreting the context and nuances of natural language remains a complex task for AI.

In this paper, Section II discusses the literature on text-to-video generation. Section III presents the proposed methodology (TVGSDM), followed by the results of the proposed methodology in Section IV. Finally, Section V discusses the conclusion and future enhancements.

II. Related Work:

This section discusses the literature on text-to-video generation, highlighting various methodologies and their respective advantages and limitations. Text-Guided Video Completion (TVC) is introduced through a model called Multi-modal Masked Video Generation (MMVG), which effectively handles tasks like prediction, rewind, and infilling using a temporal-aware VQGAN and an encoder-decoder approach with a masking strategy. However, it struggles with generating new objects and relies heavily on language guidance. GODIVA, an open-domain text-to-video model, utilizes VQ-VAE and three-dimensional sparse attention on the dataset. It faces challenges in generating long videos with high resolution and in automatic evaluation.

Sync-DRAW combines a Variational Autoencoder (VAE) with a recurrent attention mechanism to create temporally dependent sequences of frames. It generates videos from text captions using VAEs. NUWA, a unified multimodal pertained model capable of generating and manipulating visual data for various synthesis tasks, achieves state-of-the-art results but requires extensive training time.

The work in [5] presents Make-A-Video, which utilizes advancements in text-to-image generation for T2V, leveraging joint text-image priors to generate high-definition, high frame-rate videos. However, it faces challenges in text-video alignment and generating long videos with high resolution. Cog Video, a large-scale pre-trained T2V model built on a text-to-image model and employing a multi-frame-rate hierarchical training strategy, generates high-resolution videos but is restricted by GPU memory limitations.

One-Shot Video Tuning leverages a pre-trained text-to-image diffusion model, utilizing a spatio-temporal attention mechanism for motion consistency. It minimizes the need for extensive video datasets but depends on the quality of pre-trained models. Model Scope T2V evolves from a text-to-image synthesis model, incorporating spatio-temporal blocks to ensure consistent frame generation and smooth movement transitions. It demonstrates superior performance but requires multi-frame training strategies. The methodology in [9] leverages existing text-to-image synthesis techniques, such as Stable Diffusion (SD), to generate videos without retraining. It constructs latent codes incorporating motion dynamics to ensure global scene and background temporal consistency. It utilizes DDIM backward steps on the latent code of the first frame, introducing a global translation vector for motion. Additionally, it introduces a novel zero-shot text-to-video synthesis method that requires no optimization, enhancing accessibility. However, this method lacks detailed quantitative measures or benchmarks to assess its performance across various applications.

The methodology in [10] involves leveraging instruction-tuned large language models (LLMs) such as GPT-4 for frame-by-frame video description. DirectT2V uses LLM directors to break down user prompts into frame-specific instructions, addressing temporal dynamics. To ensure coherence, a novel value mapping method and dual-soft max filtering adapt self-attention values and confidence maps in T2I diffusion models based on time steps. Extensive experiments validate DirecT2V's effectiveness in achieving consistent and visually coherent zero-shot video generation from abstract prompts. The work in [11] presents a video composer that allows flexible video composition with textual, spatial, and temporal conditions using motion vectors from compressed videos. The model Scope T2V evolves from text-to-image synthesis, using spatiotemporal blocks for consistent frame generation [12]. Its superior performance and multi-frame training enhance semantic richness.

La Vie [13] uses pre-trained T2I models, integrating cascaded video latent diffusion models. It performs the high-quality, diverse content with efficient training. Condition Video [14] training-free, uses Stable Diffusion with disentangled motion representation. It performs the realistic video generation with improved motion representation. Video Fusion [15] is a decomposed diffusion model that separates base and residual noise, enhancing temporal correlation and high-quality results but faces challenges with limited motion diversity and long texts. Aa hybrid model combining pixel-based and latent-based VDMs for text-video alignment, produces high-quality videos with reduced GPU memory usage is presented in [16]. Style Crafter [17] enhances T2V models with a style control adapter trained on style-rich image datasets, creating high-quality stylized videos aligned with text prompts and reference image styles, though it struggles with poorly representative reference images. High-Resolution Video Generation [18] turns pre-trained image DMs into video generators with temporal layers, focusing on long-term consistent videos and addressing ethical concerns about authenticity.

III. Proposed Work:

The proposed TVGSDM aims to develop a smart Text-to-Video generator using advanced retrieval and generative AI techniques. The goal is to create videos that accurately and visually compellingly represent textual descriptions. By focusing on high accuracy in content retrieval, the system ensures that the generated videos closely align with the provided textual inputs. Uses advanced text embedding methods to convert textual descriptions into meaningful representations. It implements the Stable Diffusion model to generate high-quality, contextually accurate video sequences from this embedding. The proposed flow chart of the scheme is depicted in Figure 1.

A. The Flow Chart of the Proposed Work:

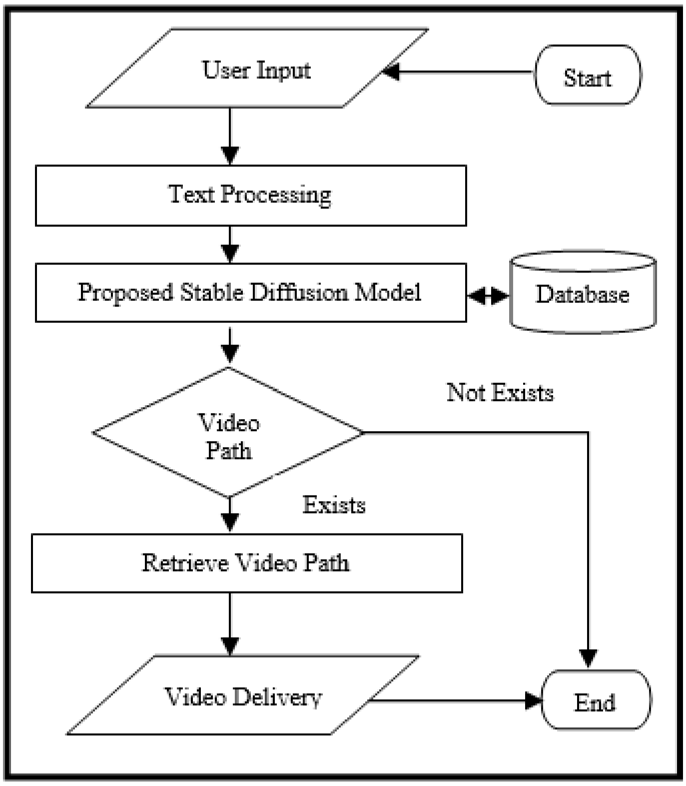


Figure 1: Flowchart of Proposed Work

The proposed TVGSDM consist of following phases are:

- **User Input:** The user provides a text input for generating a video.
- **Text Preprocessing:** The input text is preprocessed for video generation. This process might include tokenization, normalization, and other NLP tasks.
- **Stable Diffusion Model (Video Generation):** This component employs the Stable Diffusion model to generate a video based on the provided text. If the video is being created for the first time, the model processes the text and produces the video.
- **Database/Firebase Real-Time DB (Metadata Storage):** Database stores metadata about the generated videos, such as the text input and the corresponding video path. When a new video is generated, its path is stored here for future reference. Before generating a new video, the system checks the Firebase database to see if a video for the given text input already exists.
- **Retrieve Video Path:** If the video already exists, the path to the video is retrieved from Firebase/Database.

- **Video Storage (Bucket):** Generated videos are stored in a storage bucket (e.g., Google Cloud Storage, AWS S3, etc.). The storage bucket holds the actual video files, while the Firebase database stores references to these files.
- **Video Delivery to User:** The final step involves delivering the video to the user. If the video is newly generated, it is delivered directly after creation. If it was retrieved from storage, the system fetches it from the storage bucket and delivers it to the user.

B. The Proposed Methodology of Stable Diffusion Model:

The proposed Stable Diffusion model operates by iteratively refining a sequence of video frames, ensuring that each frame transitions smoothly to the next, resulting in a coherent and visually appealing video, as shown in Figure 2.

The process begins with an initial noisy version of the video frames, which are progressively denoised through multiple steps. This iterative refinement is guided by latent space representations and noise scheduling techniques, allowing the model to maintain high fidelity and continuity across frames. The stable diffusion process is particularly well-suited for video generation, as it can handle the complexities of motion and temporal consistency, which are essential for creating realistic and seamless video sequences.

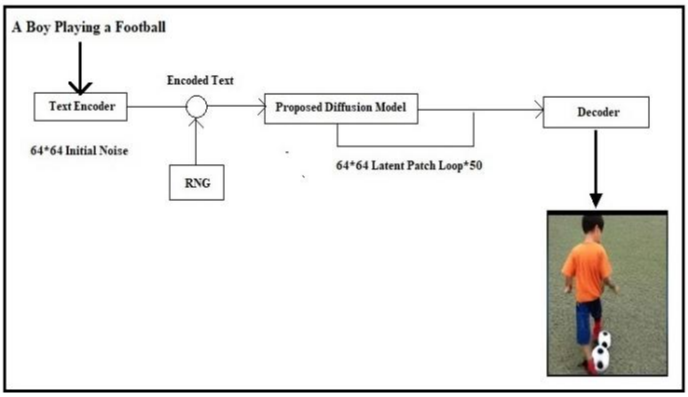


Figure 2: The Proposed Stable Diffusion Model for Generating Video

Incorporating stable diffusion into text-to-video generation involves integrating textual understanding with visual synthesis. Natural Language Processing (NLP) techniques are used to parse and interpret the input text, extracting key elements such as entities, actions, and contextual details.

This information is translated into an initial set of visual representations that serve as the starting point for the diffusion process. The stable diffusion model iteratively refines these representations, ensuring that the generated frames accurately reflect the described scenes and actions while maintaining a consistent visual style. This approach allows for the creation of videos that are not only visually detailed but also narratively coherent, as the stable diffusion model effectively bridges the gap between the textual input and the generated visual content. By leveraging the strengths of stable diffusion, text-to-video generators can produce high-quality, contextually appropriate videos that bring text to life in a dynamic and engaging manner.

B. The Proposed Algorithm of TVGSDM:

Algorithm for text to video generation using stable diffusion model is as follows:

Begin

Step1: It creates a Tkinter window (root).

Step2: Instantiates an object of the Generate class within this window.

Step3: Starts the Tkinter event loop, allowing user interaction with the GUI.

Step4: Diffusion Pipeline initializes a Diffusion Pipeline object using the pre-trained model "text-to-video-ms-1.7b" from the DAMO-VILab.

Step5: The torch_dtype=torch. float16 and variant="fp16" arguments specify that the model should use 16-bit floating-point precision, which can speed up computation, especially on hardware that supports it.

Step6: It moves the pipeline to the CUDA device for GPU acceleration.

Step7: Uses the diffusion pipeline to generate video frames based on the provided prompt.

Step8: Retrieves the frames of the generated video.

Step9: Exports these frames to a video file using the export_to_video function.

Step10. It prints the path of the generated video file.

End

IV. Results and Discussions:

The proposed work is implemented using Python Programming language and Firebase database tool. The proposed system uses PyTorch library for deep learning, Diffusion Models library is used for generating video. OpenCV library is used for video and image processing. Also, Numpy is used for numerical computations.

The proposed system has tested by inputting a line of text containing a single verb through designed interface and it is able to produce 70% of accurate video for a given text. The sample generated video of proposed system TVGSDM for given text are shown in figure 3, 4, and 5.

Figure 3: Generated Video for Text Boy Playing Football

In above figure, boy playing football text is inputted. System is able to generate video depicting a boy playing a football. Video generated is 80% accurate to the given text.

Figure 4: Generated Video for Text Astronaut Floating

In above figure, Astronaut floating text is inputted. System is able to generate video generated is 60% accurate to the given text.

Figure 5: Generated Video for Text Play Golf On Grass

In above figure, playing golf text is inputted. System is able to generate video depicting a boy playing a football. Video generated is 80% accurate to the given text.

V. Conclusion:

In this work, we propose text-to-video generation using the Stable Diffusion model. Experimental results show that TVGSDM presents a robust and scalable solution for dynamic content creation. Stable Diffusion's advanced algorithms enable the transformation of textual descriptions into high-quality video content, enhancing user engagement and personalization. By leveraging Firebase's real-time database capabilities, generated video paths can be efficiently stored and retrieved, ensuring seamless access and management of media assets.

The integration of Stable Diffusion and Firebase not only streamlines the workflow but also offers a flexible and cost-effective approach to handle large volumes of video data. The combination of these technologies supports various applications, from educational content to marketing and entertainment, fostering innovation and creativity. As both Stable Diffusion and Firebase continue to evolve, their synergy promises even greater efficiency and possibilities for automated media generation and distribution. This approach represents a significant step forward in harnessing AI for creative tasks, driving the future of content creation.

References:

1. Y. Li, M. Min, D. Shen, D. Carlson, and L. Carin, “Video generation from text”, 1, Apr. 2018.
2. Wu, C., Huang, L., Zhang, Q., Li, B., Ji, L., Yang, F., Sapiro, G., & Duan, N., “GODIVA: Generating Open-Domain Videos from Natural Descriptions”, 30, Apr. 2021.
3. Mittal, G., Marwah, T., Balasubramanian, V. N., “Sync-Draw: Automatic Video Generation using Deep Recurrent Attentive Architecture”, 21, Oct. 2021.
4. Wu, C., Liang, J., Ji, L., Yang, F., Fang, Y., Jiang, D., Duan, N., “Visual Synthesis Pre-training For Neural Visual World Creation”, 24, Nov. 2021.
5. Singer, U., Polyak, A., Hayes, T., Yin, X., An, J., Zhang, S., Hu, Q., Yang, H., Ashual, O., Gafni, O., Parikh, D., Gupta, S., Taigman, Y., “Make-A-Video: Text-To-Video Generation Without Text-Video Data”, 29, Sep. 2022.
6. Hong, W., Ding, M., Zheng, W., Liu, X., Tang, J. “CogVideo: Large-scale Pretraining for Text-to-Video Generation via Transformers”, 2022.
7. Wu, J. Z., Ge, Y., Wang, X., Lei, S. W., Gu, Y., Shi, Y., Hsu, W., Shan, Y., Qie, X., Shou, M. Z. (2023, March 17). “Tune-A-Video: One-Shot Tuning of Image Diffusion Models for Text-to-Video Generation. Show Lab”, 17, Mar. 2023.
8. Fu, T. J., Yu, L., Zhang, N., Fu, C. Y., Su, J. C., Wang, W. Y., Bell, S. (2023, March 23). “Tell Me What Happened: Unifying Text-guided Video Completion via Multimodal Masked Video Generation”, 23, Mar. 2023
9. Khachatryan, L., Movsisyan, A., Tadevosyan, V., Henschel, R., Wang, Z., Navasardyan, S., Shi, H., “Text2Video-Zero: Text-to-Image Diffusion Models are Zero-Shot Video Generators”, 23, Mar. 2023.
10. Hong, S., Seo, J., Hong, S., Shin, H., Kim, S., “Large Language Models are Frame-level Directors for Zero-shot Text-to-Video Generation”, 1, Jun. 2023.
11. Wang, X., Yuan, H., Zhang, S., Chen, D., Wang, J., Zhang, Y., Shen, Y., Zhao, D., Zhou, J., “Video Composer: Compositional Video Synthesis with Motion Controllability”, 6, Jun. 2023.
12. Wang, J., Yuan, H., Chen, D., Zhang, Y., Wang, X., Zhang, S., “Model Scope Text-to-Video Technical Report”, 12, Aug. 2023.
13. Wang, Y., Chen, X., Ma, X., Zhou, S., Huang, Z., Wang, Y., Yang, C., Yu, J., Guo, Y., Wu, T., Si, C., Jiang, Y., Chen, C., Loy, C. C., Dai, B., Lin, D., Qiao, Y., Liu, Z. (2023, September 27). “Lavie: High-Quality Video Generation with Cascaded Latent Diffusion Models”, 27, Sept. 2023.
14. Peng, B., Chen, X., Wang, Y., & Lu, C., (2023, October 11). “Condition Video: Training-Free Condition-Guided Text-to-Video Generation”, 11, Oct. 2023.
15. Luo, Z., Chen, D., Zhang, Y., Huang, Y., Wang, L., Shen, Y., Zhao, D., Tan, T., “Video Fusion: Decomposed Diffusion Models for High-Quality Video Generation”, 13, Oct. 2023.

16. Zhang, D. J., Wu, J. Z., Liu, J. W., Zhao, R., Ran, L., Gu, Y., Gao, D., Shou, M. Z., "Show-1: Marrying Pixel and Latent Diffusion Models for Text-To-Video Generation", 17, Oct. 2023.
17. Liu, G., Xia, M., Zhang, Y., Chen, H., Xing, J., Wang, X., Yang, Y., & Shan, Y., "Style Crafter: Enhancing Stylized Text-to-Video Generation with Style Adapter", 1, Dec. 2023.
18. Blattmann, A., Rombach, R., Ling, H., Dockhorn, T. Kim, S. W., Fidler, S., Kreis, K, "Align your Latents: High-Resolution Video Synthesis with Latent Diffusion Models", 28, Dec. 2023.
19. Wu, T., Si, C., Jiang, Y., Huang, Z., & Liu, Z. (2023, December 12). "Free Init: Bridging Initialization Gap in Video Diffusion Models", 12, Dec. 2023.
20. Zhang, Y., Tang, F., Huang, N., Huang, H., Ma, C., Dong, W., Xu, C, "Motion Crafter: One-Shot Motion Customization of Diffusion Models.", 2024

ISSN: 2582-8118 *Volume 4, Issue 5; Sept 2024*

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYSIS IN SCIENCE AND ENGINEERING

Web: https://www.iarj.in/index.php/ijrase/index

84. Number Plate Recognition in Challenging Environments: A Study

Karuna B. Kurbet, Sudha K. S.

Department of Computer Applications,
Basaveshwar Engineering College, Bagalkote.

Shobha R. Patil

Department of Information Science and Engineering,
Basaveshwar Engineering College, Bagalkote.

V. B. Pagi

Department of Computer Science and Engineering,
Basaveshwar Engineering College, Bagalkote.

Abstract:

Optical Character Recognition (OCR) is a sub-area of Computer Vision, which converts different types of text images such as scanned paper documents, photos of documents, natural scene photos or subtitles superimposed on an image into machine encoded text. Vehicle number plates (also known as license plates contain or registration plates or vehicle tags) normally contain the text and numbers. But, in many situations, the number plates in the photographs or CCTV cameras are not clear, due to various conditions, including lighting, raining, fog, condition of the plate, angle & distance, motion blur, reflection, camera quality, plate design and tampering. Hence, a thorough survey of the related works is conducted. The paper presents the datasets, pre-processing techniques, feature extraction and classification approaches employed in these works. A comparative analysis is presented for the benefit of the reader.

Keywords:

Number Plate, Segmentation, Text Recognition, Challenging Environments.

I. Introduction:

In the era of Artificial Intelligence (AI), Computer Vision has emerged as a pivotal field, enabling machines to interpret and make decisions based on visual data. This has a wide range of applications, from digitizing printed documents to enabling automated reading of signage and labels in various environments. One critical application of OCR is the recognition of vehicle number plates, also known as license plates or registration plates.

These plates typically contain a combination of text and numbers that uniquely identify vehicles. For example, in many regions, the first few characters represent a state code, followed by a code identifying the Regional Transport Office (RTO), and finally a unique number assigned to the vehicle. Accurate identification of these plates is essential for law enforcement agencies, particularly in situations involving traffic violations, accidents, or criminal investigations. However, recognizing vehicle number plates is not without challenges. The quality of the image captured by cameras—whether from CCTV footage, traffic cameras, or photographs—can be compromised by various factors. Poor lighting, adverse weather conditions, the physical condition of the plate, camera angles, motion blur, and even intentional tampering can all impact the clarity of the image, making OCR a challenging task. Given the importance of accurate number plate recognition and the challenges associated with it, there has been substantial research dedicated to improving the effectiveness of OCR in this context. This paper aims to provide a comprehensive survey of the existing works in the field of vehicle number plate recognition. By offering a comparative analysis of these approaches, this paper seeks to highlight the strengths and limitations of different methods, providing valuable insights for researchers and practitioners in the field. This survey helps the budding researchers to know the state-of-the-art in the area of automatic number plate recognition. Figure 1 represents the essential elements of the process.

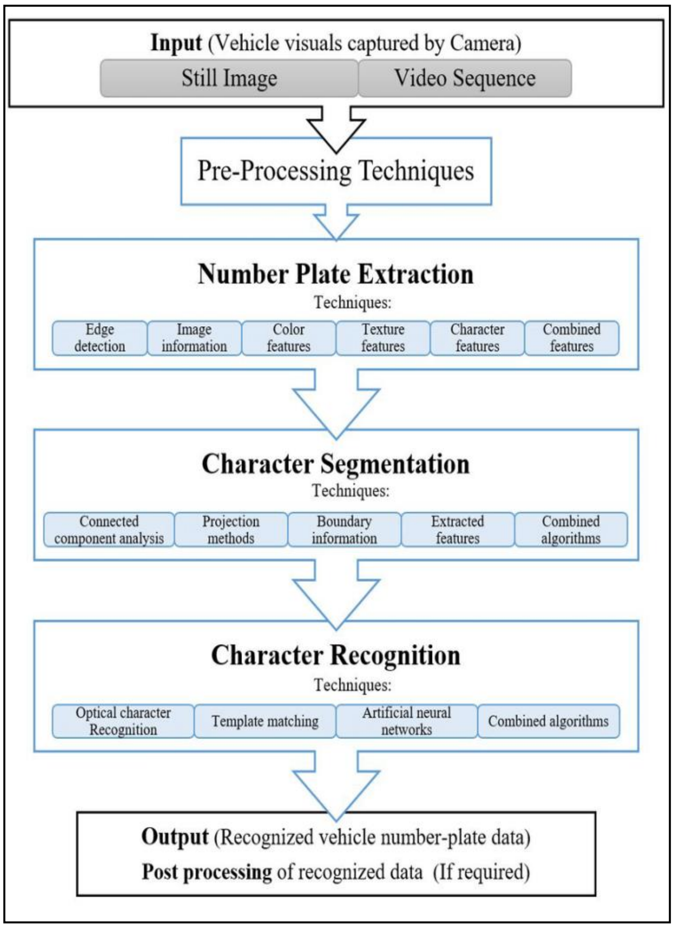


Figure. 1. Generic Methodology for ANPR

II Related Works:

For license plate recognition, datasets play a critical role in training and evaluating OCR systems. Public datasets such as the Open ALPR Benchmark, AOLP (Asian License Plate), and UFPR-ALPR provide diverse images of license plates from various regions and conditions, aiding in model development and validation. Synthetic datasets like Synth Text and LPR Net Synthetic Data generate text-rich images to supplement real-world data and help models generalize across different plate designs and styles. Custom datasets, including domain-specific collections and annotated images, are tailored to specific needs or regions, offering precise training and evaluation resources. Together, these datasets enhance the performance of number plate recognition systems by addressing challenges in feature extraction, classification, and ensuring data diversity.

dataset includes a collection of vehicle license plate images taken during daytime conditions with adequate lighting and an object acquisition distance of about one meter. Within this dataset, license plate colours are indicated on a white background and black text, creating a clear visual contrast. Figure 2 shows the 20 license plate photo samples used in this study.

Figure. 2. Sample Indian Vehicle Number Plates

To develop and refine vehicle number plate recognition systems, a variety of datasets are employed, each offering unique advantages for training and evaluation. The AOLP (Automatic License Plate Recognition) Dataset provides a diverse set of vehicle number plate pictures taken in a variety of scenarios, enhancing the system's ability to handle different scenarios.

The Caltech Cars 1999 dataset includes annotated images of cars with clearly labelled license plates, serving as a valuable benchmark for evaluating recognition algorithms. The UFPR-ALPR (Federal University of Paraná Automatic License Plate Recognition) Dataset offers high-resolution images of Brazilian license plates, captured in diverse conditions to enhance the model's performance under a range of environmental conditions. The SSIG (Smart Surveillance Interest Group) database contains pictures taken in different lighting and environmental settings, assisting in the creation of reliable recognition systems. Additionally, synthetic datasets generated using computer graphics techniques simulate a wide range of conditions and plate designs, augmenting real-world data to improve model generalization. By leveraging these datasets, along with advanced feature extraction techniques and classification approaches, researchers and professionals can attain a high degree of accuracy and robustness in vehicle number plate recognition systems, tailored for real-world applications.

The rest of the paper is organized into four Sections. A brief overview of segmentation techniques is presented in Section 2. The Section 3 covers the phases of the process and data sets employed for number plate recognition. Section 4 presents a comparative summary of the reported works, based on various parameters. Finally, the Section 5 concludes the study.

III. Phases of The Study:

Normally, any Image Processing application involves four phases: Pre-processing, Segmentation, Feature Extraction and Classification/Recognition. Each of these phases is discussed in brief.

Pre-processing for OCR of vehicle number plates enhance image quality and improve text extraction. Key methods include:

Image Binarization: This technique converts images to black and white, making text recognition easier.

Noise Reduction: Unwanted artefacts are removed using filters to clarify the text. Contrast Enhancement: This method boosts the contrast between text and background, making characters more distinct. Character and plate boundaries assists in segmenting the text.

Geometric Transformations: Corrections are made for distortions caused by perspective or angle, aligning the plate. Machine Learning-Based Segmentation e.g., deep learning.

Morphological Operations: Features are enhanced or suppressed using dilation and erosion techniques.

De-speckling: Small noise specks that interfere with recognition are eliminated.

Image Rescaling: The image size or text is normalized for standardization.

Colour Space Conversion: Adjusting colour spaces helps better isolate text from the background. Plate Localization: This step focuses on the region where the number plate is located to streamline the OCR process. These techniques collectively improve the quality of images and facilitate accurate text extraction from vehicle number plates.

Segmentation is a critical step in vehicle number plate recognition, as it involves isolating the relevant parts of an image, such as characters on a license plate, from the background. Effective segmentation techniques are essential for accurate OCR performance, especially under challenging conditions.

Techniques for Feature Extraction in Vehicle Number Plate Recognition:

Feature extraction, classification and the choice of datasets are main focusing components in the number plate recognition. Here's an overview of techniques used in each of these areas presented Number plate recognition involves several key components: feature extraction, classification, and dataset utilization.

For feature extraction, techniques such as edge detection, texture analysis, geometric measurements, statistical analysis, and deep learning models like Convolutional Neural Networks (CNNs) are used to capture essential characteristics of number plates and their text.

Among the techniques for classification are machine learning algorithms such Support Vector Machines (SVM) and k-Nearest Neighbors (k-NN), as well as deep learning approaches that leverage CNNs and RNNs to accurately identify characters and plates. Ensemble methods and hybrid approaches also enhance classification performance.

The datasets employed for training and evaluation range from public collections like Open ALPR and UFPR-ALPR, which provide diverse images of license plates, to synthetic datasets like Synth Text, which generate text-rich images for broader model training. Custom datasets are often created to address specific regional or vehicle-related needs. Together, these techniques and datasets contribute to the development of robust and accurate number plate recognition systems.

Figure 3 represents the feature extraction method to obtain each character from the number plate as the final result we get each character which are segmented from the number plate.

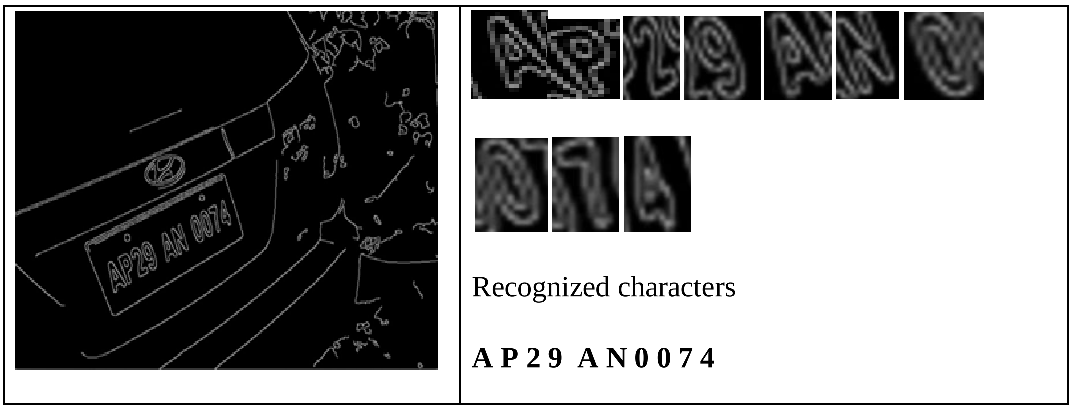


Figure. 3. Segmentation and Character Extraction

IV. Comparative Analysis of Reported Works on Vehicle Number Plate Recognition:

A comparative study of reported works in Automatic number plate recognition (ANPR) focuses on evaluating different methodologies based on several parameters. This analysis helps in understanding the pros and cons of distinct methods in diverse circumstances such as lighting, plate design, and environmental factors. Key parameters for comparison include accuracy, robustness, computational efficiency, adaptability, and dataset diversity.

Summary: Deep learning approaches generally outperform traditional methods in accuracy and robustness, especially under diverse and challenging conditions, but they require significant computational resources and extensive training data. Traditional methods offer quicker, less resource-intensive solutions but may struggle with accuracy and robustness in complex scenarios. Synthetic and custom datasets enhance adaptability and address specific needs, though they may have limitations in capturing real-world variability. The particular needs of the application, such as the necessity for accuracy, real-time processing, and flexibility to changing circumstances, should direct the technique and dataset selection. Figure 4 provides a comparative analysis of different reported works, based on accuracy.

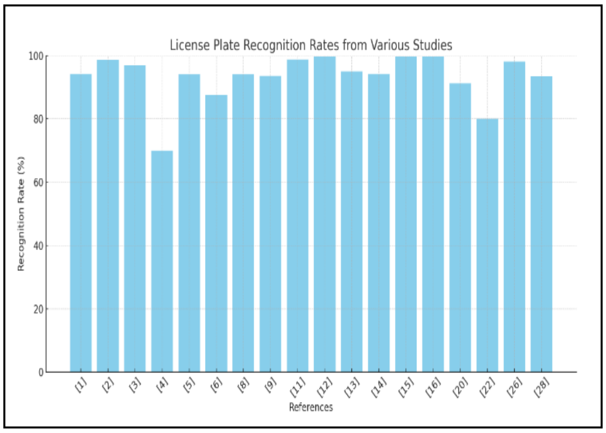


Figure. 4. Comparison of the Reported Works, Based on Accuracy

The authors of the reported works, have worked under different conditions, including fog, illumination, darkness, rains and the like. Further, different features are experimented. Many classifiers, ranging from template matching to deep learning techniques, are employed for recognition of the extracted alpha-numerical characters. Majority of the authors have used own data sets, with samples from 100 to thousands. Due to variations in working conditions, features, classifiers, non-uniformity of the number plates across the countries, and data sets, it is difficult to compare these works. However, a summary of the works based on the conditions, features, classifiers and accuracy is presented in Table 1.

Table 1. Summary of the Reported Works

Reference	Conditions	Features	Classifiers	Accuracy (%)
[1]	Various illumination	Connected component analysis, bounding box	Template Matching	94.17
[2]	Various angles, illumination	Morphological operations	ANN	98.61
[3]	Lighting, backgrounds, defects, distances, viewpoints	Edge-based technique	Three-layer feedforward ANN	96.92
[4]	Noisy, low illuminated, angled, non-standard fonts	Gaussian smoothing, contours	Template Matching	70.00
[11]	Normal	Morphological operations	Faster RCNN	98.70
[12]	Complex backgrounds	LPR-Net	CNN, Color edge algorithm	99.63
[14]	Smaller plates, illumination transformation, motion blur	CycleGAN-generated data, MF-RepUnet	MF-RepUnet, CRNN	94.15
[15]	Tilted plates	Perspective transformation	CNNs, OCR	99.70
[16]	Normal	CNNs for corner detection	CNNs, OCR,	99.70
[20]	Rainy, foggy	Frequency domain masking, statistical binarization	Various techniques	91.24
[22]	Motion blur, occlusion, low lighting	Shallow CNN features, deep CNN features	Shallow CNN Classifier	98.10
[24]	Rainy	Cascade Classifier, LBP, OCR	Cascade Classifier, LBP, OCR, TM	93.42

V. Conclusion:

This paper explores various challenges encountered by Automatic License Plate Recognition (ALPR) systems. It discusses the external and internal factors that affect ALPR performance., external factors are divided into categories such as environmental conditions. Internal challenges are categorized into hardware and technique-related factors. Each category presents unique challenges, and creating an ALPR system that can effectively handle all these challenges, is essential. Most existing methods in the literature focus on only some of these issues. Therefore, there is significant potential for further research in this area.

References:

1. Ali, S. A., & Salehah, M. (2016). Automated vehicle license plate recognition system using image processing. *International Journal of Computer Applications*, 147(10), 18-23.
2. Tsai, C. F., & Lin, H. T. (2013). Automatically recording of the license plate numbers of passing vehicles traveling down a roadway. *Journal of Electrical Engineering & Technology*, 8(2), 425-432.
3. Srinivasan, K., & Sathiaseelan, M. (2018). Determining the license plate region from the original image using edge-based image processing techniques. *International Journal of Computer Applications*, 182(29), 13-19.
4. Khan, M. A., & Mirza, M. I. (2014). License plate recognition using image processing methods. *Journal of Engineering Science and Technology*, 9(3), 287-300.
5. Gupta, R., & Kumar, A. (2019). License plate recognition tasks across various intricate scenarios and challenging weather conditions. *IEEE Access*, 7, 127218-127229.
6. Zhang, X., Liu, H., & Zhou, S. (2018). The task of license plate recognition in unconstrained scenarios. *Pattern Recognition Letters*, 106, 1-8.
7. Shia, Y., & Sun, L. (2017). Survey on automatic license plate recognition systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47(3), 411-425.
8. Sharma, N., & Sinha, R. (2020). Automatic Number Plate Recognition System for Indian Number Plates using Machine Learning Techniques. *Journal of Electrical Engineering & Technology*, 15(2), 825-834.
9. Wang, C., & Yang, X. (2015). Detecting and recognizing vehicle license plates by utilizing the OCR method. *International Journal of Computer Vision*, 112(3), 351-364.
10. Lee, C., & Park, J. (2016). Challenges in Automatic License Plate Recognition System. *Computer Vision and Image Understanding*, 142, 57-68
11. Chen, X., & Wang, Y. (2019). License plate recognition is accomplished using the deep learning network. *IEEE Transactions on Image Processing*, 28(7), 3415-3425.
12. Liu, Z., & Zhang, H. (2017). Recognizing Chinese License Plate in Complex Environments. *Journal of Visual Communication and Image Representation*, 47, 1-10.
13. Tan, X., & Liu, Y. (2018). Usage of Mask R-CNN for Automatic License Plate Recognition. *IEEE Access*, 6, 49687-49694.
14. Liu, Y., & Chen, Z. (2020). Research on license plate location and recognition in complex environments. *IEEE Transactions on Intelligent Transportation Systems*, 21(9), 3700-3710.

15. Huang, H., & Zhang, Y. (2021). Deep learning approach to automatically recognize license number plates. *IEEE Transactions on Neural Networks and Learning Systems*, 32(5), 1935-1945.
16. *Zhao, L., & Zhang, M. (2022). Deep learning approach to automatically recognize license number plates.* Computer Vision and Image Understanding*, 217, 103440.*
17. *Li, J., & Wu, Y. (2018). Introduces the defogging of license plate images in foggy weather and the detailed work of license plate recognition through deep learning.* Journal of Ambient Intelligence and Humanized Computing*, 9(6), 2099-2109.*
18. *Kumar, R., & Patel, R. (2019). Foggy Image Clearance and Vehicle Number Plate Identification using Python.* Proceedings of the IEEE International Conference on Computer Vision*, 1-8.*
19. Singh, V., & Kumar, A. (2020). Automatic License Plate Detection in Hazardous Condition. *IEEE Transactions on Image Processing*, 29(3), 1234-1244.
20. Gupta, P., & Rathi, S. (2017). Implementation of Image Processing in the Recognition of Official Vehicle License Plates. *Journal of Computer Vision and Image Processing*, 22(4), 75-89.
21. Xu, L., & Li, Y. (2021). License Plate Detection with Shallow and Deep CNNs in Complex Environments. *IEEE Transactions on Neural Networks and Learning Systems*, 32(6), 2708-2720.
22. Zhang, T., & Lin, J. (2019). License Plate Recognition for Moving Vehicles Case: At Night and Under Rain Condition. *Proceedings of the IEEE International Conference on Computer Vision*,2140-2148.
23. *Singh, K., & Sharma, N. (2018). Automatic License Plate Recognition (ALPR): A State-of-the-Art Review.* International Journal of Computer Applications*, 179(18), 13-21.*
24. *Wu, H., & Liu, X. (2022). Vehicle License Plate Recognition.* Journal of Machine Learning Research*, 23, 1-25.*

ISSN: 2582-8118 *Volume 4, Issue 5; Sept 2024*

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYSIS IN SCIENCE AND ENGINEERING

Web: https://www.iarj.in/index.php/ijrase/index

85. Comparative Study on Automated Kidney Stone Detection using Different Techniques

Rubina
Dept.Computer Science and Engineering
Basaveshwar Engineering College, Bagalkot, India.

Dr. Vishwanath Kagawade
Dept. Artificial Intelligence and Machine Learning
Basaveshwar Engineering College, Bagalkot, India.

ABSTRACT:

Present lifestyles and food habits are the main reasons for the evolution of human diseases. One of the common and fastest growing diseases is kidney stones in male and female of any age. Diagnosing a kidney stone requires a physical examination followed analysis using medical equipment's. Even the medical history of the patient under observation has to be taken into consideration by the physician. Typically testing of urine sample (urinalysis), an abdominal x-ray, a Computerized Tomography (CT) scan, or an ultrasound will be done for the diagnosis. Medical diagnostic procedures are costly and time consuming. So, in order to help the doctors and also the patients, the automated kidney stone detection models are implemented using deep learning models. These deep learning models are accurate, efficient and robust. In this paper variety of techniques that uses different imaging modalities for detecting kidney stones have been discussed and compared in detail for better understanding and clarity.

KEYWORDS:

Kidney Stone, Computerized Tomography (CT), Ultrasound, X-ray.

I. Introduction:

Kidney stones are also known as renal calculi, nephrolithiasis orurolithiasi. There will be deposition of minerals and acid salts inside the kidney which further leads to formation of kidney stones. The other Factors which increase the risk of stone formation include diabetes, obesity, lifestyle, hypertension and food habits. The most basic method used by medical professionals for diagnosis is usage of imaging techniques with the help of medical equipment's. After the diagnosis, the treatment may include medication or surgical corrections.

Kidney stone may recur after treatment and may develop into life risking chronic health conditions. Kidney stones cause huge pain and health imbalance. Kidney malfunctions can be life-threatening because the kidney is responsible for filtering toxicants from the blood and eliminate them through the urine. The kidney stone diagnosis is the most challenging task. Various imaging methods have to use for evaluation of disease at different stages of the kidney stone detection. Imaging techniques used can be Ultrasonography or Ultrasound (sonography), Computed Tomography (CT), and Kidney-Ureter-Bladded (KUB) X-ray imaging. The safest and quick method which can provide valuable evidence among all imaging procedure is Sonography. The sonography can be a confirmative diagnosis along with Computed Tomography (CT). Considering the complexity and seriousness of the disease there is an urgent need to address the problem not only from the medical perspective but also from the technical perspective. There is a chance that low-quality images are being produced by the traditional imaging techniques and hence making interpreting results or diagnosis much difficult. Such limitations have made a way for exploration of new possible approaches. There is a considerable increase in the quality of research using Machine Learning (ML) and Deep Learning (DL) for automated disease detection. The deep leaning methods facilitate the doctors as well as patients for confirmation of disease. For kidney detection using deep learning method involves 4 basic stages. Image acquisition followed by Image processing and feature extraction and finally classification. Deep learning-based methods help avert the percentage of inaccuracy that may be due to human errors or wrong diagnosis. Kidney images are the input for the different deep learning models. Deep learning techniques detect and localize the kidney stones. The steps involved in kidney stone detections are elaborated in Fig.1 for better understanding of the process.

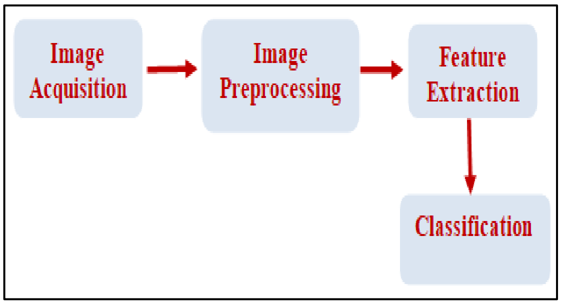


Figure.1 Generic Procedure for Kidney Stone Detection

A. Image Acquisition:

The process involves capturing of an image using CT images, MRI or sonography or images may be importing an existing image already captured into a computer. Some of the most common imaging techniques are as follows*:*

- **Intravenous pyelogram (IVP):** Excretory urogram or Intravenous pyelogram or intravenous urogram is a method getting radiographic imaging of a patient who is suspected for kidney stones. X-rays are used to create images that not only show the shape and size of the kidney but also the presence of kidney stones/tumours.
- **Ultrasound:** For the medical diagnosis the sound waves of frequency above 20 KHz which are not audible to human ears are used. The examination of all ureters and bladder

is done using this non-invasive (treatment which does not involve cutting of body) method. The method is safe for the pregnant woman and patients who cannot handle x-rays. Low cost, rapid examination and diagnosis is possible. The quality of Ultrasound images will be of poor quality, blurry with low contrast. This may be disrupting during segmentation process.

- **Magnetic Resonance Imaging (MRI):** The anatomical and clear statistics of the kidney can be obtained by MRI. The MRI images can accurately visualize the kidney's condition and its structures showing the places like renal cortex, medulla, and pelvis. MRI can also detect kidney stones and tumours. High cost is one of the disadvantages.
- **Computed Tomography (CT scan):** A narrow beam of x-rays is aimed at the patient. X-rays are rotated rapidly around the patient's body. The process produces signals that are captured and processed by the connected computer. The process is basically forming computerized image with the help of x-rays.

B. Image Pre-Processing: The images have to be pre-processed before fed to processing. Both image enhancement and Image restoration are done at this stage.

i. Image enhancement: Image quality can be increased by increasing contrast and reducing noise.
ii. Image restoration: Degradation is removed from an image. Examples include image blurring and distortion.

C. Image Segmentation: Image is divided into segments. Each segment represents an object or a particular feature of an image. The task of segmentation is to make the image look simpler, meaningful and easier for analysis. Types of methods include Edge based, ANN based, Fuzzy based, Region based and threshold based.

D. Feature Extraction: The process in which features are extracted. While choosing a feature extractor the necessary points to be considered are availability of data, dimensions of data, data label and choice of algorithm (supervised or unsupervised). Figure. 2 gives a brief introduction of different methods of feature extraction techniques.

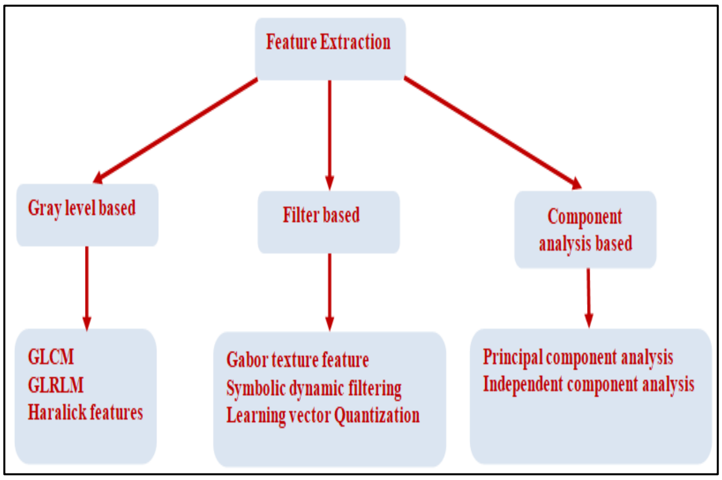


Figure. 2 Methods of Feature Extraction

E. Classification:

The process in which the model tries to predict correct label for a given input. Examples for classifications algorithm include ResNet-41, CNN, U-Net architecture and VGG 16.

II. Literature Review:

Paper [1] introduced a modified VGG16 model to identify kidney stones in KUB x-ray images. The model uses transfer learning (TL) method and uses a pre-trained VGG16 with explainable artificial intelligence (XAI) and Layer-Wise Relevance Propagation (LRP) technique to implement a system. The system takes KUB X-ray images and accurately categorizes them as kidney stones or normal cases. This combination helped in increased model fairness and transparency. In paper [2] Detection of kidney and Estimation of its size using Image Segmentation Techniques proposed two stage process on CT scan images. In stage1 median filter, top hat filter, morphological erosion and dilation methods to remove noise and improve image quality.

Segmentation is performed using thresholding, watershed methods. Feature extraction is achieved using binary masking and K-means algorithms. Six parameters Structure Similarity Index Measure (SSIM), Peak Signal-to-Noise Ratio (PSNR), Entropy, Feature Similarity Index Measure (FSIM), Normalized Cross Correlation (NCC) and Normalized Absolute Error (NAE) to check the set of CT images. Paper [3], proposed deep learning CNN model for the detection of kidney stone. The work localizes the area of kidney stone in CT image by using coronal CT images. 1799 images were utilised by capturing different cross-sectional CT images of each person. An accuracy of 96.82% is achieved in detecting stones.

Paper [4] proposed medical image processing methods like HU windowing, image screening, and labelling methods for improving segmentation performance. The model uses U-Net architecture, ResNet-41 or Efficient Net as an encoder and Feature Pyramid Network (FPN) as a decoder. Optimization is achieved by windowing selection, data screening, and labeling selections to obtain the optimal hyper parameters.

This encoder and decoder combinations gave best performance which is validate by five-fold cross-validation processes for evaluating the accuracy of kidney and tumor detection. Paper [5] proposed a method known as MBA Net architecture for improving accuracy of segmentation for kidney ultrasound (KUS) images. MBANet model uses context information to segment the kidneys correctly. An architecture composed of multi-scale feature pyramid (MSFP), multi-branch encoders (MBE) and master decoder is used to address the challenges of KUS images segmentation.

Paper [6] presented a three-branch convolutional neural network for detection of renal parenchyma within an MR image patch. Paper [7] implemented Transfer learning network to segment kidneys using, and boundary distance regression and pixel classification networks. Distance loss function is used by the Boundary Detection Network (Bnet) for training. A combination of distance loss function and a SoftMax loss function is utilised by segmentation networks.

Paper [8] proposed a convolutional neural network architecture. The work proves that the images taken in sagittal-plane on CT device have a higher diagnostic accuracy rate in comparison to the images taken on other planes.

The xResNet50 architecture that consists of an input stem consisting of four xResNet50 blocks, and an output stem. Input stem accepts input and the model processes the input to produce output in the output stem. The result obtained represents the percentage of kidney stone presence or absence. Paper [9] developed a 3D U-Net model was employed to segment the kidneys. It is further followed by gradient-based anisotropic de-noising, thresholding, and region growing. Paper [10] addressed a transfer learning-based image classification method (ExDark19) is proposed to detect kidney stones using CT images. Paper [11] implemented automatic segmentation algorithm. Images are pre-processed using contorourlet during stage1. Later the level set method is used for automatic segmentation of pre-processed images. The metrics such as number of stones and their sizes is obtained.

After a rigorous survey, it is observed that various works are carried out to detect kidney stones in various dimensions. Table.1 shows the latest techniques used to detect the kidney stones along with the accuracy they achieved during kidney stone detection.

Table 1: Comparative Studies in Kidney Stone Detection

Authors	Dataset used	Type of Dataset	Technique	Accuracy obtained	Rate of misclassification
Chiang et al.	Number of healthy samples-105 Number of patient samples-151(calcium oxalate stones)	Handcrafted	ANN and DA	ANN accuracy-89%, DA accuracy-75%	ANN-11% DA-25%
Dussol et al.	Number of healthy samples-96 Number of patient samples-119	Handcrafted	ANN (LDA and MVDA)	ANN + LDA - 75.8% ANN + MVDA - 74.4%	LDA: 24.2% MVDA: 25.6%
Cauderella et al.	Number of patient samples-80	Handcrafted	ANN and LR	ANN - 88.8% LR - 67.5%	ANN: 11.2% LR: 32.5%
De Perrot et al.	Number of patient samples- 416	Handcrafted	ML	ML-85.1%	14.9%
Kahani et al.	KUB x-ray images-100	Image-based	LASSO with ML classifiers	LASSO + ML-96%	LASSO + ML - 4%

Authors	Dataset used	Type of Dataset	Technique	Accuracy obtained	Rate of misclassification
Kumar et al.	Number of patient samples-936	Handcrafted	LVQ, MLP, and RBF	LVQ-84%, MLP-92%, and RBF-87%	LVQ: 16% MLP: 8% RBF: 13%
Ebrahimi et al.	KUB CT scan slides of 39 patients	Image-based	Image processing techniques with geometry principles	84.61%	15.39%
Annameti Rohith et al.	114 ultrasound images	Image-based	Median and rank filters	Median filter-86.4% Rank filter-82.2%	Median filter-13.6%, Rank filter-17.8%
Jungmann et al.	1714 LDCT images	Image-based	NLP	72%	28%
Kazemi et al.	Numeric characteristics from 936 patients	Handcrafted	Ensemble learning model	97.1%	2.9%
Authors	**Dataset used**	**Type of Dataset**	**Technique**	**Accuracy obtained**	**Rate of misclassification**
Suresh et al.	KUB ultrasound images	Image-based	Image processing techniques	92.57%	7.43%
Li et al.	1874 CT KUB reports	Handcrafted	NLP	85%	15%
Yildirim et al.	1799 coronal CT images	Image-based	XResNet-50	96.82%	3.18%
Valencia et al.	CT scans of 40 patients	Image-based	Image processing	92.5%	7.5%
Tsitsiflis et al.	Medical data of 716 patients	Feature-based	ANN	81.43%	18.57%
Cui et al.	625 CT images	Image-based	DL + threshold-based model	90.30%	9.70%
Fahad Ahmed et al	14,265 KUB x-ray images	Image-based	VGG16 + LRP	97.41%	2.59%

III. Comparison of Results:

It is observed that different approaches or techniques have been proposed for kidney stone detection and achieved better accuracy. Figure 3 shows the bar chart constitutes the techniques applied and the percentage of accuracy achieved.

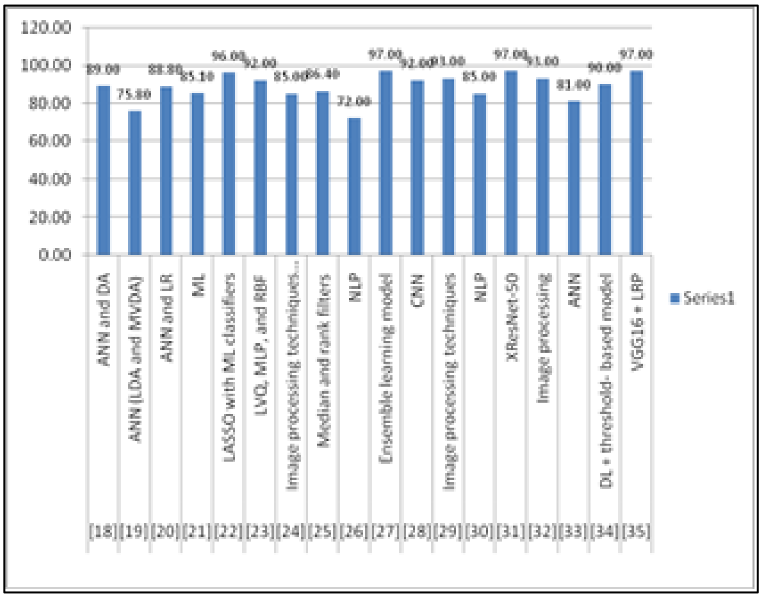


Figure 3: Results

IV. Challenges:

The challenges that may encounter for kidney stone detection are listed as below.

A. Unavailability of diverse kidney stone dataset:

As per 2019 National Medical Commission Act, any doctor must keep patient data confidential. It is much difficult to get vast dataset with such a restriction. Even if the data is available, the data set may not be having diverse or variety of possible data. Absence of variety of data will definitely affect the learning process of the model and hence the prediction accuracy may be reduced as well. Some rules and regulations exist pertaining to handling the digital health data, the same is regulated by Digital Information Security in Healthcare Act (DISHA). The act regulates the way in which health data is collected, stored, transmitted and used to ensure reliability.

B. Diverse Imaging modality:

Images may be generated by measuring the changes in energy. The change in energy may be due to absorption, scatter or refraction. The type and frequency of energy employed decides the depth of image, contrast and spatial resolution.

Such characteristics will act like the deciding factor for choosing a particular type of deep learning model for a specific scenario of kidney stone detection. Specific algorithms and methods used may change depending on the type of medical imaging and the equipment used for image acquisition.

C. Processing power:

Deep learning models need bigger data set for producing accurate results. Processing of such high quantity of data set surely requires huge processing power. The availability of such processing power will be available in well-established industries and highly equipped research laboratories. It is little difficult to have access to such resources.

D. Dependencies on medical practitioners/ knowledge about medical terms:

The authenticity and accuracy of the results produced by research model have to certified by the medical practitioners/experts. Any misinterpretation or human error by the medical practitioners may lead to wrong progress of the model/system under development.

E. Enhancement in imaging modality for evaluation and treatment of kidney stones:

In order to get high clarity and quality images, there is a need for high standard/ quality medical equipment's to be used for image capturing. But, availability of such medical equipment's is not possible in all places of diagnosis. Hence, the quality of image followed by the diagnosis is compromised. Low quality images affect diagnosis which in turn affect treatment

F. Building a technique that works on heterogeneous data type:

Building a heterogeneous model requires expertise. Obtaining an expertise implies that the expert is well versed with all the latest techniques, algorithms and trends. Technically such an ideal scenario will not exist.

V. Discussions:

CNNs have shown promising performance during classification, segmentation, localization, and detection in various medical images, thereby supporting doctors for increased accuracy in disease diagnosis.

The model can perform well for good quality of the training data and hence quality of data is also one prime factor for deciding accuracy of the model.

Depending on the characteristics of distinct medical images, proper pre-processing methods such as de-noising and contrast enhancement are selected to eliminate noise and suppress interference. Another preliminary step is intensity normalization, in which the intensity distribution is transformed or filtered and set in a specific range to improve the optimization. The pre- processing methods and deep learning models are big challenges.

VI. Conclusion:

The major challenge for research is the availability of high-quality, clarity and diversity/variety of medical image or picture of kidney stones. The limitations can be overcome by continued efforts to collect, curate, and make a broader range of medical image or picture more readily accessible to improve model performance. Deep learning models outperform other models and produce reliable and efficient ways for kidney stone detection. Images taken by Sagital-plane on CT have higher accuracy as compared to images taken using other planes. Deep learning methods assist in early detection of the disease and can reduce diagnostic cost and time. Traditional medical diagnostic procedures involve time and money. Survey data revealed that the cost involved in performing a CT scan is approximately double that of a renal ultrasound scan and approximately one third that of an MRI. This situation leads AI models to be implemented for reduced cost, high efficiency, and for deciding among imaging preference choices.

References:

1. Ahmed et al., "Author Correction: Identification of kidney stones in KUB X-ray images using VGG16 empowered with explainable artificial intelligence," Scientific reports, vol. 14, no. 1, May 2024, doi: https://doi.org/10.1038/s41598-024-60554-0.
2. None Shraman Jain, None Rajini G. K, None Rahul S.G, None Rajkumar R, None Velmurugan S, and J. Pemeena, "Detection of Kidney Stone and Estimation of its Size using Image Segmentation Techniques," Journal of Advanced Research in Applied Sciences and Engineering Technology, vol. 36, no. 2, pp. 91–109, Dec. 2023, doi: https://doi.org/10.37934/araset.36.2.91109.
3. K. Yildirim, P. G. Bozdag, M. Talo, O. Yildirim, M. Karabatak, and U. Rajendra. Acharya, "Deep learning model for automated kidney stone detection using coronal CT images," Computers in Biology and Medicine, vol. 135, p. 104569, Aug. 2021, doi: https://doi.org/10.1016/j.compbiomed.2021.104569.
4. C.-H. Hsiao et al., "A deep learning-based precision volume calculation approach for kidney and tumor segmentation on computed tomography images," Computer methods and programs in biomedicine, vol. 221, pp. 106861–106861, Jun. 2022, doi: https://doi.org/10.1016/j.cmpb.2022.106861.
5. Chen, Y. Dai, J. Zhang, X. Yin, and L. Cui, "MBANet: Multi-branch aware network for kidney ultrasound images segmentation," Computers in Biology and Medicine, vol. 141, p. 105140, Feb. 2022, doi: https://doi.org/10.1016/j.compbiomed.2021.105140.
6. Klepaczko, M. Majos, L. Stefańczyk, E. Eikefjord, and A. Lundervold, "Whole kidney and renal cortex segmentation in contrast-enhanced MRI using a joint classification and segmentation convolutional neural network," Biocybernetics and Biomedical Engineering, vol. 42, no. 1, pp. 295–311, Jan. 2022, doi: https://doi.org/10.1016/j.bbe.2022.02.002.
7. S. Yin et al., "Automatic kidney segmentation in ultrasound images using subsequent boundary distance regression and pixelwise classification networks," Medical Image Analysis, vol. 60, p. 101602, Feb. 2020, doi: https://doi.org/10.1016/j.media.2019.101602.

8. Caglayan, M. O. Horsanali, K. Kocadurdu, E. Ismailoglu, and S. Guneyli, "Deep learning model-assisted detection of kidney stones on computed tomography," International braz j urol, vol. 48, pp. 830–839, Aug. 2022, doi: https://doi.org/10.1590/S1677-5538.IBJU.2022.0132.
9. Chen, Z., Xiao, C., Liu, Y. et al. Comprehensive 3D Analysis of the Renal System and Stones: Segmenting and Registering Non-Contrast and Contrast Computed Tomography Images. Inf Syst Front (2024). https://doi.org/10.1007/s10796-024-10485-y
10. M. Baygin, O. Yaman, P. D. Barua, S. Dogan, T. Tuncer, and U. R. Acharya, "Exemplar Darknet19 feature generation technique for automated kidney stone detection with coronal CT images," Artificial Intelligence in Medicine, vol. 127, p. 102274, May 2022, doi: https://doi.org/10.1016/j.artmed.2022.102274.
11. P. T. Akkasaligar and S. Biradar, "Automatic Segmentation and Analysis of Renal Calculi in Medical Ultrasound Images," Pattern Recognition and Image Analysis, vol. 30, no. 4, pp. 748–756, Oct. 2020, doi: https://doi.org/10.1134/s1054661820040021.
12. N. Sharma et al., "Automated medical image segmentation techniques," Journal of Medical Physics, vol. 35, no. 1, p. 3, 2010, doi: https://doi.org/10.4103/0971-6203.58777.
13. C. Elton, E. B. Turkbey, P. J. Pickhardt, and R. M. Summers, "A deep learning system for automated kidney stone detection and volumetric segmentation on noncontrast CT scans," Medical Physics, Feb. 2022, doi: https://doi.org/10.1002/mp.15518.
14. G, P. Sai G B and J. J. Hilda, "Automated Kidney Stone Detection using Deep Learning Technique," 2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE), Vellore, India, 2024, pp. 1-6, doi: 10.1109/ic-ETITE58242.2024.10493536.
15. M. Shetty, S. B. Shetty, G. P. Sequeria and V. Hegde, "Kidney Stone Detection Using CNN," 2024 Second International Conference on Data Science and Information System (ICDSIS), Hassan, India, 2024, pp. 1-6, doi: 10.1109/ICDSIS61070.2024.10594342.
16. Nestler, T., Stoll, R., Schmelz, H. et al. Comparison of automated kidney stone size measurement and volumetry in photon counting CT compared to 3rd generation dual energy CT and physically measurements – an ex vivo study. World J Urol 42, 433 (2024). https://doi.org/10.1007/s00345-024-05114-x
17. S. Asif, X. Zheng, and Y. Zhu, "An optimized fusion of deep learning models for kidney stone detection from CT images," Journal of King Saud University. Computer and information sciences/Maǧalaẗ ǧam'aẗ al-malīk Saud: ùlm al-ḥasib wa al-ma'lumat, pp. 102130–102130, Jul. 2024, doi: https://doi.org/10.1016/j.jksuci.2024.102130.
18. N, B. S, D. p, D. L and H. P, "Deep Learning Model for Automated Kidney Stone Detection using VGG16," 2023 Second International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2023, pp. 1158-1164, doi: 10.1109/ICEARS56392.2023.10085509.
19. Chiang, H.-C. Chiang, W.-C. Chen, and F.-J. Tsai, "Prediction of stone disease by discriminant analysis and artificial neural networks in genetic polymorphisms: a new method," BJU International, vol. 91, no. 7, pp. 661–666, Apr. 2003, doi: https://doi.org/10.1046/j.1464-410x.2003.03067.x.
20. Bertrand Dussol, J.-M. Verdier, Jean-Marc Le Goff, P. Berthezene, and Y. Berland, "Artificial neural networks for assessing the risk of urinary calcium stone among men," *Urological research*, vol. 34, no. 1, pp. 17–25, Jan. 2006, doi: https://doi.org/10.1007/s00240-005-0006-4.

21. Caudarella R, Tonello L, Rizzoli E, Vescini F. Predicting five-year recurrence rates of kidney stones: an artificial neural network model. Arch Ital Urol Androl. 2011 Mar;83(1):14-9. PMID: 21585163.
22. K. Kumar and A. Abhishek, "Artificial Neural Networks for Diagnosis of Kidney Stones Disease," International Journal of Information Technology and Computer Science, vol. 4, no. 7, pp. 20–25, Jul. 2012, doi: https://doi.org/10.5815/ijitcs.2012.07.03.
23. "Journal of Image and Graphics," Journal of Image and Graphics, Mar. 2023, doi: https://doi.org/10.18178/joig.
24. Y. Kazemi and S. A. Mirroshandel, "A novel method for predicting kidney stone type using ensemble learning," *Artificial Intelligence in Medicine*, vol. 84, pp. 117–126, Jan. 2018, doi: https://doi.org/10.1016/j.artmed.2017.12.001.
25. Y. Li and N. Elliot, "Natural language processing to identify ureteric stones in radiology reports," *Journal of Medical Imaging and Radiation Oncology*, vol. 63, no. 3, pp. 307–310, Feb. 2019, doi: https://doi.org/10.1111/1754-9485.12861.
26. De Perrot, T. et al. Differentiating kidney stones from phleboliths in unenhanced low-dose computed tomography using radiomics and machine learning. Eur. Radiol. 29, 4776–4782 (2019).
27. Kahani, M., Hariri Tabrizi, S., Kamali-Asl, A. & Hashemi, S. A novel approach to classify urinary stones using dual-energy kidney, ureter and bladder (DEKUB) X-ray imaging. *Appl. Radiat. Isot.* **164**(May 2019), 109267. https:// doi. org/ 10. 1016/j. aprad iso. 2020.109267 (2020).
28. Jungmann, F. *et al.* Towards data-driven medical imaging using natural language processing in patients with suspected urolithiasis.*Int. J. Med. Inform.* **137**, 1–21. https:// doi. org/ 10. 1016/j. ijmed inf. 2020. 104106 (2020).
29. Rohith, A. Detection of kidney stones in ultrasound images using median filter compared with rank filter. *Rev. Gestao Inovacao eTecnol.* **11**(4), 1096–1111. https:// doi. org/ 10. 47059/ revis tagei ntec. v11i4. 2171 (2021).
30. Suresh, M. B. & Abhishek, M. R. Kidney stone detection using digital image processing techniques. In *Proceedings of the 3rd International Conference on Inventive Research in Computing Applications, ICIRCA 2021, Coimbatore, India* 556–561 (IEEE, 2021).
31. https:// doi. org/ 10. 1109/ ICIRC A51532. 2021. 95446 10.
32. Jendeberg, J., Thunberg, P. & Lidén, M. Differentiation of distal ureteral stones and pelvic phleboliths using a convolutional neural network. *Urolithiasis* **49**(1), 41–49. https:// doi. org/ 10. 1007/ s00240- 020- 01180-z (2021).
33. Cui, Y. *et al.* Automatic detection and scoring of kidney stones on noncontrast CT images using S.T.O.N.E. nephrolithometry: Combined deep learning and thresholding methods. *Mol. Imaging Biol.* **23**(3), 436–445. https:// doi. org/ 10. 1007/ s11307- 020-01554-0 (2021).
34. Yildirim, K. *et al.* Deep learning model for automated kidney stone detection using coronal CT images. *Comput. Biol. Med.* **135**,1–7. https:// doi. org/ 10. 1016/j. compb iomed. 2021. 104569 (2021).
35. Tsitsiflis, A. *et al.* The use of an artificial neural network in the evaluation of the extracorporeal shockwave lithotripsy as a treatment of choice for urinary lithiasis. *Asian J. Urol.* **9**(2), 132–138. https:// doi. org/ 10. 1016/j. ajur. 2021. 09. 005 (2022).
36. Valencia, F. A. C., Muñoz, J. J. A. M. & Montaño, F. M. J. Stone detection in kidney with image processing technique: CT images. *J. Posit. Sch. Psychol.* **6**(6), 7643–7653 (2022).

37. Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C. & Liu, C. A survey on deep transfer learning. In *27th International Conference on*
38. *Artificial Neural Networks, Rhodes, Greece* 270–279 (Springer, 2018). https:// doi. org/ 10. 1007/ 978-3- 030- 01424-7.
39. Ahmed et al., "Author Correction: Identification of kidney stones in KUB X-ray images using VGG16 empowered with explainable artificial intelligence," Scientific reports, vol. 14, no. 1, May 2024, doi: https://doi.org/10.1038/s41598-024-60554-0.

ISSN: 2582-8118 *Volume 4, Issue 5; Sept 2024*

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYSIS IN SCIENCE AND ENGINEERING

Web: https://www.iarj.in/index.php/ijrase/index

86. AI in Robotics: Objectives, Innovations, and Ethical Implications

Dr. Shivanand Pujar

Associate Professor, HOD CSE,
Computer Science and Engineering Department
Bharat Ratna Indira Gandhi College of Engineering,
Solapur, Maharashtra India.

Prof. Shaik Mohammadsamir

Assistant Professor, Computer Science and Engineering Department
Bharat Ratna Indira Gandhi College of Engineering,
Solapur, Maharashtra India.

Prof. N. Ramesh Babu

Associate Professor, Electronics and
Communication Engineering Department, J.B. Institute of Engineering &
Technology, Hyderabad Telangana, India.

Dr. Doddavarapu Sreenu Babu

Professor, Electronics and Telecommunication Department
Bharat Ratna Indira Gandhi College of Engineering,
Solapur, Maharashtra India.

ABSTRACT:

Artificial Intelligence (AI) represents the frontier of technological innovation, where machines are endowed with the ability to simulate human intelligence. It encompasses a diverse array of techniques and methodologies aimed at enabling computers to perceive, comprehend, and act upon data in ways that mimic human cognitive functions.

From speech recognition and natural language processing to visual perception and decision-making, AI is revolutionizing industries and everyday life. Its applications span from virtual assistants that respond to voice commands to advanced systems that autonomously navigate vehicles and diagnose medical conditions. As AI continues to evolve, its potential to transform industries such as healthcare, finance, and transportation becomes increasingly evident, promising unprecedented efficiency gains and ethical considerations as society navigates this new era of intelligent machines.

KEYWORDS:

Expert Systems, Limited Theory, Artificial Intelligence, Robotics

I. Introduction:

Creating systems capable of doing activities that normally require human intelligence is the goal of artificial intelligence (AI), a fast-developing branch of computer science. Making decisions, identifying patterns, gaining knowledge from experience, and comprehending natural language are some of these duties.

Symbolic thinking, logic, statistical procedures, machine learning, and other methodologies are all in the broad category of artificial intelligence (AI).

Artificial Intelligence (AI) aims to create machines that can mimic human thought processes and thought patterns. Planning, reasoning, perception, problem-solving, and language comprehension are all included in this.

The uses of AI are many and include everything from medical diagnostics to self-driving automobiles to mastering sophisticated games like Go or Chess.

Key subfields within AI include:

1. Algorithms used in machine learning allow computers to learn from data and gradually become more efficient over time without the need for explicit programming.
2. Natural Language Processing (NLP): Methods that allow computers to comprehend, interpret, and produce human language.
3. Computer Vision: Techniques for giving computers the ability to understand visual data from the outside environment, such as pictures and films.
4. Robotics: Automating jobs in the physical environment by fusing artificial intelligence with mechanical mechanisms.
5. Expert Systems: AI systems that mimic a human expert's decision-making process in a certain field.

AI has the power to transform entire sectors, increase productivity, and find solutions to challenging issues that were previously thought to be unsolvable. It also brings up moral and societal issues, such as the consequences of autonomous systems, algorithmic biases, and the loss of jobs due to automation.

Researchers are investigating novel techniques like deep learning, neural networks, and reinforcement learning as AI develops further to build more powerful and intelligent systems. This is a dynamic and quickly growing topic, and its future social influence is being shaped by continuing discussions and advancements.

In computer science and engineering, artificial intelligence (AI) is a rapidly evolving area that aims to build intelligent computers that can carry out tasks that traditionally require human intelligence.

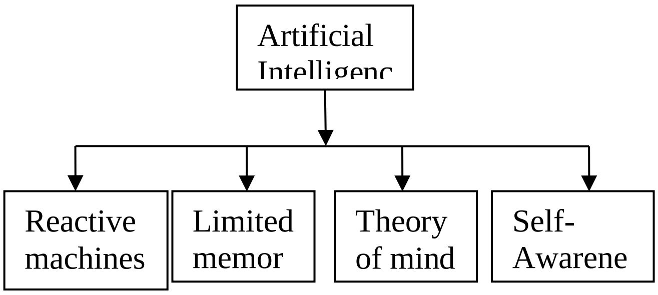
These tasks cover a broad range of skills, such as comprehending plain language, identifying patterns in data, picking up knowledge from experience, using reasoning to make decisions, and even operating physical devices on their own.

Evolution and Origins:

The notion of artificial intelligence (AI) has its roots in ancient philosophy and mythology when the creation of artificial entities possessing human-like capacities was discussed. Still, as breakthroughs in mathematics, computer science, and cognitive psychology converged, the formal subject of artificial intelligence (AI) arose in the middle of the 20th century.

Important individuals like Alan Turing, who is frequently referred to as the father of artificial intelligence, established fundamental theories and coined the well-known Turing Test to gauge machine intelligence

Types of AI:



a) Reactive Machines:

Reactive machines take in information from the outside world and organize their actions accordingly. The machines are only able to comprehend the work at hand and carry out specific functions.

Under repeated conditions, the machines behave consistently. IBM created Deep Blue, a reactive chess player that could anticipate moves by figuring out where each piece should go on the board, in the 1990s.

b) Limited Memory:

With little memory using recent observations, machines can make deft decisions. Concerning their pre-programmed conceptual framework, the machines take into account observational data.

After being kept for a brief while, the observational data is subsequently deleted.

c) Theory of Mind:

Concept of mind to engage in social contact, computers must be able to develop thoughts and make decisions regarding emotional context. Although many of the machines still need to be developed, they already show certain human-like traits. Voice assistant apps, for instance, are unable to carry on a conversation but are capable of understanding simple spoken commands and instructions.

d) Self-Awareness:

By generating ideas, forming wants, and comprehending their internal states, self-awareness robots exhibit intelligent behavior. To find machines that might behave indistinguishably from humans, Alan Turing created the Turing Test in 1950.

While There Are Many Advantages to Artificial Intelligence (AI), There Are Also Several Difficulties and Disadvantages.

a) Fairness and Bias: AI systems have the potential to produce biased results based on the biases in the data they are trained on, particularly in delicate domains such as law enforcement, lending, and employment.
b) AI-driven automation may result in job displacement in some industries by automating tasks, which would affect employment in manufacturing, customer service, and transportation, among other sectors.
c) Privacy Concerns: AI systems frequently rely on enormous volumes of personal data, which raises concerns about data breaches and the misuse of private information.
d) Lack of Transparency: A lot of AI models, particularly deep learning models, function as "black boxes," making it challenging to comprehend how they make judgments. It can be troublesome when there is a lack of openness, particularly in industries like law and healthcare.
e) Security Risks: Assaults and malevolent usage of AI systems are possible; examples include adversarial assaults on image recognition systems and the manipulation of AI-driven autonomous vehicles.
f) Ethical Issues: Discussions concerning the moral implications of AI are still going on. These include questions regarding responsibility, moral judgment on the part of autonomous systems, and the possibility of abuse in surveillance or combat.
g) Over-reliance on AI: If AI systems are relied upon for crucial jobs and they malfunction or are corrupted, there could be systemic dangers.
h) Social Impact: For example, by enforcing already-existing disparities or generating new ones between those who have access to AI and those who do not, AI technologies have the potential to worsen social divisions.
i) Environmental Impact: It takes a lot of computer power to train complex AI models, which can increase carbon emissions and harm the environment.
j) Regulatory Challenges: AI is developing faster than most regulatory frameworks, which presents problems for safety standards, liability, and governance, among other areas.

Thorough evaluation of ethical standards, strict regulation, openness in AI research, and continual investigation into reducing prejudices and hazards related to AI technology are all necessary to address these shortcomings

Applications of AI:

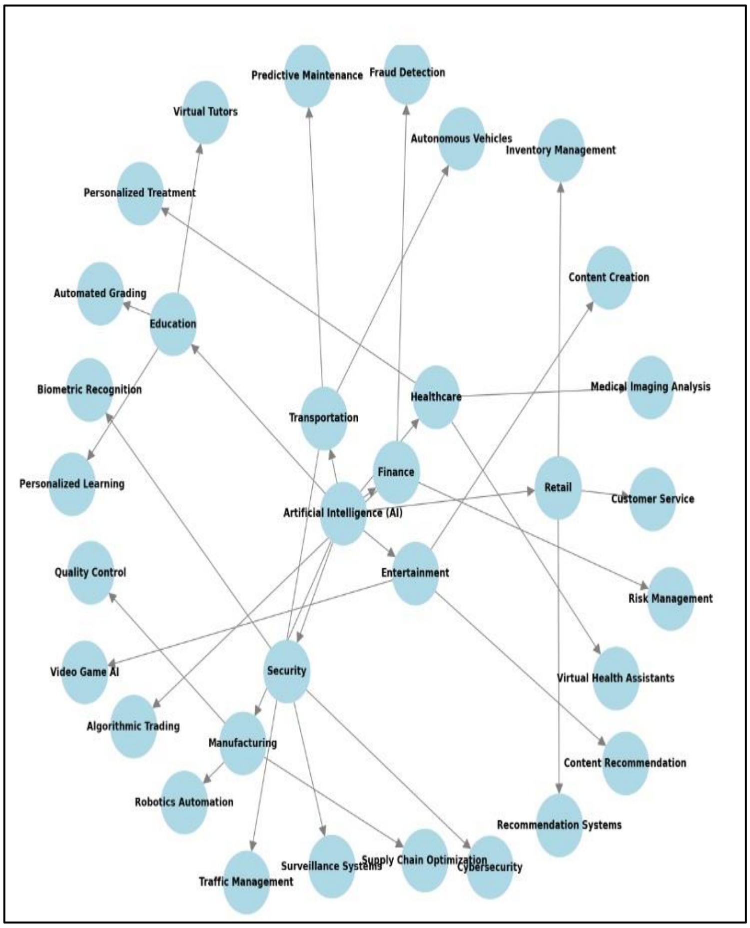


Figure 1: Real World Examples of AI

a) Healthcare: Medical imaging, tailored medicine, and diagnosis.
b) Finance: Risk assessment, fraud detection, and algorithmic trading.
c) Transportation: autonomous cars and traffic control.
d) Education: Adaptive learning systems, personalized tutoring.
e) Retail: Customer service, demand forecasting, recommendation systems.
f) Ethical Considerations.
g) Ethical frameworks: Principles for developing and deploying AI responsibly.
h) Accountability and Liability: Challenges in assigning responsibility for AI actions.
i) AI and Human Rights: Implications for privacy, equality, and social justice.
j) Dual-Use Dilemmas: Ethical concerns in military, surveillance, and autonomous weapon systems.
k) Societal Impacts.
l) Economic Disparities: Potential exacerbation of inequalities.
m) Social Division: Access to AI technologies and digital divides.
n) Cultural and Behavioral Shifts: Changing perceptions and societal norms.

II. Literature Survey:

Current advanced applications make extensive use of artificial intelligence (AI); nevertheless, because many AI models are opaque, it can be difficult to understand and rely Current advanced applications make extensive use of artificial intelligence (AI); nevertheless, because many AI models are opaque, it can be difficult to understand and rely on their results. Typically, it is critical to comprehend the logic that underlies an AI model's judgment. Consequently, there is a need for explainable AI (XAI) techniques to increase user confidence in AI models. In the field of AI, XAI has gained popularity as a study topic in recent years. There are currently no reviews that have examined the assessment techniques, available resources, XAI datasets, and other relevant topics. Previous survey publications have addressed the principles of XAI, its general terminologies, and post-hoc explain ability methodologies. So, using a case study example, we give readers an overview of the latest findings and trends in this quickly developing field in this extensive study. Next, XAI concerns—legal requirements, user perspectives, and application orientation—are discussed about the importance of explain ability [2].

With so many different things being searched for online, it is becoming increasingly necessary in today's society to use e-commerce materials like music, movies, and electronic goods. The implementation of filtering strategies enables the item search to yield relevant results since it identifies pertinent information for item recommendations. Instead of accessing all of the data on the internet to obtain related results, a variety of filtering strategies are available for the data. Finding suitable results based on user choices is a difficult task that requires both efficiency and effective data access. The main benefits and implementation procedures of the suggested Constrained Social Network Rating-based Neural Network Technique (CSNR-NNT) are discussed in this study [8].

The research results on the use of AI techniques by OR Society members are presented in this publication. A total of 26% of the UK-based members of the OR Society responded to the poll, which comprised mailing a questionnaire to one-third of the members. The poll examined the general characteristics of the participants to allow for comparison with earlier surveys; it also identified the sectors in which artificial intelligence (AI) techniques are now and increasingly being used, as well as their advantages and disadvantages [17].

An international outbreak of COVID-19 has resulted from the SARS-CoV-2 virus's quick global dissemination. The majority of businesses, governments, and institutions of higher learning are taking part in the COVID-19 effort to stop the pandemic's spread. Artificial intelligence (AI) technologies are frequently deployed as a potent weapon against COVID-19 in the fight against this pandemic. In this review, we examine the primary uses and contributions of AI in the fight against COVID-19 from the perspectives of illness detection and diagnosis, virology and pathogenesis, medication and vaccine research, and epidemic and transmission forecasting. Furthermore, we provide an overview of the information and materials that are currently accessible and useful for COVID-19 studies utilizing AI. This concludes the discussion of the primary obstacles and possible paths for AI in the fight against COVID-19. In an attempt to encourage researchers to keep maximizing the benefits of AI and big data in the battle against COVID-19, this survey provides medical and AI researchers with a thorough understanding of the current and future uses of AI technology in this regard [10].

Recent years have seen a notable increase in the use of generative AI, with several applications spanning a wide range of industries. This study offers a thorough analysis of over 350 generative AI applications, along with succinct explanations and a structured taxonomy of different unimodal and even multimodal generative AIs. A wide range of unimodal generative AI applications, including text, graphics, video, games, and brain information, are covered in the sections that make up the survey. To improve comprehension of the state-of-the-art and encourage more innovation in the area, our survey intends to be an invaluable tool for researchers and practitioners navigating the quickly changing world of generative AI [7].

Artificial intelligence research is always developing, and this study is by no means an exhaustive account of the advances achieved over the past ten years. AI encompasses a wide range of fields. Building powerful AI may benefit from much of the research discussed in this study. It is crucial to build a machine that can comprehend the ideas behind words because doing so would enable better translation and more conversational exchanges. Detecting human emotions using auditory and visual clues is another exciting area of research. This work offers a thorough analysis of current advancements in the field of artificial intelligence and its applications, in particular. Newcomers to the field of artificial intelligence are the intended audience for this book [15].

There have been numerous definitions of artificial intelligence's various forms published recently, and it has had a huge impact on every field. The majority of literary pieces' center on the remarkable powers of artificial intelligence. Recently, there have been reports of difficulties with the development of intelligent systems, including security, safety, fairness, robustness, and energy consumption. The number of new issues rises with the use of intelligent technologies. Naturally, perspectives on issues like security will shift as artificial intelligence progresses from narrow intelligence to super intelligence. Furthermore, it is improper to discuss the difficulties involved in creating intelligent systems without taking into account the recent advancements in human-level intelligence [2].

A large number of samples are needed for optimizing utilizing meta-heuristic optimization techniques. Specifically, there is a high number of individuals and little use of preexisting samples while producing a sub-generation. Thus, the development of surrogate-based optimization has resulted in a significant reduction in both the number of individuals in the sub-generation and the optimization cost. This research created a double-objective air supply optimization method based on the Kriging surrogate model and Non-Dominated Sorting Genetic Algorithms-II. Single-objective optimization results may not be comprehensive in complex air supply scenarios. The outcomes demonstrated that the approach can rapidly arrive at an approximate boundary condition prediction (the average error was 6.8% and there were 82 simulations when predictions were based on experimental data) [3].

The study of computational models of problem-solving, when the issues to be addressed are of the complexity of problems solved by humans, is known as artificial intelligence (AI) in the computing science domain. The field of artificial intelligence studies ways to teach computers to perform tasks that people now perform more effectively. The goal of this area of computer science is to develop artificial intelligence in machines. Artificial Intelligence is another term for the study and creation of intelligent agent systems.

Reasoning, knowledge, planning, learning, perception, communication, and object movement and manipulation are among the qualities that constitute the core issues with AI [14]. Automating tasks that now require human intelligence is the goal of the diverse field of artificial intelligence (AI). Artificial intelligence (AI) is a technology that is transforming all facets of life, despite its general lack of familiarity. This essay seeks to enlighten laypeople about artificial intelligence (AI) and inspire them to use it as a tool across a variety of disciplines to reconsider how we gather, process, and evaluate data. In this piece, we outlined artificial intelligence (AI), its functions, and some potential applications for it in daily life [4].

Intelligent machines will eventually supplement or perhaps surpass human abilities in a variety of domains. The intelligence displayed by software or robots is known as artificial intelligence. That is a branch of computer science. The discipline of computer science artificial intelligence is growing in popularity since it has improved human lives in many ways.

The performance of the industrial and service systems has significantly increased over the past 20 years thanks to artificial intelligence. Expert systems are a fast-expanding technology that is the result of research in the field of artificial intelligence. Because expert systems are now widely utilized to tackle complicated issues in a variety of domains, including science, engineering, business, medicine, and weather forecasting, artificial intelligence is having a significant impact on a wide range of areas of life [16].

Our lives have changed dramatically and will continue to do so as a result of artificial intelligence (AI). "Artificial intelligence" (AI) is the term used to describe a recent wave of technology advancements that may mimic human intelligence and interact with the surroundings. E. This paper aims to explore the foundations and history of artificial intelligence (AI), the current state of AI research, the development of AI, machine learning, and some of the applications of AI that are being used across multiple industries. Artificial Intelligence in Pharmaceuticals is also discussed in this article [6].

important and cutting-edge technologies of the modern era, has fundamentally changed the way people work, play, and live. From voice assistants like Siri and Alexa to self-driving cars and personalized recommendations, artificial intelligence (AI) has influenced many aspects of our everyday lives. But what precisely is AI, and how does it work? The creation of computer systems that can do tasks that typically require human intelligence is known as artificial intelligence. Automation of financial processes enhanced retail consumer experiences, improved healthcare diagnosis and treatment, and increased manufacturing productivity are all being achieved with artificial intelligence (AI)-driven solutions [1].

Artificial intelligence is advancing at an extremely innovative rate in the modern world. From developing tiny handheld gadgets for tiny tasks like face recognition to developing hard and strict abilities like driving a car. There was a time when the difficulty of performing routine tasks and physical labor inspired people to invent new technologies. This led to the development of artificial intelligence, or AI, with the lofty objective of imbuing machines with intelligence comparable to that of humans. The main goals of AI are the development of expert systems and the incorporation of human intellect into robots [13].

III. Objectives of AI towards Robotics:

AI is essential to the advancement of robotics because it serves multiple purposes that motivate the incorporation of AI into robotic systems:

a) **Autonomy**: Provide autonomous, decision-making, and change-adapting robots that don't require human assistance in dynamic, unstructured environments.
b) **Perception and Sensing**: Robots can better perceive their environment and interact with it if AI-driven sensor fusion, computer vision, and spatial awareness are implemented.
c) **Manipulation and Control**: AI-driven control algorithms can increase robot dexterity and precision when handling items, enabling tasks requiring intricate manipulation and fine motor abilities.
d) **Human-Robot Interaction (HRI)**: Allow for intuitive and natural interactions between people and robots by integrating AI into collaborative task planning, gesture detection, and speech recognition.
e) **Learning and Adaptation**: Create robots that can learn from their mistakes and adjust their behavior in response to feedback and changes in their surroundings. Over time, this will increase their efficiency in doing jobs.
f) **Safety and Reliability**: To ensure the safe operation of robots in a variety of settings and environments, implement AI algorithms for fault tolerance, collision avoidance, and real-time risk assessment.
g) **Multi-agent Systems**: Use AI-based coordination and communication protocols to arrange several robots to work cooperatively and effectively, maximizing task distribution and resource management.
h) **Robustness to Uncertainty**: Give robots artificial intelligence (AI) skills like probabilistic thinking and decision-making under uncertainty so they can effectively handle erratic circumstances and insufficient information.
i) **Task Autonomy and Flexibility**: Give robots the ability to independently plan and carry out difficult tasks with flexibility, modifying their plans in response to changing objectives and restrictions as they arise.
j) **Ethical and Social Considerations**: Ensure responsible deployment and social integration by addressing ethical concerns about AI in robotics, such as privacy, transparency, and the effect on employment.

AI-powered robotics is being developed with these goals in mind: to construct intelligent machines that can interact safely with humans, carry out a wide range of jobs, and autonomously adapt to changing situations

Some of the real-world examples where AI is being used are:

1. Industrial Automation and Manufacturing:

- **Robotic Arms**: In the manufacturing industry, robotic arms with AI capabilities are utilized to precisely assemble goods, weld, paint, and carry out repetitive operations. Robots are used in automotive assembly lines by companies like KUKA and Fanuc to increase efficiency and decrease errors.

- **Predictive Maintenance**: Artificial intelligence (AI) systems use machine sensor data to forecast when an equipment component might break, enabling maintenance to be done before a breakdown happens. This lowers maintenance expenses and downtime.

2. **HealthCare:**

- **Surgical-Robots**: Artificial intelligence (AI)-powered surgical robots, such as the da Vinci Surgical System, let doctors carry out difficult operations more precisely, with fewer incisions and quicker patient recovery.
- **Rehabilitation Robots**: AI-driven robots support physical therapy, offering tailored workouts and tracking patients' progress to aid in their recovery from accidents or strokes.

3. **Logistics and Supply Chain:**

- **Autonomous Mobile Robots (AMRs)**: Companies like DHL and Amazon deploy AI-powered robots in warehouses to transfer goods effectively. These robots are capable of navigating in dynamic settings, picking the best path while dodging obstacles.
- **Drones**: AI-enabled drones are used for inventory management in large warehouses, scanning barcodes and counting stock. Companies like Walmart have tested drone technology to manage inventory.

4. **Agriculture:**

- **Autonomous Tractors and Harvesters**: Precision agriculture uses AI-driven tractors and harvesters to sow seeds, apply fertilizer, and harvest crops. Businesses such as John Deere have created AI-powered autonomous farming equipment that maximizes agricultural techniques.
- **Crop Monitoring**: Drones and robots with AI capabilities keep an eye on crops for symptoms of disease, pests, and nutrient shortages. This enables farmers to quickly implement remedial measures, increasing crop yields.

5. **Service Robots:**

- **Customer Service Robots**: AI robots like SoftBank's Pepper are used in retail and hospitality to interact with customers, answer questions, and provide information. These robots are capable of understanding and responding to human emotions.
- **Cleaning Robots**: Autonomous vacuum cleaners, like iRobot's Roomba, use AI to navigate homes, avoid obstacles, and efficiently clean floors. These robots have become popular in households around the world.

Although artificial intelligence (AI) has the potential to be extremely beneficial in a number of fields, it also poses serious ethical questions.

Privacy and the effect on employment are two of the most urgent challenges. Here's a deeper dive into these subjects.

1. Privacy Concern:

For AI systems to work well, a lot of data is frequently required. Numerous privacy concerns are raised by the possibility that this data contains sensitive, private, and perhaps biometric information.

Data Collection and Surveillance:

- **Mass Surveillance**: Large populations can be monitored by AI systems, especially in the surveillance domain. Facial recognition technology, for example, can identify people without their permission when placed in public areas, allowing for widespread surveillance. This information could be misused by governments or businesses to violate citizens' rights and personal freedoms.
- **Data Ownership and Control**: Large-scale data collection by AI systems raises concerns regarding data ownership and access rights. People frequently don't have much control over how their data is used, which raises questions regarding permission and misuse possibilities.
- **Predictive analytics**: AI is able to forecast behavior, like customer preferences or possible criminal action, by analyzing personal data. While this can be helpful for law enforcement or targeted marketing, if it is not properly regulated, it can also result in invasive profiling and prejudice.

2. Impact on Jobs: AI has the power to drastically alter the labor market, which would have tremendous impact on jobs, the economy, and social equality.

- **Automation and Employment Losses:**

Manufacturing and Blue-Collar Jobs: Jobs in warehousing, manufacturing, and other blue-collar industries are being rapidly replaced by AI-driven automation. Because robots and AI systems can complete repetitive activities faster and cheaper than human labor, there will be a significant loss of jobs as a result.

- **Employment in the Service Industry**: AI technologies have the potential to automate employment in the service sector, including those of cashiers, customer service agents, and even delivery drivers. The necessity for human workers in these roles is already decreasing due to self-checkout systems, chatbots, and autonomous cars.
- **Impact on Skilled Jobs:** Jobs like financial analysis, journalism, and legal research that were previously thought to be immune to automation are also being threatened by AI. Artificial intelligence (AI) has the potential to replace human experts in a number of industries by analyzing legal documents, producing news stories, and making investing judgments.

IV. Conclusion:

To sum up, the incorporation of Artificial Intelligence (AI) into robotics is a significant development that has wide-ranging effects on several sectors of the economy and society. We have looked at the goals of artificial intelligence (AI) in robotics throughout this paper, as well as the cutting-edge technologies that result from these efforts and the moral issues that must surround their application. Artificial intelligence (AI) has given robots previously unheard-of powers, like autonomy in decision-making, improved awareness through sophisticated sensing technologies, and the ability to learn and communicate adaptably. These developments have made it possible for robots to carry out difficult jobs in dynamic situations, such as industrial automation, autonomous navigation, customized healthcare, and educational support. But as robotics driven by AI develops further, important ethical issues need to be taken into account. It is necessary to carefully consider and take proactive mitigation measures in response to issues including algorithmic prejudice, automation-related job displacement, privacy concerns, and the societal impact of AI-driven technology. Strong ethical frameworks, openness in AI development and application, and inclusive dialogues with participants from all industries are essential. In the future, maximizing AI's potential in robotics while ethically handling ethical dilemmas would require continued research and interdisciplinary collaboration. We can use AI to improve human capacities, and quality of life, and successfully address global concerns by promoting innovation in tandem with ethical accountability. AI in robots will continue to change our future, and maximizing advantages while limiting hazards will require interdisciplinary cooperation, regulatory frameworks, and public participation. In this era of intelligent machines, we can make sure that technical developments respect human values, foster sustainable growth, and contribute constructively to society by working toward ethical AI deployment.

References:

1. Dhivyabharathi, U., 2024. A Survey on Artificial Intelligence its Applications, International Journal of Research Publication and Reviews.
2. Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., 2023e. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence, Information Fusion.
3. Guo, Y., Wang, Y., Cao, Y., Long, Z., 2023. A New Optimization Design Method of Multi-Objective Indoor Air Supply Using the Kriging Model and NSGA-II. Applied Sciences 13, 10465. https://doi.org/10.3390/app131810465.
4. Chiu, T.K.F., Xia, Q., Zhou, X., Chai, C.S., Cheng, M., 2023. Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. Computers and Education Artificial Intelligence 4, 100118. https://doi.org/10.1016/j.caeai.2022.100118.
5. Mahto, A.K., 2023. ARTIFICIAL INTELLIGENCE FOR THE REAL WORLD. International Research Journal of Modernization in Engineering Technology and Science. https://doi.org/10.56726/irjmets42512.
6. Burad, S., Markad, K., Kulkarni, N., 2023. An Overview of Artificial Intelligence and Its Applications. International Journal of Innovative Research in Technology 9, 1186–1187.
7. Patil, N.H., Patel, S.H., Lawand, S.D., 2023. Research Paper On Artificial Intelligence and Its Applications. Journal of Advanced Zoology 229–238.
8. Saini, N., 2023. RESEARCH PAPER ON ARTIFICIAL INTELLIGENCE & ITS APPLICATIONS, International Journal for Research Trends and Innovation.
9. Gozalo-Brizuela, R., Garrido-Merchán, E.C., 2023c. A Survey of Generative AI Applications.
10. Joksimovic, S., Ifenthaler, D., Marrone, R., De Laat, M., Siemens, G., 2023. Opportunities of artificial intelligence for supporting complex problem-solving:

Findings from a scoping review. Computers and Education Artificial Intelligence 4, 100138. https://doi.org/10.1016/j.caeai.2023.100138.

11. Su, J., Ng, D.T.K., Chu, S.K.W., 2023. Artificial Intelligence (AI) Literacy in Early Childhood Education: The Challenges and Opportunities. Computers and Education Artificial Intelligence 4, 100124. https://doi.org/10.1016/j.caeai.2023.100124.
12. Diwan, C., Srinivasa, S., Suri, G., Agarwal, S., Ram, P., 2023. AI-based learning content generation and learning pathway augmentation to increase learner engagement. Computers and Education Artificial Intelligence 4, 100110. https://doi.org/10.1016/j.caeai.2022.100110.
13. Bewersdorff, A., Zhai, X., Roberts, J., Nerdel, C., 2023. Myths, mis- and preconceptions of artificial intelligence: A review of the literature. Computers and Education Artificial Intelligence 4, 100143. https://doi.org/10.1016/j.caeai.2023.100143.
14. Liu, T., 2023. AI proctoring for offline examinations with 2-Longitudinal-Stream Convolutional Neural Networks. Computers and Education Artificial Intelligence 4, 100115 https://doi.org/10.1016/j.caeai.2022.100115.
15. De Araujo, A., Papadopoulos, P.M., McKenney, S., De Jong, T., 2023. Automated coding of student chats, a trans-topic and language approach. Computers and Education Artificial Intelligence 4, 100123. https://doi.org/10.1016/j.caeai.2023.100123.
16. Ottikunta, L., 2022b. Improved constrained social network rating-based neural network technique for recommending products in an E-commerce environment. International Journal of Intelligent Networks 3, 80–86. https://doi.org/10.1016/j.ijin.2022.07.001
17. Shanthamallu, U.S., Spanias, A., Tepedelenlioglu, C., Stanley, M., 2021. A Brief Survey of Machine Learning Methods and their Sensor and IoT Applications, **ISBN:** 978-1-5386-3731-9.
18. Chen, J., Li, Kenli, Zhang, Z., Li, Keqin, 2021. A Survey on Applications of Artificial Intelligence in Fighting Against COVID-19. arXiv 1–37.
19. Batarseh, F.A., Freeman, L., Huang, C., 2021. A survey on artificial intelligence assurance, J Big Data. https://doi.org/10.1186/s40537-021-00445-7
20. Zawacki-Richter, O., Marín, V.I., Bond, M., Gouverneur, F., 2019. A systematic review of research on artificial intelligence applications in higher education – where are the educators? International Journal of Educational Technology in Higher Education. https://doi.org/10.1186/s41239-019-0171-0.
21. Agarwal, V., 2019. Impact of Artificial Intelligence on Humans: A Survey, Journal of Emerging Technologies and Innovative Research.
22. Apoorva, R., Deekshitha, Arasa, Shriyanka, 2018. A Survey on Artificial Intelligence (Window to Mankind), International Journal of Engineering Research & Technology (IJERT).
23. Gupta, N., 2017. A Literature Survey on Artificial Intelligence, International Journal of Engineering Research & Technology (IJERT).
24. Pannu, A., 2015. Artificial Intelligence and its Application in Different Areas, International Journal of Engineering and Innovative Technology (IJEIT).
25. Doukidis, G.I., Paul, R.J., 1992. A Survey of the Application of Artificial Intelligence Techniques within the OR Society, Operational Research Society Ltd.

ISSN: 2582-8118 *Volume 4, Issue 5; Sept 2024*

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYSIS IN SCIENCE AND ENGINEERING

Web: https://www.iarj.in/index.php/ijrase/index

87. An Insight into Key Algorithms for Sentiment Analysis

Raju A. Nadaf

Research Scholar,

Department of Computer Science and Engineering

Basaveshwar Engineering College, Bagalkot, India.

Veerappa. B. Pagi

Professor and Head

Department of Computer Science and Engineering

Basaveshwar Engineering College, Bagalkot, India.

ABSTRACT:

The modern technology era has emerged with lots of advancement in Natural Language Processing (NLP), Artificial Intelligence (AI), cybersecurity, Internet of Things (IoT) and other latest technological trends. The sentiment analysis has been considered as one of the important components of NLP.

The area is evolving with the usage of transformers and other latest algorithms with considerable increase in accuracy. Hence, there is a need to explore the avenues in the area of sentiment analysis for further investigation. With Recent advancements in text sentiment analysis there is an improvement in the performance and versatility of models. Such an adaptability is due to the usage of variety of techniques, in particular deep learning and transformer-based architectures. In this paper, the basics of sentiment analysis are covered along with the illustration of the latest algorithms, for better understanding of the area.

KEYWORDS:

NLP, Sentiment Analysis, Transformers.

1. Introduction:

Sentiment analysis is the process of analyzing digital text to determine if the emotional tone of the message is negative, positive, or neutral. Sentiment analysis is the most popular task in Natural language processing. It is basically analyzing the text for classification, based on mentality or mood, expressed in the text. The algorithms used for analysing text sentiment can be rule-based, hybrid or automatic.

Different types of Sentiment Analysis are

1. Fine-Grained Sentiment Analysis

It depends on polarity. This category has very positive, positive, neutral, negative, or very negative values. Rating can be from scale of 1 to 5.

2. Emotion detection

Also known as Lexicon method in which the sentiments can be happy, angry, sad, upset, angry, jolly and pleasant.

3. Aspect-Based Sentiment Analysis: It is a focus-based analysis. Focus may be on a particular aspect. Example of aspects can be battery, camera quality during purchase of mobile.
4. Multilingual Sentiment Analysis

Multilingual consists of different languages. It is difficult and challenging task.

Different levels of sentiment Analysis are:

1. Document Level: Analysis is performed on whole document and a single polarity is given to the whole document.
2. Sentence Level: Analysis is performed on each sentence to find out corresponding polarity.
3. Phrase level: Analysis is performed where word's opinions are mined at phrase level.
4. Aspect Level: Analysis is done for each sentence at aspect level. Finally, aggregated sentiment has to be calculated for a sentence to decide polarity.

Sentiment Analysis process involves data selection, data scraping, data analysis, the approach and finally the result.

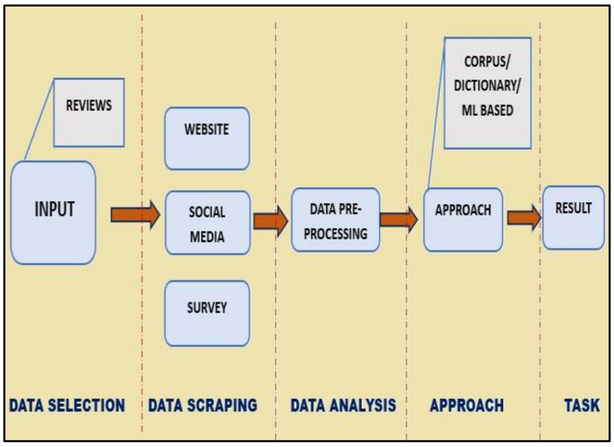


Figure 1: Generic Procedure for Sentiment Analysis

II. Literature Review:

[6] presents a multifaceted reality of sentiment analysis and [12] make acquaintance with a survey of sentiment analysis. [17] presents a review of recent text classification models. [21] elaborates the complete reviews, comparisons, and different approaches used during sentiment analysis along with advantages and drawbacks.

[5][18] elaborates an overview of the preprocessing techniques, feature extraction methods and classification techniques used in sentiment analysis. [33] Sentiment evaluation methods are explored and categorized.

[7] performs a critical assessment of different modules of a sentiment analysis framework. Various shortcomings associated with the existing methods have been discussed in detail. AI-powered sentiment analysis may be used to learn what the competitors' customers think of them by considering major aspects of the businesses [8].

Sentiments can be extracted towards entities using the decoder-based generative transformers [1]. A hybrid approach for Aspect Based Sentiment Analysis using transfer Learning is proposed [2] and [3] uses the Zero Sample Defender (ZDDR) for adversarial sample detection and recovery without relying on prior knowledge. [4] proposes a model that achieved results with the micro F1-score of 75.53% and 86.60% for the two datasets with different levels for the restaurant domain.

III. Key Algorithms and Models:

1. RoBERTa (Robustly optimized BERT approach):

This is an improved version of Bidirectional Encoder Representations from Transformers (BERT)The original BERT is added with the features:

a. Dynamic masking

Static masking is used in BERT models. In RoBERTa the masking is dynamic. The masking is generated uniquely every time a sequence is sent to model. As a result, the data is not duplicated during training and hence model gets more opportunity to work on more data and more masking patterns.

b. Omitting the next sentence prediction objective

By removing the Next Sentence Prediction (NSP) loss slightly improves downstream task performance.

c. Training on longer sentences

The large batches improve perplexity on masked language modelling objective and increases end-task accuracy. Large batch sizes can be trained parallelly.

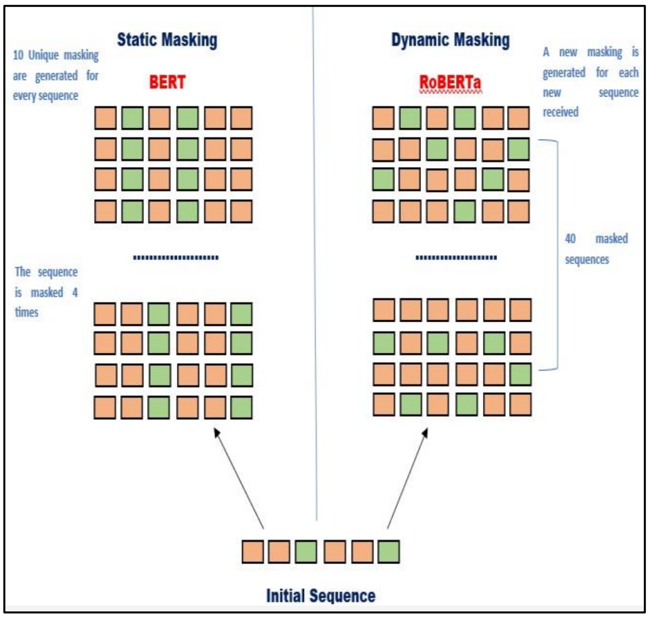


Figure 2: Working of RoBERTa model

2. T5 (Text-to-Text Transfer Transformer):

It is actually an **encoder-decoder model**. It converts the Natural Language Processing (NLP) problems into a text-to-text format. It is trained using teacher forcing. Teacher forcing involves that for training there is a need for an input sequence and a corresponding output sequence

It is pre-trained on a multi-tasking with a mixture of unsupervised and supervised tasks with the idea of clubbing multiple tasks while pre-training. T5 is trained with the Masked Language Model (MLM) objective.

3. ALBERT (A Lite BERT):

It improves efficiency by reducing the number of parameters and maintains performance. It is actually done by reduction of two parameter technique.

Factorized embeddings parameterization is the first technique. The large vocabulary embedding matrix is decomposed into two small possible matrices.

Hence, the size of the hidden layers is explicitly separated from the size of vocabulary embedding. This in turn grows the hidden size. Even there will not be considerable increase in the parameter size of the vocabulary embeddings.

Cross-Layer parameter sharing is a second technique. Hence, there will be prevention of growth of the parameter with the depth of the network.

4. XLNet:

Unlike BERT, which uses Masked Language Modelling (MLM), where certain words are masked during prediction based on requirement and context, XLNet uses Permutation Language Modelling (PLM). All the possible permutations of the input sequence are used during training. This approach helps to capture bidirectional context without masking.

In Generative Pre-Trained Transformer (GPT)based models, a token in the input sequence is predicted based on the previous token. Bidirectional dependencies will not be captured properly.

XLNet handles this problem by training the model to predict a token along with its context by considering all possible permutations of its context.

5. Efficiently Learning an Encoder that Classifies Token Replacements Accurately (ELECTRA):

The ability to understand and generate a useful and meaningful text representation is improved by using a training method in which a small subset of input is replaced by generated tokens. ELECTRA is a new pre-training approach that aims to use significantly less compute resources for the pre-training stage. The pre-training task in ELECTRA is based on detecting replaced tokens in the input sequence. This setup requires two Transformer models, a generator and a discriminator.

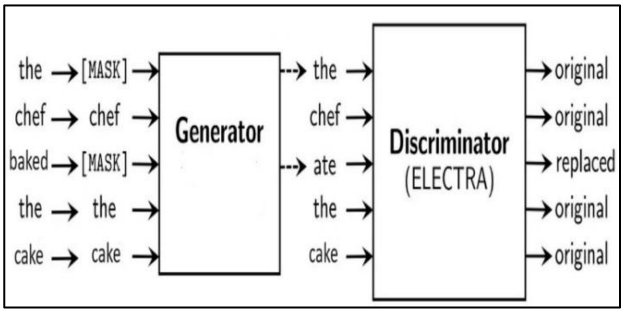


Figure 3: Working of ELECTRA model

For every token in the input sequence, the discriminator predicts whether it is an original token or whether the token is replaced by the generator.

Comparison of Key algorithms:

There is no best model that can be applied to all possible cases. It absolutely depends on the situation and scenario. Some of the important criteria to considered for choosing an algorithm are:

- Availability of hardware
- Long text in dataset

- Training from scratch or fine tuning the existing pretrained weights
- Downstream tasks Natural Language Understanding (NLU) based or Natural Language Generation (NLG)

The key features of the RoBERTa, T5, ALBERT, XLNet and ELECTRA are shown in Table 1.

Table 1: key features of the RoBERTa, T5, ALBERT, XLNet and ELECTRA

Name of the Model	Training method	Key Features
RoBERTa	A pretrained model with MLM task without NSP	The changes made to hyper-parameters are: Training time is long. Huge training data (16GB to 160GB). Big batch size, ranging from 256 to 8000. The NSP task is removed. Larger vocabulary size ranging from 30000 to 50000. Input sequences are long enough (limited to 512 tokens) Dynamic masking used
T5	Text-To-Text Transfer Transformer	A unified framework that converts all text-based language problems into a text-to-text format Achieved good performance on the GLUE, SuperGLUE, SQuAD, benchmarks. Produces the exact text of the answer for WebQuestions (37.4%), Natural Questions (34.5%) and TriviaQA (50.1%). T5 can be applied to regression tasks.
ALBERT	Uses modified MLM. The masked inputs are generated with n-gram spam of 3	NSP is replaced by Sentence Order Prediction (SOP). SOP does coherence prediction of sentences.
XLNet	Builds upon both BERT model and Transformer-XL model	Some weaknesses of BERT being handled here are: The assumption that the masked tokens are independent. As similar to autoregressive style it is fact that the prediction of one masked position can depend on other masked positions. XLNet does not replace tokens with [MASK]. The short context limitation. XLNet employs the recurrence mechanism and Relative Positional Encoding from Transformer-XL to for improving performance on long context, recurrence mechanism and Relative Positional Encoding is used.
ELECTRA	Pretrained on the Replaced Token Detection (RTD) task	The helper model generator is used to generate input for main model. This model is termed as Discriminator because the main purpose is to discriminate tokens to check whether each token is original or made up by generator

In order to compare the key algorithms, the following data sets are used. Quality of data is one of the prime factors affecting the accuracy of the algorithm. The mentioned sentiment analysis algorithms need large amount of diverse data for training so that they can produce more accurate results.

Table-2: The data set used for comparison of key algorithms

Sr No	Algorithm Name	Data Set Used
1	RoBERTa	160GB text data
2	T5	750GB text data
3	ALBERT	Combination of data being used for XLNet and RoBERTa
4	XLNet	158GB text data
5	ELECTRA	158GB text data

Here is a comparative analysis showing the performance of RoBERTa, T5, ALBERT, XLNet and ELECTRA models.

Table-3: Performance comparison of ROBERTA, T5, ALBERT, XLNET and ELECTRA

Model	Parameter						
	What's new	Size	Data	Training method	Hardware Used for training	Speed of Training	Performance
RoBERTa	Optimizes some parameters for BERT	Same as BERT	160GB	MLM	1024 32GB V100 GPUs	1 day	Better than BERT. Comparable to XLNet
T5	Encoder-Decoder Model	60-770M	750GB	Teacher Forcing	Four NVIDIA A100 GPUs with 40 GB of memory each	Less than a day	Better than BERT, RoBERTa and XLNet
ALBERT	*Sentence order Prediction *Cross layer parameter sharing *Factor based parameter embedding	Base:12M Large:18M Xlarge:60M Xxlarge:235M	Combination of data being used for XLNet and RoBERTa	MLM and Prediction of Sentence	64 to 512 TPU V3	90 hours	Better than BERT, RoBERTa and XLNet
XLNet	Combination of Bi-directional and autoregression	Almost same as BERT	158GB	PML	512 TPU V3 chips	5.5 days	Better than BERT

Model	**Parameter**						
	What's new	**Size**	**Data**	**Training method**	**Hardware Used for training**	**Speed of Training**	**Performance**
ELECTRA	Replacement of token Detection	14M-335M	158GB	RTD	V100 GPU	1/4th as that of RoBERTa and XLNet	Better than ALBERT, RoBERTa and XLNet

RoBERTa, ALBERT, XLNet and ELECTRA are trained with almost same sized data set. But for T5 algorithm around 750GB of text data set is used for training. T5 outperforms all other algorithms as far as speed of training is concerned.

Speed is not the only deciding factor for choosing an algorithm. Performance also matters. For measuring the performance, GLUE, BLEU, RACE, SQuAD and accuracy metrics can be used. In this paper, GLUE and SQuAD metrics are used for measuring the performance of mentioned algorithms. Empiricially, it is verified that with much smaller architecture models XLNet, RoBERTa, ALBERT and Electra achieve on-par performance with T5–3B. Fig-4 depicts the number of parameters used for each model.

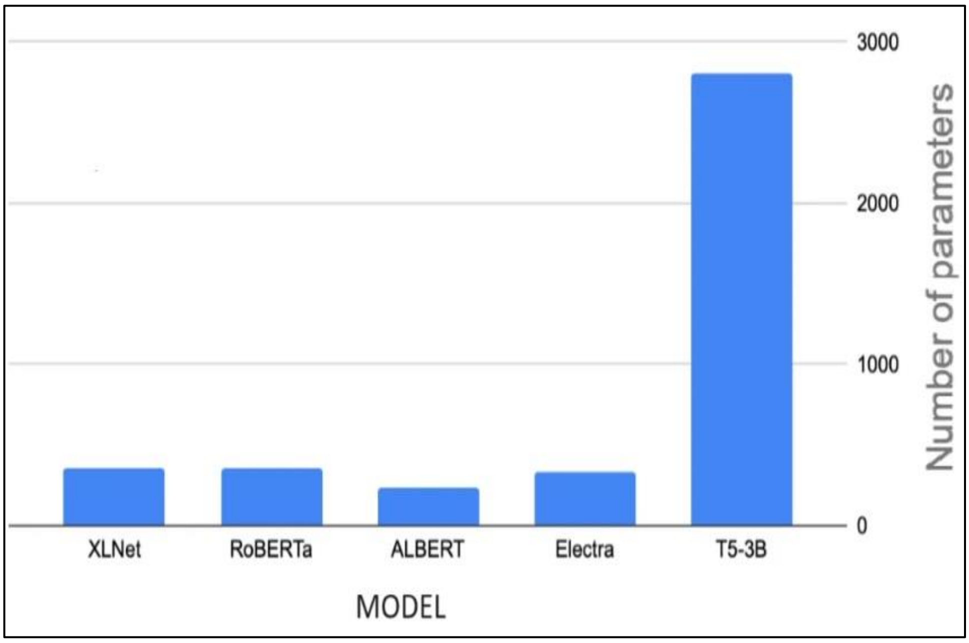


Figure 4: Number of Parameters Used for Each Model

Figure-5 shows the performance of XLNet, RoBERTa, ALBERT, Electra and T5 for small architecture (small number of parameters). Hence, mentioned algorithms perform almost consistent and equal for small number of parameters.

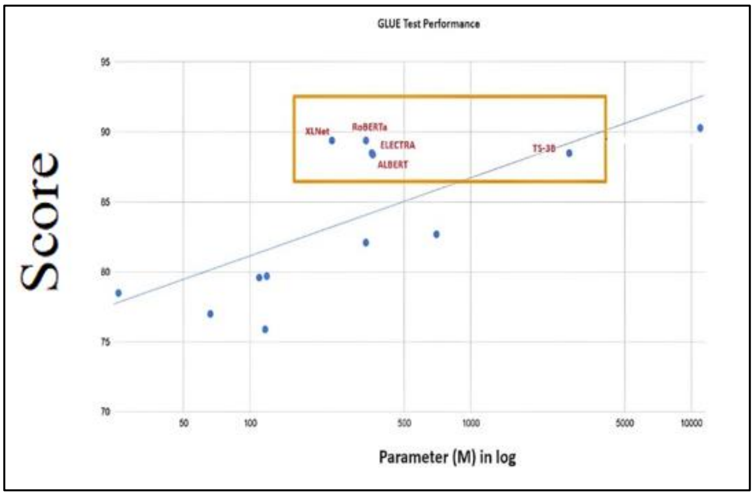


Figure-5: Performance Comparison of XLNet, RoBERTa, ALBERT, Electra and T5 for Small Architecture.

GLUE, is General Language Understanding Evaluation. It is an evaluation benchmark created to measure the performance of language understanding. From Fig-6, it is clear that the T5 algorithm performs well even when the architecture is big (a large number of parameters). The mentioned algorithms are tested using GLUE metrics.

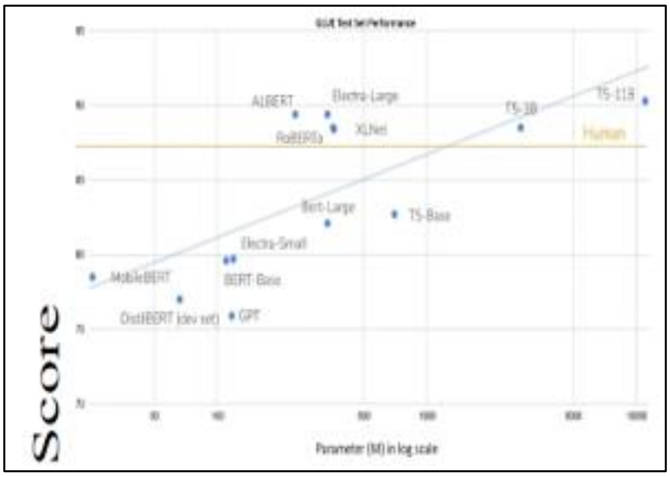


Figure-6: Performance of Algorithms on GLUE Test Set

SQuAD, is Stanford Question-Answering Dataset. It is a dataset designed for training and evaluating question-answering systems. From Figure-7, it is clear that the T5 algorithm performs well even when the architecture is big (over 10,000 parameters). The mentioned algorithms are tested using SQuAD 1.1 metrics.

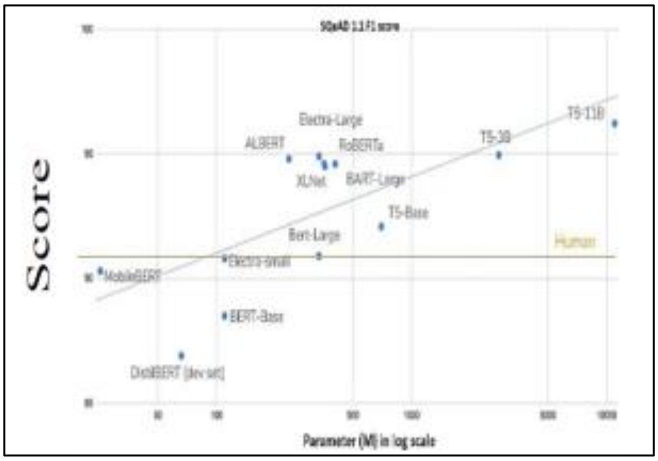


Figure 7: Performance of Algorithms on SQuAD 1.1 F1 Test Set

Applications of Sentiment Analysis:

Sentiment analysis can be used for business analysis, product analysis, politics, public reactions, healthcare and finance and the like. It is the most popular way to analyse text, survey responses, customer reviews, live chats and online reviews.

Such an analysis can help companies to perform well in business by understanding needs of customers. Some other Useful applications are Social media monitoring, Listen to voice of the customer (VoC), Product analysis and Market research and competitive research.

The following list of sectors in which sentiment analysis is used:

a. Business Analysis:

Businessmen may use sentiment analysis data to understand market demands, market trends, public opinion about product. The customer comments are useful for building flexible market strategies.

b. Health Care:

The healthcare data and reviews can be used to analyze the quality of service provided. The survey, twitter data, blogs, reviews provide sufficient amount of data for making analysis and for making a decision.

c. Product Review Analysis:

The data collected for any product will be utilized for product review. Such a review will help to improve the product quality and improved product quality may increase the business.

d. The travel:

The machine learning based approach along with web-based customer rating system has led to the data driven decision making system. Such a system with AI implementations may increase ease of life.

e. *Voice of Customer:*

Call center data, emails, chats, survey, web data can be categorized using sentiment analysis. Such an analysis finds pattern and solution for problems or any concerns.

f. Monitoring social media:

Social media provides real-time, round the clock data. Comments, trolls and forums and favourable opinions and positive comments help to increase reputation of the organizations.

VI. Conclusion:

The advancement in research has introduced new algorithms in the market with different flavours. In this paper, a comparative study of latest and most useful algorithms like RoBERTa, T5, ALBERT, XLNet and ELECTRA, is presented, for enhanced clarity and understanding.

The detailed analysis shows that, with much smaller architectural models, XLNet, RoBERTa, ALBERT and Electra achieve performance on-par with T5–3B. But, when the model size is bigger, T5 algorithm tends to perform better than other competitors for sentiment analysis. The present work has opened avenues for analysis of the data sets, different algorithms on empirical basis and scope for parameter tuning.

Some of the challenges in this field include sarcasm, irony, double negations, slangs, multilingual sentiment analysis and aspect-based sentiment analysis. The sentiment analysis fused with AI based domains promises a new horizon and insights.

References*:*

1 N. Rusnachenko, A. Golubev, and N. Loukachevitch, “Large Language Models in Targeted Sentiment Analysis,” *arXiv (Cornell University)*, Apr. 2024, doi: https://doi.org/10.48550/arxiv.2404.12342.

2 G. Negi, R. Sarkar, O. Zayed, and P. Buitelaar, “A Hybrid Approach To Aspect Based Sentiment Analysis Using Transfer Learning,” *arXiv (Cornell University)*, Mar. 2024, doi: https://doi.org/10.48550/arxiv.2403.17254.

3 M. Chen, G. He, and J. Wu, “ZDDR: A Zero-Shot Defender for Adversarial Samples Detection and Restoration,” *IEEE access*, pp. 1–1, Jan. 2024, doi: https://doi.org/10.1109/access.2024.3356568.

4 Dang Van Thin and N. Luu, “Aspect-Category based Sentiment Analysis with Unified Sequence-To-Sequence Transfer Transformers,” *VNU Journal of Science: Computer*

Science and Communication Engineering, Aug. 2023, doi: https://doi.org/10.25073/2588-1086/vnucsce.662.

5 K. L. Tan, C. P. Lee, and K. M. Lim, "A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research," *Applied Sciences*, vol. 13, no. 7, p. 4550, Apr. 2023, doi: https://doi.org/10.3390/app13074550.

6 M. Rodríguez-Ibánez, A. Casánez-Ventura, F. Castejón-Mateos, and P.-M. Cuenca-Jiménez, "A Review on Sentiment Analysis from Social Media Platforms," *Expert Systems with Applications*, vol. 223, p. 119862, Mar. 2023, doi: https://doi.org/10.1016/j.eswa.2023.119862

7 M. Bordoloi and S. K. Biswas, "Sentiment analysis: A survey on design framework, applications and future scopes," *Artificial Intelligence Review*, vol. 56, Mar. 2023, doi: https://doi.org/10.1007/s10462-023-10442-2.

8 H. Taherdoost and M. Madanchian, "Artificial Intelligence and Sentiment Analysis: A Review in Competitive Research," *Computers*, vol. 12, no. 2, p. 37, Feb. 2023, doi: https://doi.org/10.3390/computers12020037.

9 Thanakorn Chaisen, Phasit Charoenkwan, C. G. Kim, and Pree Thiengburanathum, "A Zero-Shot Interpretable Framework for Sentiment Polarity Extraction," *IEEE Access*, pp. 1–1, Jan. 2023, doi: https://doi.org/10.1109/access.2023.3322103.

10 Peerawat Chomphooyod, Atiwong Suchato, Nuengwong Tuaycharoen, and Proadpran Punyabukkana, "English grammar multiple-choice question generation using Text-to-Text Transfer Transformer," *Computers and education. Artificial intelligence*, vol. 5, pp. 100158–100158, Jan. 2023, doi: https://doi.org/10.1016/j.caeai.2023.100158.

11 S. N. Aakur and S. Sarkar, "Leveraging Symbolic Knowledge Bases for Commonsense Natural Language Inference using Pattern Theory," *IEEE transactions on pattern analysis and machine intelligence*, pp. 1–18, Jan. 2023, doi: https://doi.org/10.1109/tpami.2023.3287837.

12 J. Cui, Z. Wang, S.-B. Ho, and E. Cambria, "Survey on sentiment analysis: evolution of research methods and topics," *Artificial Intelligence Review*, vol. 56, Jan. 2023, doi: https://doi.org/10.1007/s10462-022-10386-z.

13 Ugranada Channabasava, Golu Kumar Ram, Bhanu Bhakta Jaishi, C. Raj, and Keshav Kumar Shandliya, "Aspect-Based Sentiment Analysis: A Survey of Deep Learning Methods," *Lecture notes in electrical engineering*, pp. 347–354, Jan. 2023, doi: https://doi.org/10.1007/978-981-99-2058-7_32.

14 J. Hartmann, M. Heitmann, C. Siebert, and C. Schamp, "More than a Feeling: Accuracy and Application of Sentiment Analysis," *International Journal of Research in Marketing*, vol. 40, no. 1, Jun. 2022, doi: https://doi.org/10.1016/j.ijresmar.2022.05.005.

15 Q. A. Xu, V. Chang, and C. Jayne, "A systematic review of social media-based sentiment analysis: Emerging trends and challenges," *Decision Analytics Journal*, vol. 3, p. 100073, Jun. 2022, doi: https://doi.org/10.1016/j.dajour.2022.100073.

16 Q. Li *et al.*, "A Survey on Text Classification: From Traditional to Deep Learning," *ACM Transactions on Intelligent Systems and Technology*, vol. 13, no. 2, pp. 1–41, Apr. 2022, doi: https://doi.org/10.1145/3495162.

17 Gasparetto, M. Marcuzzo, A. Zangari, and A. Albarelli, "A Survey on Text Classification Algorithms: From Text to Predictions," *Information*, vol. 13, no. 2, p. 83, Feb. 2022, doi: https://doi.org/10.3390/info13020083.

18 M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artificial Intelligence Review*, vol. 55, no. 55, Feb. 2022, doi: https://doi.org/10.1007/s10462-022-10144-1.

19 K. Pipalia, R. Bhadja, and M. Shukla, "Comparative Analysis of Different Transformer Based Architectures Used in Sentiment Analysis," *IEEE Xplore*, Dec. 01, 2020. https://ieeexplore.ieee.org/abstract/document/9337081.

20 K. Subramanyam Katikapalli, R. Ajit, and S. Sivanesan, "AMMUS: A Survey of Transformer-based Pretrained Models in Natural Language Processing," *arXiv.org*, Aug. 2021, doi: https://doi.org/10.48550/arXiv.2108.05542.

21 M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowledge-Based Systems*, vol. 226, no. 1, p. 107134, Aug. 2021, doi: https://doi.org/10.1016/j.knosys.2021.107134.

22 P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," *Social Network Analysis and Mining*, vol. 11, no. 1, Aug. 2021, doi: https://doi.org/10.1007/s13278-021-00776-6.

23 Y. Wang and S. Zhao, "Cross-Modal Representation Reconstruction for Zero-Shot Classification," Jun. 2021, doi: https://doi.org/10.1109/icassp39728.2021.9413572.

24 S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, "Deep Learning--based Text Classification," *ACM Computing Surveys*, vol. 54, no. 3, pp. 1–40, May 2021, doi: https://doi.org/10.1145/3439726.

25 Ligthart, C. Catal, and B. Tekinerdogan, "Systematic reviews in sentiment analysis: a tertiary study," *Artificial Intelligence Review*, vol. 54, Mar. 2021, doi: https://doi.org/10.1007/s10462-021-09973-3.

26 Siva, F. Dernoncourt, N. Lipka, and T. Solorio, "Exploring Conditional Text Generation for Aspect-Based Sentiment Analysis," *arXiv (Cornell University)*, Jan. 2021, doi: https://doi.org/10.48550/arxiv.2110.02334.

27 Cortiz, "Exploring Transformers in Emotion Recognition: a comparison of BERT, DistillBERT, RoBERTa, XLNet and ELECTRA," *arXiv (Cornell University)*, Jan. 2021, doi: https://doi.org/10.48550/arxiv.2104.02041.

28 G. E. Abdul, I. A. Ali, and C. Megha, "Fine-Tuned T5 for Abstractive Summarization," *International Journal of Performability Engineering*, vol. 17, no. 10, p. 900, 2021, doi: https://doi.org/10.23940/ijpe.21.10.p8.900906.

29 H. Wang, X. Hu, and H. Zhang, "Sentiment analysis of commodity reviews based on ALBERT-LSTM," *Journal of Physics: Conference Series*, vol. 1651, p. 012022, Nov. 2020, doi: https://doi.org/10.1088/1742-6596/1651/1/012022.

30 S. Garg, D. S. Panwar, A. Gupta, and R. Katarya, "A Literature Review On Sentiment Analysis Techniques Involving Social Media Platforms," *IEEE Xplore*, Nov. 01, 2020. https://ieeexplore.ieee.org/document/9315735

31 S. Ranjan and S. Mishra, "Comparative Sentiment Analysis of App Reviews," Jul. 2020, doi: https://doi.org/10.1109/icccnt49239.2020.9225348.

32 F. A. Acheampong, C. Wenyu, and H. Nunoo-Mensah, "Text-based emotion detection: Advances, challenges, and opportunities," *Engineering Reports*, vol. 2, no. 7, May 2020, doi: https://doi.org/10.1002/eng2.12189.

33 K. Chakraborty, S. Bhattacharyya, and R. Bag, "A Survey of Sentiment Analysis from Social Media Data," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 2, pp. 450–464, Apr. 2020, doi: https://doi.org/10.1109/tcss.2019.2956957.

34 Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and Radu Soricut, "ALBERT: A Lite BERT for Self-Supervised Learning of Language Representations," *arXiv (Cornell University)*, Apr. 2020, doi: https://doi.org/10.48550/arxiv.1909.11942.

35 Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and Radu Soricut, "ALBERT: A Lite BERT for Self-Supervised Learning of Language Representations," *arXiv (Cornell University)*, Apr. 2020, doi: https://doi.org/10.48550/arxiv.1909.11942.

36 K. Clark, M.-T. Luong, Q. V. Le, and C. D. Manning, "ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators," *arXiv (Cornell University)*, Mar. 2020, doi: https://doi.org/10.48550/arxiv.2003.10555.

37 K. Clark, M.-T. Luong, Q. V. Le, and C. D. Manning, "ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators," *arXiv (Cornell University)*, Mar. 2020, doi: https://doi.org/10.48550/arxiv.2003.10555.

38 Y. Xian, C. H. Lampert, B. Schiele, and Z. Akata, "Zero-Shot Learning—A comprehensive evaluation of the good, the bad and the ugly," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 9, pp. 2251–2265, Sep. 2019, doi: 10.1109/tpami.2018.2857768.

39 R. A. Nadaf, Rubina M, P. Sujata, and V. M. Bonal, "Smart Mirror Using Raspberry Pi for Human Monitoring and Intrusion Detection," Jul.2019, doi: https://doi.org/10.1109/icait47043.2019.8987294.

40 R. A. Nadaf, F. A. Katnur, and S. P. Naik, "Android Application Based Solid Waste Management," Proceeding of the International Conference on Computer Networks, Big Data and IoT (ICCBI - 2019), pp. 555–562, 2020, doi: https://doi.org/10.1007/978-3-030-43192-1_63

ISSN: 2582-8118 ***Volume 4, Issue 5; Sept 2024***

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYSIS IN SCIENCE AND ENGINEERING

Web: https://www.iarj.in/index.php/ijrase/index

88. Impact of Lumpy Skin Disease Virus on Indian Cattle: A Review of Mitigation Strategies

Raghavendra Y. Jadhav

Asst Professor,
Dept of CSE KLS Gogte Institute of Technology
(Affiliated to VTU), Belgavi, India.

Dr. Sadhana P. Bangarashetti

Professor, Dept of ISE,
Basaveshwar Engineering College, Bagalkot, India.

Abstract:

The cause of lumpy skin disease (LSD) is the LSD virus (LSDV). This virus belongs to the Poxviridae family and the Capripoxvirus genus, primarily affecting large ruminants, particularly domestic water buffalo and agricultural animals. The first recorded outbreak of Lumpy Skin Disease (LSD) occurred in Zambia in 1929. The disease subsequently spread across Africa and eventually reached Asia and Europe.

In Southeast Asia, LSD was first detected in Vietnam and Myanmar in 2020, and by 2021, it had reached Thailand and Laos. Consequently, LSD is a relatively new disease in Southeast Asia, necessitating further research on its pathophysiology, transmission modes, diagnosis, distribution, prevention, and treatment. Lumpy Skin Disease (LSD) is a viral infection that significantly affects cattle, causing considerable economic losses in the livestock industry.

This review investigates the impact of LSD on Indian cattle and explores how machine learning (ML) and artificial intelligence (AI) tools can be utilized to develop effective mitigation strategies. Integrating AI technologies in veterinary epidemiology can improve disease prediction, monitoring, and control efforts. This paper investigates the mechanisms of LSD transmission, current and future control strategies, and the potential role of AI and machine learning in disease management. The findings emphasize the importance of combining advanced technologies with traditional methods to improve the effectiveness of LSD control.

Keywords:

Lumpy Skin Disease Virus (LSDV), Capripoxvirus, sheep pox virus (SPPV), goat pox virus (GTPV).

I. Introduction:

Lumpy Skin Disease is caused by the Lumpy Skin Disease Virus (LSDV), which belongs to the genus Capripoxvirus as well as sheep pox virus (SPPV) and goat pox virus (GTPV) [1]. This virus can cause infection mainly in cattle (*Bos* spp.) and buffaloes (*Bubalus* spp.); there are also reports in other wild ruminant species, such as giraffes, bulls, and springboks [2].

The disease is characterized by fever, enlarged lymph nodes, and distinctive skin nodules that can lead to severe discomfort, secondary infections, and sometimes death. LSD has been increasingly reported in several Indian states, causing major concerns for livestock health and economic stability. LSD outbreaks occur frequently in various regional states of the country, despite intensive vaccination campaigns.

The aim of this research is to know the overview of LSD in Indian regions, reasons for outbreak of the disease and controlling methods. Vaccination, movement control and slaughter of infected and in-contact animals are considered as options for the control of LSD. However, it is widely agreed that vaccination is the most manageable and realistic approach to control the disease in endemic and resource poor countries [4].

II. Mode of Transmission:

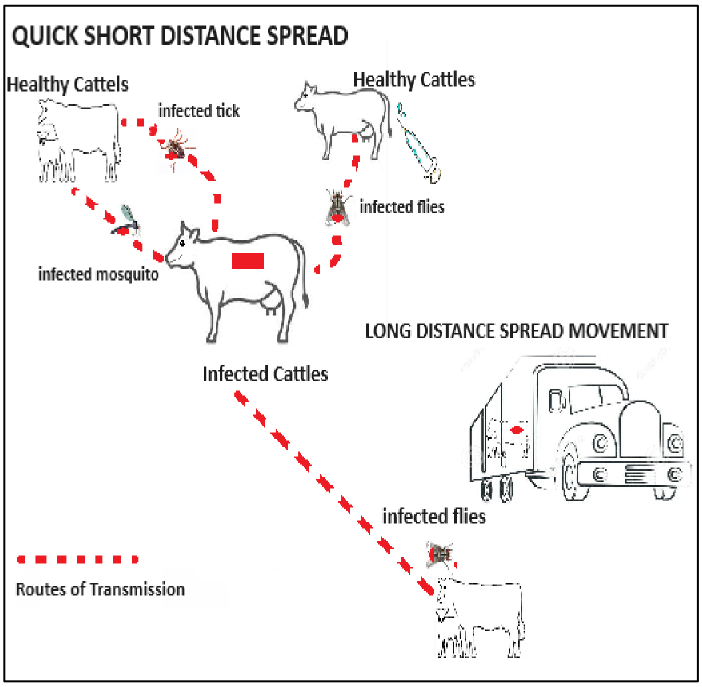


Figure 1: Spreading of Lumpy Skin Disease Viruses.

Figure 1 illustrates the transmission modes of Lumpy Skin Disease viruses

III. Impact on Indian Cattle and Buffaloes:

i). Clinical Symptoms and Morbidity:

- **Incubation Period**: 4-12 days.
- **Initial Symptoms**: Fever (40-41.5°C) lasting 1-3 days, increased nasal and pharyngeal secretions, lachrymation, lymph node enlargement, anorexia, dysgalactia, general depression, and reluctance to move [5] .
- **Skin Nodules**: Appear within 1-2 days, becoming harder and necrotic, causing severe discomfort, pain, and lameness. In 2-3 weeks, nodules may regress or lead to necrosis and infection [5] .
- **Other Effects**: Decreased milk production, loss of hide quality, temporary or permanent infertility in bulls, and increased risk of bacterial infections or myiasis. Morbidity ranges from 50-100%, and mortality from 1-5% [5] .

ii) Economic Impact:

- **Milk Production**: Significant reduction in milk yield.
- **Hide Quality**: Damage to skin reduces the market value of hides.
- **Draft Power**: Loss of draft power due to the disease coincides with critical farming seasons, affecting agricultural productivity [2] .
- **Economic Losses**: The 2016 Balkan outbreak of LSD led to losses of approximately 20.9 million Euros, highlighting the potential economic impact in India as well [2] .

iii) Epidemiology in India:

- **Outbreaks**: Documented in Odisha, Tamil Nadu, Andhra Pradesh, and Jharkhand, outbreaks are attributed to the movement of infected animals and vector transmission via biting flies and mosquitoes. [8].
- **Host Range**: Lumpy Skin Disease impacts cattle (Bos indicus and Bos taurus) and buffaloes (Bubalus bubalis), with Bos taurus being more vulnerable. Wild animals generally exhibit resistance to LSDV under natural conditions. [5].

IV. Mitigation Strategies:

i) Vaccination:

- **Vaccines**: Use of vaccines based on sheep and goat pox viruses, which provide cross-protection against LSDV. Vaccination programs are actively promoted by veterinary authorities to control LSD [5] .
- **Coverage**: Ensuring widespread vaccination coverage, particularly in outbreak-prone regions.

ii) Movement Restrictions:

- **Regulations**: Implementing movement restrictions on cattle and buffaloes from affected areas to prevent the spread of the virus [5].

- **Quarantine Measures**: Establishing quarantine zones around outbreak areas.

iii) Vector Control:

- **Insecticides and Repellents**: Using insecticides and repellents to control populations of biting flies and mosquitoes.
- **Environmental Management**: Managing cattle environments to reduce vector breeding sites [2] .

iv) Surveillance and Reporting:

- **Monitoring Systems**: Establishing robust surveillance systems for early detection and reporting of LSD cases.
- **Data Collection**: Maintaining comprehensive records of outbreaks to analyse trends and improve response strategies [7] .

v) Public Awareness and Training:

- **Education Programs**: Conducting educational programs to inform farmers and veterinary personnel about LSD symptoms, transmission, and prevention strategies.
- **Training**: Providing education on optimal practices for disease management and control [8] .

vi) Minimizing contacts: Minimizing contacts between wildlife and cattle during LSD outbreaks might help limit cross-species transmission. Continued monitoring is needed to assess the impact of LSDV on gazelles and other wild and domestic ruminants in India.

vii) Make Use of traditional medicine: Use of traditional medicine in supporting the recovery of infected cattle. While these treatments do not guarantee recovery, they may aid in the healing process. The Department of Animal Husbandry has signed an MoU with the Ministry of AYUSH to explore integrating traditional practices into veterinary care [9]

V. LSD Risk factors:

Small farms are more susceptible due to increased chances of insect bite, and nearness to the lake is associated with increased insect vector activity. Areas with great odds of being positive for LSD were mostly covered with croplands, grassland or shrubland, with higher cattle density, as well as areas with higher annual mean temperature and higher temperature diurnal range.

VI. Role of AI and ML in Disease Surveillance:

i) Predictive Analytics:

- **Epidemiological Modelling**: AI and ML algorithms can analyze historical outbreak data to predict future LSD outbreaks. Predictive models help in identifying high-risk areas and timing of potential outbreaks, allowing for proactive measures [8] .

- **Climate and Environmental Data Integration**: ML models can integrate climate data, such as temperature and humidity, which influence vector populations, to predict LSD spread patterns [8] .

ii) Real-time Surveillance:

- **Remote Sensing and IoT**: Integration of remote sensing technologies and Internet of Things (IoT) devices with AI algorithms can provide real-time monitoring of cattle health and environmental conditions. Early detection of anomalies can trigger immediate responses [2] .

- **Geospatial Analysis**: AI-driven geospatial analysis of disease spread helps in visualizing and managing the spatial distribution of LSD outbreaks, facilitating targeted interventions [5]

VII. AI and ML in Diagnostic Tools:

i) Image Recognition and Analysis:

- **Automated Detection**: ML models trained on images of healthy and infected cattle can accurately identify signs of LSD from photographs, enabling early diagnosis and reducing the need for manual inspections [2].

- **Mobile Applications**: Farmers may now send photos of their cattle to be evaluated for signs of LSD using smartphone apps that have been developed with AI capabilities. This allows for quick diagnostic feedback. [8][13].

ii) Genomic Sequencing and Analysis:

Virus Detection and Typing: AI algorithms can process genomic data to identify and classify LSDV strains rapidly. This assists in understanding virus evolution and planning vaccination strategies [8] .

VIII. Systems for Supporting Decisions:

i) Resource Allocation:

- **Optimized Vaccination Strategies**: ML models can optimize vaccination campaigns by analyzing data on cattle populations, disease prevalence, and logistics, ensuring efficient use of resources and maximum coverage [2].
- **Outbreak Response Planning**: AI-driven decision support systems can offer data-driven advice to veterinary authorities, assisting them in the planning and execution of outbreak responses. [5] [15].

ii) Risk Assessment and Management:

- **Risk Maps**: AI can generate risk maps based on various data inputs, helping authorities focus on high-risk areas for targeted interventions.

- **Economic Impact Analysis**: ML models can estimate the economic impact of LSD outbreaks and control measures, aiding in policy-making and resource allocation [8].

IX. Challenges and Future Directions

i) Data Quality and Availability:

- **Data Integration:** Requirements for using AI and ML effectively include large-scale, well-curated datasets. It is essential to integrate data from multiple sources, such as remote sensing, temperature data, and veterinary records. [2][13][14].
- **Standardization**: AI and ML models are more accurate and useful when data collecting and reporting procedures are standardized across geographical boundaries. [5] .

ii) Technical and Infrastructure Barriers:

- Access to Technology: Ensuring that farmers and veterinary services have access to AI and ML tools is essential. This includes providing necessary hardware, software, and training [2] [11][12][13].
- Scalability: Developing scalable AI solutions that can be adapted to different regions and cattle management systems is vital for widespread adoption [5] .

iii) Ethical and Privacy Concerns:

- Data Privacy: Managing and protecting sensitive data is critical. Establishing robust data privacy frameworks ensures the ethical use of AI and ML technologies [8] .
- Bias and Fairness: Addressing biases in AI models to ensure fair and accurate disease management practices is crucial. Continuous monitoring and updating of algorithms are necessary to maintain fairness [2].

X. Data Sources:

1. **Lumpy skin disease outbreak data:** Outbreak locations for Lumpy Skin Disease were sourced from the Food and Agriculture Organization's (FAO) Global Animal Disease Information System (https://empres-i.review.fao.org//) [10].
2. **Meteorological data:** Monthly cloud cover (percentage), diurnal temperature range (degrees Celsius), frost day frequency (days per month), wet day frequency (days), potential evapotranspiration (millimeters per day), precipitation (millimeters per months), daily mean temperature (degrees Celsius), monthly average maximum and minimum temperature (degrees Celsius), and vapor pressure (hectopascal) data can be obtained from the Indian Weather Repository https://mausam.imd.gov.in/bengaluru/ [10].
3. **Animal density data:** Cattle and buffalo population density data can be obtained from

https://dahd.nic.in/documents/ statistics/livestock-census [10].

4. **Land Use/Land Cover:** This data can be obtained from https://www.nrsc.gov.in/EO_LULC Objective? language_ content_ entity= en [10].
5. **Elevation data:** Global geospatial elevation dataset (GRAY_50M_SR. VERSION 2.1.0) can be downloaded from Natural Earth database (free vector and raster map data @ natural earth- data.com) [10]

XI. Research and Development:

- Research Funding: Increased investment in research is needed to develop more effective vaccines and treatments for LSD.
- Surveillance Systems: Establishing robust surveillance systems can aid in early identification of outbreaks and facilitate timely management responses.

XII Conclusion:

The prediction, monitoring, and control of Lumpy Skin Disease can be greatly improved by integrating machine learning and artificial intelligence (AI) methods. Veterinary authorities in India may create more effective mitigation plans by utilizing these technologies, which will ultimately lessen the effect of LSD on the cattle business.

References:

1. Kanokwan Ratyotha, Suksanti Prakobwong, Supawadee Piratae, "Lumpy skin disease: A newly emerging disease in Southeast Asia", Vet World. 2022 Dec, 15(12): 2764–2771, Veterinary world, 2022 Dec 5, doi: 10.14202/vetworld.2022.2764-2771
2. Naveen Kumar, Yogesh Chander,Ram Kumar,Nitin Khandelwal, Thachamvally, Riyesh ,Khushboo Chaudhary, Karuppusamy Shanmugasundaram, Sanjit Kumar, Anand Kumar, Madhurendu K. Gupta,Yash Pal, Sanjay Barua ,Bhupendra N. Tripathi, "Isolation and characterization of lumpy skin disease virus from cattle in India", PLOS ONE, Volume: 16(1):e0241022, January 11, 2021, https://doi.org/10.1371/journal. pone.0241022.
3. Gari G, AbieG, GizawD, WubeteA, KidaneM, AsgedomH, BayissaB, Ayelet G, Oura CA, Roger F, Tuppurainen ES (2015), "Evaluation of the safety, immunogenicity and efficacy of three capripoxvirus vac cine strains against lumpy skin disease virus. Vaccine 33:3256 3261. https://doi.org/10.1016/j.vaccine.2015.01.035
4. E.S.M. Tuppurainen, E.H. Venter, J.A.W. Coetzer, L. Bell-Sakyi. "Lumpy skin disease: Attempted propagation in tick cell lines and presence of viral DNA in field ticks collected from naturally-infected cattle", Ticks and Tick-borne Diseases, Volume 6, Issue 2, March 2015, Pages 134-140, https://doi.org/10.1016/j.ttbdis.2014.11.002
5. Tania Gupta, Vanita Patial, Diksha Bali, Shivani Angaria, Mandeep Sharma, Rajesh Chahota, "A review: Lumpy skin disease and its emergence in India", Veterinary Research Communications (2020) Volume 44:111–118, Springer Nature B.V. 2020, 28 August 2020, 10.1007/s11259-020-09780-1
6. John Wiley and Sons Ltd, "Lumpy skin disease II. Data collection and analysis", European Food Safety Authority (EFSA), 19 February 2018,

https://doi.org/10.2903/j.efsa.2018.5176. “Isolation and characterization of lumpy skin disease virus from cattle in India”, PLOS ONE, Volume: 16(1): e0241022, January 11, 2021, https://doi.org/10.1371/journal.pone.0241022.

7. Manimuthu Prabhu, Shanmugasamy Malmarugan, Sithanandam Rajagunalan, Balakrishnan Govindan, Lakshmi Prasanth Thangavelu, Ganapathi Palanisamy, Revanaiah Yogisharadhya, Kumaragurubaran Karthi, “Isolation and molecular characterization of lumpy skin disease virus from Tamil Nadu, India during the outbreaks from 2020 to 2022”, Volume: Virus Genes 2024, April 06 2024, , https://doi: 10.1007/s11262-024-02057-8.
8. Kalyani Putty, Pachineella Lakshmana Rao, Vishweshwar Kumar Ganji, Devasmita Dutta, Subhajit Mondal, Nagendra R. Hegde, Anand Srivastava & Madhuri Subbiah, “First complete genome sequence of lumpy skin disease virus directly from a clinical sample in South India”, Volume 59, Pages 317–322 (2023). https://doi.org/10.1007/s11262-023-01967-3
9. S. Jayakumar a, N. Baskaran, “Herbal medicine as a live practice for treating livestock ailments by indigenous people: A case study from the Konar community of Tamil Nadu”, South African Journal of Botany, Volume 118, Pages 23-32, September 2018, 10.1016/j.sajb.2018.06. 002..
10. Ehsanallah Afshari Safavi, “Assessing machine learning techniques in forecasting lumpy skin disease occurrence based on meteorological and geospatial featur es”, Tropical Animal Health and Production (2022) Springer Nature, https://doi.org/10.1007/s11250-022-03073-2.
11. Mrs. Anitha Rao, Monika H R, Rakshitha B C, Seham Thaseen, “Cattle Disease Prediction Using Artificial Intelligence”, Internaional Journal for research in applied science and engineering technology, Volume 11, Issue: IV,
12. Cateni, S., Vannucci, M., Vannocci, M. and Colla, V. (2012). Variable selection and feature extraction through artificial intelligence techniques. Multivariate Analysis in Management, Engineering and the Science. 6: 103-118.
13. Harsh J. Shah, Chirag Sharma, Chirag Joshi “Cattle medical diagnosis and prediction using machine learning” International Research Journal of Engineering and Technology (IRJET) – 2022
14. Dmitry Yu. Pavkin , Alexei S. Dorokhov , Fedor E. Vladimirov , Igor M. Dovlatov and Konstantin S. Lyalin “Algorithms for Detecting Cattle Diseases at Early Stages and for Making Diagnoses and Related Recommendations” Applied Science-2021
15. Dibyendu Biswas. Shib Shankar Saha, Shankar Biswas and Md. Abu Sayeed, “Outbreak of Lumpy Skin Disease of Cattle in South-West Part of Bangladesh and its Clinical Management”, Veterinary Sciences: Research and Reviews, Volume 6 | Issue 2 | Page 101, December 2020, http://dx.doi.org/10.17582/journal.vsrr/2020.6.100.108

ISSN: 2582-8118 *Volume 4, Issue 5; Sept 2024*

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYSIS IN SCIENCE AND ENGINEERING

Web: https://www.iarj.in/index.php/ijrase/index

89. A Deep learning Model for Bird Sound Recognition

Vishwanath Kagawade

Artificial Intellience and Machine Learning
Basaveshwar Engineering College Bagalkote, India.

K. Chandrasekhar, Jayashri G. Ghantimath

Physics, Basaveshwar Engineering College
Bagalkote, India.

Abstract:

Bird sound recognition is a key tool for tracking migration patterns. Bird sound recognition is a rapidly evolving field, driven by advancements in signal processing, machine learning and deep learning techniques. Despite challenges like background noise and data scarcity, innovative techniques such as deep learning and transfer learning are enhancing the accuracy and robustness of recognition systems. Continued interdisciplinary collaboration and the integration of citizen science will further propel the field, contributing to biodiversity conservation and ecological research. This article provides insights of state-of-the-art techniques used for bird sound recognition. Deep learning technique is presented for bird sound recognition and 98.40% recognition accuracy is achieved.

Keywords:

Bird Migration, Bird sound recognition, Deep Learning model, Signal processing.

I. Introduction:

Migration of birds is important for the conservation and management of ecosystem. This conservation and management of ecosystem has significant effect on the human health and other species. Migration is a seasonal movement of species from one place to another place. Migration of birds is being monitored throughout the continent and they are better than other migration species. Migration of birds is required to disperse seeds, pollinate flowers, manage pests, nutritious vegetation [1-3].

Birds are ecosystem service providers and they are required for the environment to benefit the humans and other species. Birds play important vital role in maintaining health and consistency of ecosystem and also, they are important elements of the ecosystem [4].

Birds are sensitive to changes in climate, habitat structure and composition. Birds are good indicators of quality of habitat and bio-diversity [1]. Birds migrate from one place to another place in response to changes in climate/temperature, food supply, amount of day light and for the purpose of breeding.

Over the last few decades, research on bird's migration adopted new technologies. The study on migration of birds is essential and important to assess the ecosystem and biodiversity [1, 5-11]. The classical methods of monitoring bird's migration like observing the birds during specific seasons/periods with the help of citizens or bird watchers. The problem of such bird observation depends on the observer ability and experience to recognize birds. The observer experiences in bird's identification, detectability, species identification are very crucial [1, 7]. The observer's performance is very much influenced by his age, mental ability, hearing ability, eye sight and level of fatigue, and also very much depends on his physical health and motivation towards observation [1]. Also, the environmental variables and bird behaviors has very much affected by the observer to assess the bird's migration to the particular place. The classical observation of migration of birds to the place is the correct approach and methodology but it is lot of time consuming, expensive and also requires the experienced observers/citizens.

Using digital camera, photographs during migration events can be recorded. The recorded data can be processed manually in a computer and required information easily obtained. However, this process requires experienced observers and also consumes lot of time and effort. Further, in this method bad weather conditions will very much affect the quality of photography and may induce errors in identifying migration of birds. This manual process leads to disruptions in the completion of the process within the time [12].

The other monitoring/prediction methods, such as long-term observation at migrate hot spots, large scale weather surveillance radar networks [3]. These techniques are required to understand long-term, large-scale pattern of bird's migration. These methods are not accurate to predict migration of birds at the specific level or individual level.

Another method adopted to monitor the individual bird's migration to specific area is recording the bird call or song which provides the detailed recording of movement of the individual birds. The recordings of millions of data and analyze them requires lot of time and very expensive.

Over the many years' machine learning and deep learning have become increasingly popular to use for many problems. Considering the complexity of migration of birds to different continent and increase in the size of data, the machine learning and deep learning is the one to use for predicting the migration of birds from one place to another place [7-11]. The advancements in the machine learning and deep learning may provide the better solution to predict the migration of birds to different regions. In this direction, article presents a new deep learning model for bird sound recognition.

The remaining part of the paper has been organized as: section II covers detailed survey on migration of bird and bird sound recognition. Section III highlights various methods adapted for bird sound recognition and proposed modal is presented. Section IV includes experimentation and results, followed by conclusion and future scope in section V.

II. Literature Review:

Bird migration is a complex and fascinating natural phenomenon that involves the regular, often seasonal, movement of bird species from one region to another. This process is driven by various ecological factors, including food availability, breeding, and climatic conditions. Understanding bird migration requires insights from various fields such as ornithology, ecology, and geography. This section provided a detailed survey of bird migration, with references to key studies and sources.

A. Types of Bird Migration:

The types of bird migration can be classified into three categories namely Latitudinal Migration, Altitudinal Migration and Longitudinal Migration and shown in Figure. 1.

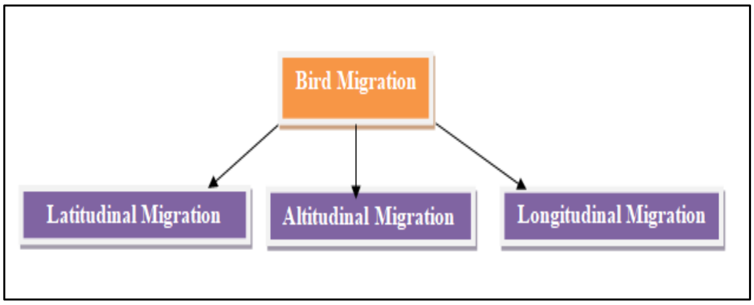


Figure 1: Types of Bird Migration

Latitudinal Migration: Many birds migrate between higher latitudes (where they breed) and lower latitudes (during winter). This type of migration is particularly common among species in the Northern Hemisphere. *Altitudinal Migration:* Some species migrate vertically up and down mountains rather than across latitudes. This is often seen in mountainous regions where birds move to lower altitudes during the winter. *Longitudinal Migration:* Less common, this type involves east-west movements, usually in response to changing environmental conditions.

Bird migration is triggered by a combination of internal and external factors such as Photoperiod, Hormonal Changes, and Weather Conditions. *Photoperiod:* Changes in day length serve as a primary cue for many migratory birds to start their journey. *Hormonal Changes:* Internal physiological changes, often influenced by the endocrine system, prepare birds for migration. *Weather Conditions:* Temperature changes and weather patterns can also act as migration cues.

Birds use Celestial Cues, Geomagnetic Fields, Landmarks and Olfaction etc. variety of mechanisms to navigate during migration. *Celestial Cues:* Many species use the position of the sun, stars, and the moon to navigate. *Geomagnetic Fields:* Birds can detect the Earth's magnetic field and use it for orientation. *Landmarks:* Visual landmarks such as rivers, mountains, and coastlines help in navigation. Olfaction: Some birds use their sense of smell to aid in migration.

B. Case Studies and Research Findings:

This section gives some of the case studies and research outcomes on long distance and short distance migrants.

a) Long-distance Migrants

- **Arctic Tern (Sterna paradisaea):** Known for the longest migration of any bird species, Arctic Terns travel between the Arctic and Antarctic regions, covering a distance of at least 25,000 miles (40,000 kilometers) annually. Research using geolocators has provided detailed insights into their migratory routes and stopover sites [13].
- **Bar-tailed Godwit (Limosa lapponica):** This species undertakes non-stop flights of over 11,000 kilometers from Alaska to New Zealand. Satellite tracking has revealed their impressive endurance and navigational capabilities [14].

b) Short-distance Migrants:

- **American Robin (Turdus migratorius):** These birds migrate from their breeding grounds in Canada and the Northern United States to wintering sites in the Southern United States and Mexico. Studies have shown that Robins rely heavily on photoperiod and temperature changes to time their migration [15].
- **European Blackbird (Turdus merula):** In Europe, blackbirds show partial migration, with northern populations moving southwards during winter. Research indicates that weather patterns and food availability are crucial in determining migration timing and routes [16].

Migratory birds face numerous threats along their migratory routes, including habitat loss, climate change, and human-made obstacles like wind turbines and skyscrapers. Conservation efforts are critical to ensure the survival of these species.

C). Conservation Strategies:

Protected Areas: Establishing and maintaining protected areas along migratory routes is essential for providing safe stopover and breeding sites.

International Cooperation: Since migratory birds cross international borders, cooperation between countries is vital for effective conservation.

Research and Monitoring: Continued research and the use of technology like satellite tracking and geolocations help in understanding migration patterns and identifying critical habitats.

Bird migration and bird sound recognition are interconnected in research as both fields provide valuable insights into avian behavior, ecology, and conservation. Bird sound recognition is a complex task that benefits from the integration of advanced audio processing techniques and machine learning models.

Through careful experimentation and analysis, it is possible to develop accurate and reliable systems for identifying bird species based on their vocalizations. The success of such systems depends on high-quality data, effective feature extraction, robust models, and thorough result analysis.

This detailed survey highlights the complexity of bird migration and underscores the importance of comprehensive conservation strategies to protect these incredible journeys.

III. Methodology:

Bird sound recognition is an interdisciplinary field involving ornithology, signal processing, machine learning, and bioacoustics. This field has garnered significant interest due to its applications in biodiversity monitoring, conservation, and understanding avian behavior.

This survey aims to provide an overview of the methodologies, challenges, and advancements in bird sound recognition.

A. Methodologies:

Bird sound recognition involves the use of various techniques and methods to identify bird species based on their vocalizations. Section provides an overview of the common techniques used in bird sound recognition [17-22]:

a) Pre-processing Techniques:

- *Noise Reduction:* Techniques such as spectral subtraction or band-pass filtering to remove background noise.
- *Segmentation:* Dividing the audio signal into smaller chunks that can be processed individually.
- *Normalization:* Adjusting the amplitude of the signal to a standard level to reduce variability.

b) Feature Extraction:

- *Mel-Frequency Cepstral Coefficients (MFCCs):* These coefficients capture the power spectrum of a sound signal and are commonly used in speech and audio processing.
- *Spectrograms*: Visual representations of the spectrum of frequencies in a sound signal as it varies with time. They are useful for visual pattern recognition and as inputs to machine learning models.
- *The Spectral Features:* Includes Frequency content, spectral centroid, bandwidth.
- *Chroma Features:* Represent the 12 different pitch classes and are useful for capturing harmonic and melodic content.
- *Zero-Crossing Rate (ZCR):* Measures the rate at which the signal changes sign, useful for distinguishing between voiced and unvoiced sounds.
- *Temporal Features:* Such as root mean square energy (RMSE), which capture the dynamics of the sound over time.

c) Machine Learning Models:

- *Traditional Models:* K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Hidden Markov Models (HMMs) and Random Forests can be used for classification based on extracted features.
- *Deep Learning Models:* Convolutional Neural Networks (CNNs) are particularly effective with spectrograms, while Recurrent Neural Networks (RNNs) and Long Short-Term Memory Networks (LSTMs) can capture temporal dependencies in audio signals.

B. Challenges:

a) *Background Noise:* Bird sound recordings often include background noise from wind, water, and other animals, complicating the extraction of clean bird sounds.
b) *Variability in Bird Calls:* Individual birds may exhibit variations in their calls due to age, health, and environmental factors, making recognition challenging.
c) *Data Scarcity:* High-quality labeled datasets are scarce, hindering the training and validation of robust models.

To address the above listed challenges, the work presents a CNN model and architecture diagram of the proposed model is illustrated in Figure. 2.

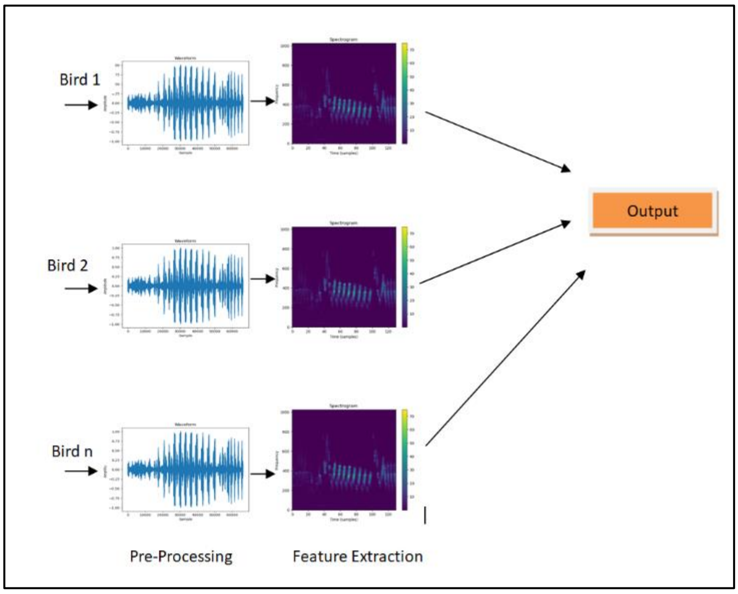


Figure 2: Architecture diagram of the proposed model

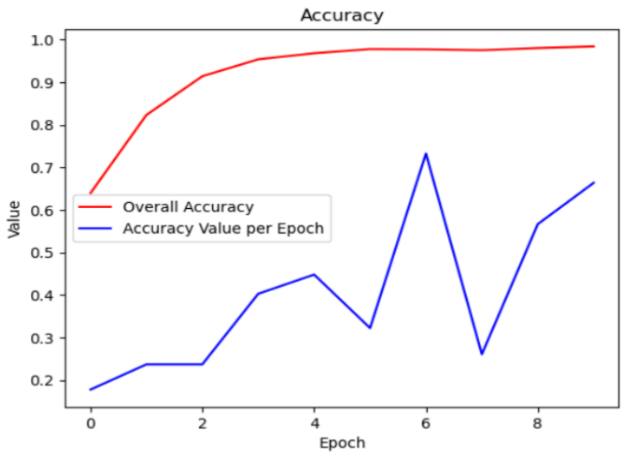
In the initial state, bird sounds wave files are read and data augmentation technique was applied to all the audio files (during pre-processing) to handle overfitting problem. During training and testing Mel spectrum features are extracted and feed into the CNN model. The CNN model is able to perform bird sound recognition task. CNN model summary is shown in Figure.3.

```
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 reshape_2 (Reshape)         (None, 128, 130, 1)       0

 conv2d_1 (Conv2D)           (None, 121, 123, 64)      4160

 leaky_re_lu_1 (LeakyReLU)   (None, 121, 123, 64)      0

 batch_normalization_1 (Bat  (None, 121, 123, 64)      256
 chNormalization)

 max_pooling2d_1 (MaxPoolin  (None, 60, 61, 64)        0
 g2D)

 conv2d_2 (Conv2D)           (None, 59, 60, 16)        4112

 leaky_re_lu_2 (LeakyReLU)   (None, 59, 60, 16)        0

 flatten (Flatten)           (None, 56640)             0

 dropout (Dropout)           (None, 56640)             0

 dense (Dense)               (None, 128)               7250048

 leaky_re_lu_3 (LeakyReLU)   (None, 128)               0

 dense_1 (Dense)             (None, 5)                 645

=================================================================
Total params: 7259221 (27.69 MB)
Trainable params: 7259093 (27.69 MB)
Non-trainable params: 128 (512.00 Byte)
_________________________________________________________________
```

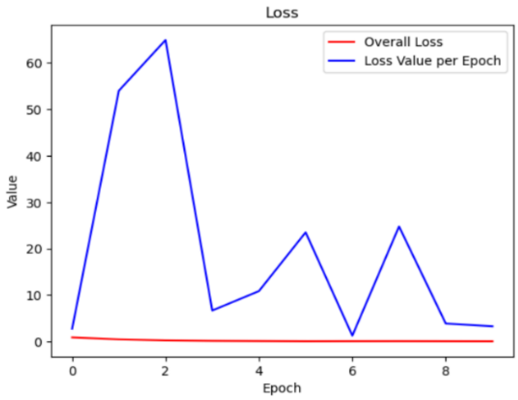
Figure 3: Proposed CNN model summary

IV. Experimentation Results:

To evaluate the performance of the model, bird sound dataset is downloaded (https://www.kaggle.com/code/collinpfeifer/5-bird-audio-classifier-model), which consist of 9107 audio files. 5000 audio files are used for training and remaining audio file are used for testing. Python language is used to implement the proposed technique. The proposed CNN model achieved 98.40 % accuracy. ROC curve on Accuracy Vs Epoch is illustrated in Fig. 4(a) and overall loss vs Epoch is shown in Figure 4(b).



(a) ROC curve on Accuracy Vs Epoch



(b) ROC curve on Loss Vs Epoch

Figure 4: Performance of the proposed Model

V. Conclusion:

Bird migration is a remarkable natural phenomenon that showcases the incredible adaptability and resilience of avian species. Through a combination of innate biological mechanisms and external environmental cues, birds undertake arduous journeys that are critical for their survival and reproduction. Continued research and conservation efforts are necessary to protect these migratory species and the ecosystems they depend on. The work presents a Convolution Neural Network model for bird sound recognition and experimentation results obtained are encouraging. This approach ensures that bird sound recognition systems are scientifically sound and practically useful, contributing to biodiversity monitoring and conservation efforts. Further, need of developing a local bird dataset and design of new algorithm to handle large number of bird classes.

References:

1. Akçay, H.G.; Kabasakal, B.; Aksu, D.; Demir, N.; Öz, M.; Erdoğan, A., "Automated Bird Counting with Deep Learning for Regional Bird Distribution Mapping". *Animals* 2020, *10*, 1207. https://doi.org/10.3390/ani10071207.
2. Stefan Kahl, Connor M. Wood, Maximilian Eibl, Holger Klinck, "BirdNET: A deep learning solution for avian diversity monitoring", Ecological Informatics, Volume 61,2021,101236, ISSN 1574-9541, https://doi.org/10.1016/j.ecoinf.2021.101236.
3. Andrea Flack, Ellen O. Aikens, Andrea Kolzsch, Elham Nourani, Katherine R. S. Snell, Wolfgang Fiedler, Nils Linek, Hans-Gunther Bauer, Kasper Thorup, Jesko Partecke, Martin Wikelski and Hannah J. Williams, "New frontiers in bird migration research", Current Biology, ISSN: 0960-9822, Vol: 32, Issue: 20, Page: R1187-R1199, October 24, 2022. https://doi.org/10.1016/j.cub.2022.08.028.
4. Sekercioglu C.H., Daily G. C., Ehrlich P. R., "Ecosystem consequences of bird declines". Proc. Natl. Acad. Sci. USA 2004, 101 (52), 18042-18047. https://doi.org/10.1073/pnas.0408049101.

5. Jiangjian Xie, Luyang Zhang, Junguo Zhang, Yanyun Zhang, Björn W. Schuller, "Cross-corpus open set bird species recognition by vocalization", Ecological Indicators, Volume 154, 2023, 110826, ISSN 1470-160X, https://doi.org/10.1016/j.ecolind.2023.110826.
6. Ricky Mohanty, Bandi Kumar Mallik, Sandeep Singh Solanki, "Automatic bird species recognition system using neural network based on spike", Applied Acoustics, Volume 161, 2020, 107177, ISSN 0003-682X, https://doi.org/10.1016/j.apacoust.2019.107177.
7. Zhang S, Gao Y, Cai J, Yang H, Zhao Q, Pan F., "A Novel Bird Sound Recognition Method Based on Multifeature Fusion and a Transformer Encoder". *Sensors*. 2023; 23(19):8099. https://doi.org/10.3390/s23198099
8. Indumathi C P, Diviyalakshmi K R, Mahalakshmi R, "Bird Sound Identification System using Deep Learning", Procedia Computer Science, Volume 233, 2024, Pages 597-603, ISSN 1877-0509. https://doi.org/10.1016/j.procs.2024.03.249.
9. Quan Tang, Liming Xu, Bochuan Zheng, Chunlin He, Transound: Hyper-head attention transformer for bird's sound recognition, Ecological Informatics, Volume 75, 2023, 102001, ISSN 1574-9541, https://doi.org/10.1016/j.ecoinf.2023.102001.
10. Shipeng Hu, Yihang Chu, Lu Tang, Guoxiong Zhou, Aibin Chen, Yurong Sun, "A lightweight multi-sensory field-based dual-feature fusion residual network for bird song recognition", Applied Soft Computing, Volume 146, 2023, 110678, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2023.110678.
11. Sylvain Christin, Éric Hervet, Nicolas Lecomte, "Applications for deep learning in ecology". Volume10, Issue10,October 2019, Pages 1632-1644. https://doi.org/10.1111/2041-210X.13256.
12. Groom, Geoff & Petersen, Ib & Anderson, Mark & Fox, A., "Using object-based analysis of image data to count birds: Mapping of Lesser Flamingos at Kamfers Dam, Northern Cape, South Africa". International Journal of Remote Sensing. 32. 4611-4639,2011. https://doi.org/10.1080/01431161.2010.489068.
13. Egevang, C., Stenhouse, I. J., Phillips, R. A., Petersen, A., Fox, J. W., & Silk, J. R. D., "Tracking of Arctic Terns Sterna paradisaea reveals longest animal migration". Proceedings of the National Academy of Sciences, 107(5), 2078-2081,2010. https://doi.org/10.1073/pnas.0909493107.
14. Gill RE, Tibbitts TL, Douglas DC, Handel CM, Mulcahy DM, Gottschalck JC, Warnock N, McCaffery BJ, Battley PF, Piersma T. "Extreme endurance flights by landbirds crossing the Pacific Ocean: ecological corridor rather than barrier?". Proc Biol Sci. 2009 Feb 7;276(1656):447-57. https://doi.org/10.1098/rspb.2008.1142. PMID: 18974033; PMCID: PMC2664343.
15. Vanderhoff, N., P. Pyle, M. A. Patten, R. Sallabanks, and F. C. James," American Robin (*Turdus migratorius*), version 1.0". In Birds of the World (P. G. Rodewald, Editor). Cornell Lab of Ornithology, Ithaca, NY, 2020, USA. https://doi.org/10.2173/bow.amerob.01.
16. Ali Zeraoula, Taqiyeddine Bensouilah, Hafid Brahmia, Zihad Bouslama, Moussa Houhamdi, Ahmed Kerfouf, "Breeding biology of the European Blackbird Turdus merula in orange orchards", Journal of King Saud University - Science, Volume 28, Issue 4, 2016, Pages 300-307, ISSN 1018-3647, https://doi.org/10.1016/j.jksus.2015.10.005.
17. Anderson, S. E., Dave, A. S., and Margoliash, D., "Template-based automatic recognition of birdsong syllables from continuous recordings". Journal of the Acoustical Society of America, 100(2), 1209-1219, 1996. https://doi.org/10.1121/1.415968.

18. Briggs F, Lakshminarayanan B, Neal L, Fern XZ, Raich R, Hadley SJ, Hadley AS, Betts MG., “Acoustic classification of multiple simultaneous bird species: a multi-instance multi-label approach”. J Acoust Soc Am. 2012 Jun;131(6):4640-50. DOI: https://doi.org/10.1121/1.4707424. PMID: 22712937.
19. Stowell, D., Wood, M. D., Pamuła, H., Stylianou, Y., & Glotin, H., “Automatic acoustic detection of birds through deep learning: The first Bird Audio Detection challenge. Methods in Ecology and Evolution”, 10(3), 368-380, 2019. https://doi.org/10.1111/2041-210X.13103
20. Salamon, J., & Bello, J. P., “Deep convolutional neural networks and data augmentation for environmental sound classification”. IEEE Signal Processing Letters, 24(3), 279-283, 2017. https://doi.org/10.1109/LSP.2017.2657381.
21. Stuart A. Brooker, Philip A. Stephens, Mark J. Whittingham, Stephen G. Willis, “Automated detection and classification of birdsong: An ensemble approach”, Ecological Indicators, Volume 117, 2020, 106609, ISSN 1470-160X, https://doi.org/10.1016/j.ecolind.2020.106609.
22. Xue Han, Jianxin Peng, “Bird sound classification based on ECOC-SVM”, Applied Acoustics, Volume 204,2023,109245, ISSN 0003-682X, https://doi.org/10.1016/j.apacoust.2023.109245.

ISSN: 2582-8118 ***Volume 4, Issue 5; Sept 2024***

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYSIS IN SCIENCE AND ENGINEERING

Web: https://www.iarj.in/index.php/ijrase/index

90. Phishing Detection Using Deep Learning: State of the Art and Future Directions

Rudragouda Hiregoudar
Department of Computer Science,
TKIET, Warananagar, Maharastra, India.

S. V. Saboji
Department of Computer Science,
Basaveshwar Engineering College,
Bagalkot, Karnataka, India.

Abstract:

One of the most common cybercrimes is phishing, which can be simply defined as observing entities and misrepresenting them to obtain sensitive information. Thus, this paper gives an overview of the techniques for phishing detection based on deep learning. This paper discusses various types of phishing attacks, their detection methodologies, and publicly available datasets for research. It assesses the effectiveness of various deep learning models for detecting phishing activities. Finally, we discuss the challenges and future directions towards improving the efficiency of phishing detection systems

Keywords:

Cyber Crime, Deep Learning, Phishing Detection.

I. Introduction:

The information technology revolution has endured for a long time, involving global communication, and nowadays, it is the dominant thing in the lives of over three billion people around the world. The internet's availability has significantly improved, its speeds have increased, it has expanded its applications, and its usefulness has increased. Political, business, cultural, and non-profit processes are greatly dependent on the internet [1].

Phishing is a form of online fraud that is akin to fishing. Like its counterpart, phishing aims to get people's personal data, just as it targets humans for their private information. No wonder that today, the latest updates on deep learning technology create a huge buzz among the Artificial Intelligence, data science, and machine learning communities, since this advanced technology extracts knowledge from the data it receives.

Over a million attacks were launched in the quarter of 2023, an increase from the prior quarter but a decline from the peak that was seen in the first quarter. The discontinuation of Freenom's free domain program earlier in the year, which was commonly exploited by fraudsters, led to a surge in phishing activities. After Freenom ceased its free domain registration services in January, there was a decline in phishing attacks using its domains by mid-2023[2].

Phishing remains one of the most common cybercrimes worldwide; Google blocks approximately 100 million phishing emails daily. Notably, 65% of groups employ spear phishing tactics for reconnaissance purposes. With the rise in phishing incidents, 96% of surveyed organizations reported experiencing at least one attack within the year, with 52% noting an increase in sophistication compared to previous threats [3].

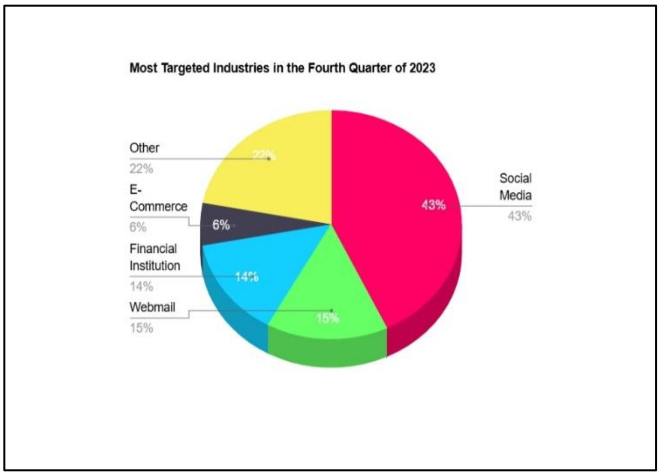


Figure 1: Industries Most Targeted During the Fourth Quarter of 2023.

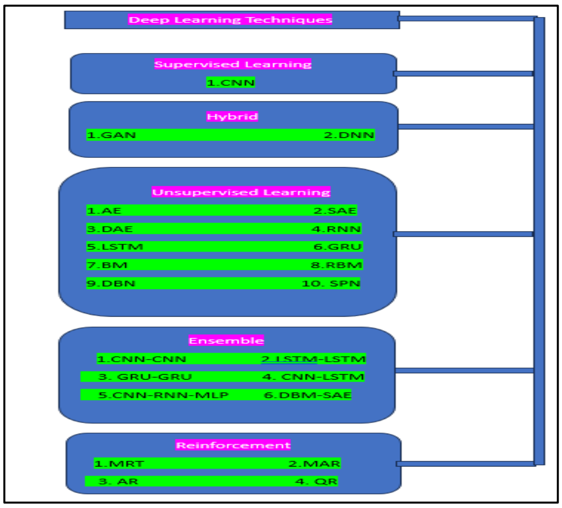


Figure 2: Taxonomy of Deep Learning phishing detection method [18]

Following are the Deep Learning (DL) techniques used in cybersecurity, particularly in phishing detection

1. Discriminative DL models:

Purpose: Supervised learning Classification, prediction, or recognition based on labeled data.

Examples: Convolutional Neural Network (CNN), Multilayer Perceptron [18]

2. Generative DL models:

Purpose: Unsupervised learning Modeling of input data distribution for generating random samples to improve the feature representation.

Examples: Autoencoder, Restricted Boltzmann Machine, Deep Belief Network.

3. Hybrid DL models:

Purpose: These combine discriminative and generative models to leverage the advantages of both.

Examples: Deep Neural Network (DNN), Generative Adversarial Network (GAN).

4. Ensemble DL Models:

Objective: Combine multiple DL algorithms in parallel or sequentially to improve overall performance.

Types:

Homogeneous ; e.g., CNN-CNN, LSTM-LSTM)

Heterogeneous ; e.g., CNN-LSTM, CNN-RNN-MLP).

5. Reinforcement Learning:

Objective: Adaptive learning approach using trial and error to achieve optimal behavior through feedback from numerical rewards.

Examples: Multi-task Reinforcement, Multi-agent Reinforcement, Asynchronous Reinforcement, Q-learning Reinforcement.

Key DL Techniques for Phishing Detection:

LSTM: Long Short-Term Memory: It deals effectively with the time-series sequence data and solves the problems of vanishing/exploding gradients from the traditional RNN.

CNN: Convolutional Neural Network: It efficiently and fastly extracts features from raw and possibly complex data, reducing network complexity and speeding the learning process. The observations from the research include the following:

Popularity: LSTM and BiLSTM are the most popular DL techniques for phishing detection, with 34% share each, followed by CNN with 30% share, DNN, and MLP at 8% each. GAN and DRL appear less frequently but are promising for further research.

Section 2 deals with the various types of phishing. The methodology of phishing detection has been presented in Section 3. The Dataset for Phishing is discusses in section 4, The literature survey is presented in Section 5. Section 6 shows the performance metrics. Section 7 finally deals with the conclusion and the future scope of the paper

II. Types of Phishing:

1.Website phishing: Hackers create fake websites that appear to be real [5]. These sites attract victims using emails, SMS, social media, and pop-ups [6]. Giving sensitive information results in financial loss.

2.Webpage Phishing: These are fake URLs and paperbacks of well-known websites. Hackers use the weaknesses of a hacked website and give it an original look. This procedure of bluffing users to reveal their personal information results in identity and financial theft.

3.Email Phishing: In this method, hackers design an email from a trusted source to fool the user. The email normally directs the user to a fraudulent website or downloads malware. Such emails are capable of stealing personal information along with viruses [6].

4.URL Phishing: The hackers create the fake URLs. Most of the strategies make use of these URLs that would drag people to open the phishing websites. This is usually done in order to steal data or to download malware.

III. Phishing Detection Deep Learning Methodology:

1. Input Data: We will collect URLs, email addresses, and the content from the web pages as input data for analysis. Examples include links from phishing websites and authentic URLs from numerous databases [8].

2. Feature Extraction: The feature extraction phase removes attributes such as the oddity of domain names, the length of URLs, and the existence of suspicious keywords. One example is machine learning's ability to find the structure and content patterns of URLs.

3. Feature Selection: This is the list of features that should be most useful in identifying the phishing attack. Example: using mutual information or Chi-square to reduce the feature set [10].

4. Classification Techniques: Classify the data using techniques such as decision trees, random forests, support vector machines, and neural networks. If high accuracy is desired, then employ SVM with Random Forests to deal with high-dimensional data [11].

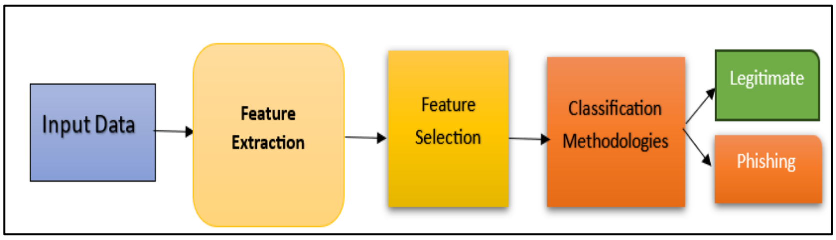


Figure 3: Phishing Detection Scenario

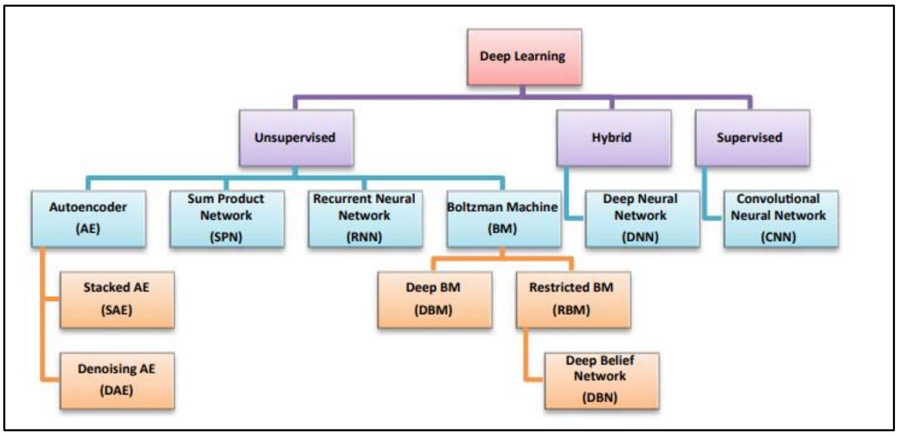


Figure 4: Classification of Deep Learning Techniques [3].

IV. literature Survey:

Yang and team employed an RNN to process multidimensional features in the IEEE dataset, achieving an accuracy of 96.50%. The model demonstrated proficiency in capturing sequential data but encountered difficulties with long-term dependencies. The emphasis was on achieving real-time detection and effectively handling sequential dependencies [10]. Alzahrani and colleagues (2023) utilized CNN with URL features from Phish Tank and manually collected legitimate sites, achieving an accuracy of 98.77%. Their study highlights the importance of effective feature extraction for high accuracy, emphasizing the need for large labeled datasets to balance model complexity [11]. HELPHED enhances phishing email detection by incorporating Ensemble Learning with hybrid features containing email content and textual traits. In this research, two techniques have been used that base their judgement on Stacking Ensemble Learning and Soft Voting Ensemble Learning. The former uses decision trees, KNN, etc. This would reduce the complexity and improve the models' performance with a high F1-score of 0.9942[12]. THEMIS [13] An advanced model of recurrent-convolution-neural-networks-based solution having attention mechanisms and multilevel vectors—achieved an accuracy of 99.848%, with a very low false positive rate of 0.043%. THEMIS is a state-of-the-art, new approach in the detection of phishing emails, developed using TensorFlow and Keras. The training set contained 5,447 clean and 699 phishing emails, while the validation set contained 2,334 clean and 300 phishing emails. THEMIS uses RCNN powered by LSTM due to the ability of the latter to overcome long-term dependencies, while the attention mechanism improves the focus on relevant elements in a manner somewhat similar to human cognition. Mughaid, A., et al. [14] developed a boosted decision tree model for the mitigation of phishing attacks by identifying the inherent characteristics of email text. This model categorized the non-phishing and phishing emails, which also validated the test data by splitting the database for training. This approach effectively utilized an imbalanced database and fewer features for the prediction of phishing emails. However, this approach failed to utilize automated tools for the extraction of new features from raw emails. Benavides-Astudillo, E., et al. [15] designed a Bidirectional Gated Recurrent Unit (BiGRU) for the detection of phishing web pages by extracting words from Hypertext Markup Language (HTML) codes of web pages. Here, the syntactic and semantic features were considered for the analysis and processing of embedding words. This approach effectively preserved the intrinsic richness of texts and recorded minimum execution time while detecting phishing or benign web pages. However, this approach failed to utilize word embedding techniques, like word2Vec and Fast Text for the computation of its performance in the Natural Language Processing (NLP) domain. Tang, L. and Mahmoud, Q.H.[16], proposed Recurrent Neural Network-Gated Recurrent Unit (RNN GRU).The system effectively detected phishing in real time without interruptions. However, the method required an extended training period. Gupta, B.B., et al. [17] (2021) Lexical-based RF The method was highly reliable, required low computational power and resources, and detected phishing URLs within a short duration. However, no sophisticated deep-learning algorithms were incorporated to boost the detection rate.

V. Phishing Detection Performance Metrics Making Use of Deep Learning:

Performance measurements are essential for assessing how well phishing detection methods work. This section goes over the standard measures that are used in academic articles to evaluate how well deep learning-based phishing detection systems are working.

1. Accuracy: The ratio of accurately anticipated cases—both phishing and legitimate—to the total number of cases is known as accuracy. It offers an indicator of the model's overall performance.

Accuracy= {TP+TN}/{FP+FN+TP+TN}

TP (True Positives): Correctly identified phishing instances.

TN (True Negatives): Correctly identified legitimate instances.

FP (False Positives): Legitimate instances incorrectly identified as phishing.

FN (False Negatives): Phishing instances incorrectly identified as legitimate.

2. Precision: Precision measures the accuracy of the positive predictions (i.e., how many of the predicted phishing instances are actually phishing).

Precision ={TP}/{TP + FP}

3. Recall (Sensitivity or True Positive Rate): Recall measures the ability of the model to identify all relevant instances (i.e., how many of the actual phishing instances are correctly identified)

Recall=TP/{TP+FN}

4. The F1-Score: It is the harmonic mean of precision and recall, providing a balance between the two metrics.

F1Score=2×Precision×Recall/Precision +Recall

5. Matthews Correlation Coefficient (MCC): MCC considers all four confusion matrix categories (TP, TN, FP, FN) and provides a balanced measure, even if the classes are of very different sizes.

Dataset	Description	Website
Spam Assassin	Contains information on phishing sites reported by users.	https://spamassassin.apache.org
Enron Email Dataset	Contains emails from Enron employees, which can be used for phishing detection research.	https://www.cs.cmu.edu/~enron/
UCI	Contains URLs labelled as phishing or legitimate	https://archive.ics.uci.edu/dataset/327/phishing+websites
Phishing Email Dataset (Mendeley)	Includes a collection of phishing and legitimate emails for research and detection.	https://data.mendeley.com/
Microsoft Malware	Contains emails and attachments	https://www.microsoft.com/en-us/research/project/msrc-malware-collection/

Dataset	Description	Website
Classification Dataset	used to classify malware, including phishing examples	

MCC={TP x TN}-{FP x FN}/

√{TP+FP}x{TP+FN}x{TN+FP}x{TN+FN}

VI. Dataset for Phishing:

TABLE I

Publically Available Dataset for Email Phishing Detection

TABLE II

Publically Available Dataset for Website Phishing Detection

Dataset	Description	Website
Phish Tank	Contains information on phishing sites reported by users.	https://www.phishtank.com
Alexa Top Sites	A list of top websites, which can be used to compare with known phishing sites.	https://www.alexa.com
UCI	Contains URLs labeled as phishing or legitimate	https://archive.ics.uci.edu/dataset/327/phishing+websites
Kaggle Phishing	A dataset with features extracted from URLs, labeled as phishing or legitimate.	https://www.kaggle.com/datasets
Open Phish	Provides data on phishing websites	https://openphish.com/

Dataset	Description	Website
	and their characteristics.	
Yahoo	Contains URLs and features for phishing detection, collected from Yahoo's search engine	http://dir.yahoo.com

VII. Conclusion and Future Work:

Phishing remains the most frequent and serious threat to cybersecurity across the world, breaking people's trust in the theft of sensitive information. This paper gave an outline of various deep-learning-based techniques in the detection of phishing attacks, focused on state-of-the-art advancement and model effectiveness, such as RNNs, CNNs, and Ensemble Learning. However, research has not been conducted on dataset quality, feature engineering, model complexity, real-time detection, and adversarial attack handling.

Future research in phishing detection should be informed by robust, upto date datasets from continuous data collection and collaboration. Optimizing the detection algorithms for real-time performance and exploring edge computing can improve the detection capability by reducing latency and bandwidth usage. Advanced feature engineering, including automated feature extraction and semantic analysis via NLP, will further improve adaptability to new phishing tactics. Subsequently, hybrid models using deep learning techniques and adversarial training will make the phishing detection systems resilient and effective in hostile environments.

References:

1. Yang, J.; Cheng, G.; Shen, M. Secure fusion of encrypted RSI based on Brovey. Sci. China Inf. Sci. 2021, 64, 1–3.
2. Boulemtafes, A.; Derhab, A.; Braham, N.A.A.; Challal, Y. PReDIHERO–Privacy-Preserving Remote Deep Learning Inference based on Homomorphic Encryption and Reversible Obfuscation for Enhanced Client-side Overhead in Pervasive Health Monitoring. In Proceedings of the 2021 IEEE/ACS 18th International Conference on Computer Systems and Applications (AICCSA), Tangier, Morocco, 30 November–3 December 2021; IEEE: New York, NY, USA; pp. 1–8.
3. Chouragade, P.M.; Ambhore, P.B. A Survey on Privacy Preserving Content Based Image Retrieval and Information Sharing in Cloud Environment. In Proceedings of the 2021 4th International Conference on Recent Trends in Computer Science and Technology (ICRTCST), Jamshedpur, India, 1–12 February 2022; IEEE: New York, NY, USA; pp. 238–245.
4. Saravanan, K.; Suganthi, K.; Kalaiselvi, R.; Divya, B.; Kumar, S.D. Privacy Preserving On Remote Sensing Data Using Reversible Data Hiding. J. Phys. Conf. Ser. 2021, 1979, 012050.

5. Wang, X.; Li, J.; Yan, H. An improved anti-quantum MST3 public key encryption scheme for RSI. Enterp. Inf. Syst. 2021, 15, 530–544.
6. Hasan, M.M.; Islam, M.U.; Sadeq, M.J.; Fung, W.K.; Uddin, J. Review on the Evaluation and Development of Artificial Intelligence for COVID-19 Containment. Sensors 2023, 23, 527.
7. Xie, D.; Chen, F.; Luo, Y.; Li, L. One-to-many image encryption with privacy-preserving homomorphic outsourced decryption based on compressed sensing. Digit. Signal Process. 2019, 95, 102587.
8. Du, G.; Zhou, L.; Li, Z.; Wang, L.; Lü, K. Neighbor-aware deep multi-view clustering via graph convolutional network. Inf. Fusion 2023, 93, 330–343.
9. Sirisha, U.; Chandana, B.S. Privacy Preserving Image Encryption with Optimal Deep Transfer Learning Based Accident Severity Classification Model. Sensors 2023, 23, 519.
10. Li, S.; Wu, L.; Meng, W.; Xu, Z.; Qin, C.; Wang, H. DVPPIR: Privacy-preserving image retrieval based on DCNN and VHE. Neural Comput. Appl. 2022, 34, 14355–14371.
11. Alsubaei, F.S., Alneil, A.A., Mohamed, A. and Mustafa Hilal, A., 2023. Block-Scrambling-Based Encryption with Deep-Learning-Driven Remote Sensing Image Classification. *Remote Sensing*, *15*(4), p.1022.
12. Alkhelaiwi, M., Boulila, W., Ahmad, J., Koubaa, A. and Driss, M., 2021. An efficient approach based on privacy-preserving deep learning for satellite image classification. *Remote Sensing*, *13*(11), p.2221.
13. Yang, X., Cao, W., Lu, Y. and Zhou, Y., 2022. Self-supervised learning with prediction of image scale and spectral order for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, pp.1-15.
14. Chen, Q., Wu, Y., Wang, X., Jiang, Z.L., Zhang, W., Liu, Y. and Alazab, M., 2023. A Generic Cryptographic Deep-Learning Inference Platform for Remote Sensing Scenes. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *16*, pp.3309-3321.
15. Kuang, W., Tu, B., He, W., Zhang, G. and Peng, Y., 2021. A spectral-spatial attention aggregation network for hyperspectral imagery classification. *International Journal of Remote Sensing*, *42*(19), pp.7551-7580.
16. Al-Khasawneh, M.A., Uddin, I., Shah, S.A.A., Khasawneh, A.M., Abualigah, L. and Mahmoud, M., 2022. An improved chaotic image encryption algorithm using Hadoop-based MapReduce framework for massive remote sensed images in parallel IoT applications. *Cluster Computing*, pp.1-15.
17. Marques, G., Agarwal, D. and De la Torre Díez, I., 2020. Automated medical diagnosis of COVID-19 through Efficient Net convolutional neural network. *Applied soft computing*, *96*, p.106691.
18. Kuyumani, E.M., Hasan, A.N. and Shongwe, T., 2023. A Hybrid Model Based on CNN-LSTM to Detect and Forecast Harmonics: A Case Study of an Eskom Substation in South Africa. *Electric Power Components and Systems*, *51*(8), pp.746-760.
19. Chen, S. and Zheng, J., 2023. Sand Cat Arithmetic Optimization Algorithm for Global Optimization Engineering Design Problems. *Journal of Computational Design and Engineering*, p. qwad094.

ISSN: 2582-8118 ***Volume 4, Issue 5; Sept 2024***

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYSIS IN SCIENCE AND ENGINEERING

Web: https://www.iarj.in/index.php/ijrase/index

91. Big Data Based Security Analytics for Protecting the Virtualized Infrastructure in Cloud Computing

Vishwanath Kagawade

Department of Artificial Intelligence and Machine Learning
Basaveshwara Engineering College, Bagalkote, India.

Shweta I. Haigar

Department of Computer Science and Engineering
Basaveshwara Engineering College, Bagalkote, India.

ABSTRACT:

Cloud computing virtualized infrastructure has become a target for cyber attackers aiming to execute sophisticated attacks such as ransomware, distributed denial-of-service (DDoS) attacks, and advanced persistent threats (APTs). Our proposed technology could be utilized to develop a big data-based security analytics technique for identifying advanced assaults in virtualized infrastructures. In the proposed work, a new system is developed to detect phishing attempts by analyzing URL structures to handle APTs attacks. As sophisticated cyber threats, including Advanced Persistent Threats (APTs), become more prevalent, there is an increasing need for robust security measures in cloud computing environments. The proposed work explores a newly developed system designed to detect phishing attempts by analyzing URL structures. The system is integrated into a web application to manage user-generated content, incorporating various security features to safeguard against cyber threats.

KEYWORDS:

Cloud, Dijango, Firebase, Security, Phishing

I. Introduction:

Cloud computing brings with its security issues that should be considered before utilizing technology to its best. Risk and compliance management concerns with the evaluation, implementation, and deployment of solutions, identities and access control, service continuity, and endpoint interoperability should be taken into account. When talking about infrastructure as a service, platform as a service, or software as a service, cloud security becomes a complex issue. The work aims to develop a comprehensive web application using

Django to manage user-generated posts, incorporating user authentication, content management, and protection features. The objectives include creating a robust notification system and integrating advanced security measures such as steganography, cryptography, and URL phishing detection. The work focuses on providing a user-friendly interface for interactions and a secure environment for content sharing. Since the internet's creation and growth, three data-driven platforms have emerged: cloud computing, big data, and virtualization. They still have the upper hand in terms of data flow, control, and preservation for a wide spectrum of big and small businesses. Users' ability to adjust to cloud technologies is limited by security concerns. Inadequate security measures have alarmed cloud service providers, and problems with data integrity, control, audit, confidentiality, and availability need to be fixed [1]. Cloud computing is a popular technology these days. Microsoft, Google, and Amazon are all enhancing their consumer services [2]. In order to utilize cloud infrastructure, users of cloud computing can also develop and launch applications on the cloud [3]. Develop a system to detect phishing attempts by analyzing URL structures. Implement machine learning models or use existing libraries for URL analysis. Messages and data are encrypted so that only the sender and the intended recipient can read them. It is a type of messaging security and the study of communications. In the end, cryptography can prevent data from being lost or stolen. Additionally, it can be applied to user authentication. To keep electronic data and messages private and only readable by the intended parties, cryptography frequently uses encryption and an algorithm. Steganography is a method that hides data inside a regular, non-secret file or message so that it won't be discovered; the hidden data is then extracted when it gets to its intended location. Data hiding or protection can be further enhanced by combining steganography with encryption. Digital content of any kind, including text, images, videos, and audio, can be hidden via steganography. Almost any other kind of digital material can include the secret data contained within. Since Google hosts content created with Sites and Firebase, it may be found via a Google URL. Users can therefore trust the website when they see the Google domain and want to confirm if a page is authentic protects people against harmful actions by focusing on detecting and stopping phishing attempts. Firebase is a collection of platforms for application development and backend cloud computing services offered by Google. Users can access resources and services through the Internet from anywhere at any time thanks to cloud computing.

A. Big Data:

The term "Big data" refers to enormous volumes of data from numerous databases that are growing exponentially. The Big Data paradigm emerged as a response to the incapacity of traditional database systems to handle enormous datasets [4]. Every day, enormous volumes of data are transferred through computerized systems from the stock exchange, banking, internet, and in-person purchases. This data is then collected and stored for use in inventory management, customer behavior, and market behavior [5]. The following user application and network logs were gathered from a virtualized architecture; this presents a big data challenge for our suggested method.

- **Big Data Security:**

Privacy and security issues are the main danger that comes with large data. The problem is that analytics engineers are tinkering with personal or unlawful data to get incorrect findings. Because Big Data is made up of a huge number of interconnected data sets, it necessitates its own security against unauthorized access.

- **Data Security Analytics:**

The goal of big data security analytics, or BDSA, is to provide users with an up-to-date and comprehensive view of IT activities so they can act swiftly and intelligently. Security and intelligence services can benefit from big data analytics as it helps to reshape security intelligence [4]. Any gadget with an internet connection is open to attack. Therefore, data pertaining to security needs to be kept for both pre- and post-analytical assessments. Big Data has a lot of difficulties [5].

B. Virtual Infrastructure:

Virtual infrastructure may save a lot of money and effort and is necessary to manage the IT environments of many firms. Physical infrastructure requires a lot of space and can be readily replaced by virtual infrastructure. Virtual machines (VMs), which are components of virtualized infrastructure, are dependent on the software-defined multi-instance resources of the hosting hardware. Virtual machine monitors, or hypervisors, support, govern, and manage software-defined multi-instance architectures [8]. Hypervisors should be installed in place of OSs since virtual infrastructure makes use of physical infrastructure. Virtual machines run a specific type of software called a hypervisor.

The remaining part of the article is organized as follows: Section II describe the state-of-the-art-techniques. Section III describes proposed system architecture. Section Results are presented in section IV. Conclusion is presented in section V.

II. Literature Survey:

This section presents systematic literature review in the domain Big Data Based Security Analytics for Protecting Virtualized Infrastructure in Cloud Computing (BDSA-PVI). Existing solutions and technologies offer various ways to manage user content and ensure security in web applications. Django, a high-level Python web framework, is widely used for its ease of use and comprehensive features, including built-in user authentication and database management. Steganography and cryptography provide advanced methods for data protection, while URL phishing detection techniques are crucial for maintaining user trust and safety in web applications. This project leverages these technologies to create a secure and efficient content management system. In the work of Dr. Deepali Gupta and Ashima Naran, 2018[1], foundation of cloud resource sharing is virtualization. Cloud security is another major issue. Research is being done to increase the use of cloud storage even more, as it is a rapidly expanding trend. Deepika Goyal, Priyanka, 2018[2], propose an underpinning preliminary inquiry towards ensuring the Protecting Virtualized Infrastructures (PVI). The suggested system can even be used by virtual clusters for secure communication, and it can be implemented as firmware on cloud resources. Olasupo Ajayi et al., 2017[3], Virtualization technology utilized in cloud computing. This makes it possible to host virtual machines on actual servers. Research is still being conducted in an effort to further push the limits of cloud storage adaption, as cloud storage adoption continues to grow at an astounding rate. In the work of Dr. M Siddappa, Siddaramaiah S R, and N L

Udaya Kumar, 2015[4], Virtualization has regained significance as a means of enhancing system security and augmenting the usage of underlying resources through the creation of virtual replicas of physical cloud services. Both the data processing procedures and the Big Data security components are examined. Finally, the potential applications of big data are discussed. A novel approach to security of big data is suggested. Darshan Tank et al., a recent thorough analysis of the risks and weaknesses associated with virtualization is provided. A recent thorough analysis of the risks and weaknesses associated with virtualization is provided. To ensure secure virtualized implementations, general security requirements.

III. System Architecture:

The different phases involved in the proposed technique is discussed in this section and depicted in Figure.1.

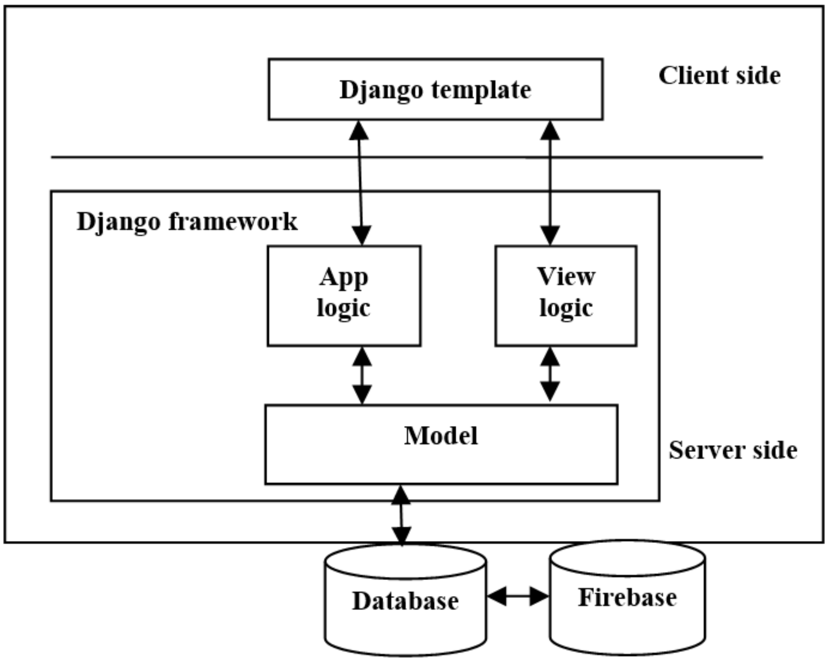


Figure.1: The Architecture Diagram of the Proposed Technique

Figure 1, depicts the high-level structure of proposed method including servers, databases, frameworks, and how they communicate. **App Logic:** Encapsulates functionality specific to each application, including models, views, forms, and URLs. **View Logic:** Handles user interactions, retrieves data from models, and renders templates. **Client-Side:** Runs in the user's browser, managing UI interactions and dynamic content updates. **Server-Side:** Runs on the server, processing data, business logic, and database interactions. **SQLite Database:** Default database for Django, lightweight and file-based, suitable for development and small projects. **Firebase:** Firebase is a platform for application development and backend cloud computing services offered by Google. **Template System:** Separates presentation from logic, using Django template language for dynamic HTML generation.

A. System Overview:

The primary goal of the proposed system is to enhance the security of cloud-based web applications by detecting phishing attempts and handling APT attacks through advanced techniques in URL analysis.

B. System Development:

- ***Web Application:***

Framework: The web application is developed using the Django framework, known for its scalability and robustness. Django facilitates the creation of a secure and efficient environment for managing user-generated posts.

Features: The application includes essential features such as user authentication, content management, and protection mechanisms to ensure secure interactions and data handling.

- ***Backend Services:***

Firebase: Chosen for its cloud computing services, Firebase supports the application's backend functionalities, including real-time database management and scalable infrastructure. This integration allows for efficient data handling and enhances the application's overall performance.

- ***Security Enhancements:***

Steganography: The system incorporates steganography to conceal sensitive information within other data formats, adding an extra layer of security by making it harder for unauthorized entities to access critical information.

Cryptography: Cryptographic techniques are employed to secure data transmissions and storage, ensuring that user data and communication remain confidential and protected from unauthorized access.

URL Phishing Detection: A novel approach is implemented to detect phishing attempts by analyzing URL structures. This feature helps identify potentially malicious URLs that could be used in phishing attacks, thus preventing unauthorized access and maintaining user trust.

IV. Results:

The proposed model successfully integrates user authentication, content management, and protection features in a Django web application. The notification system efficiently alerts users to important events, ensuring an interactive and engaging user experience. In this model, user registration module is used to create an account by providing needed information and click the create button so that user can login into the application. The login module of the work is shown in below Figure 3.

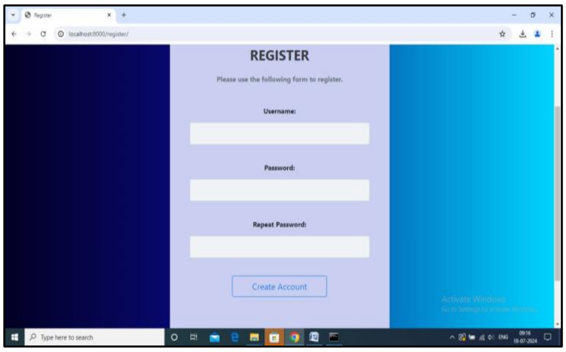


Figure 3: User Registration

Once the registration is done then the user can enter username and password in order to login into the application as shown in Figure. 4.

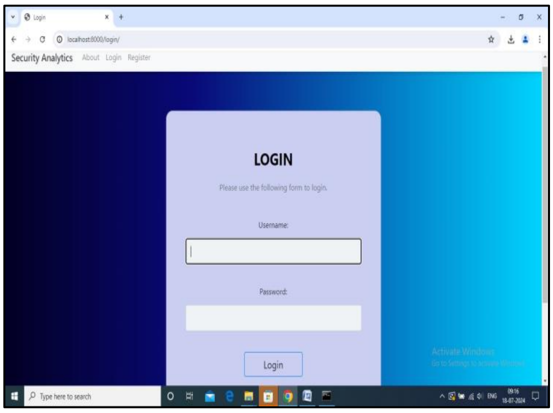


Figure 4: User Login

Figure. 5 illustrates how users communicate, share data, and build content security in the model.

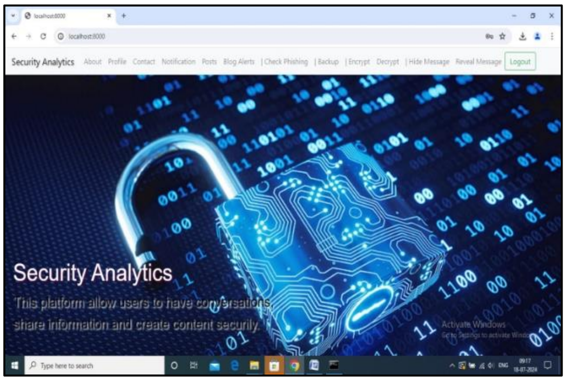


Figure 5: Content Security

Users can alter their profiles and change their passwords as shown in the below Figure. 6.

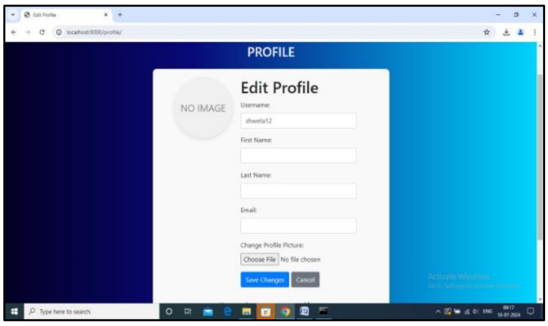


Figure 6: Profile Editor

Users of the application or website can enter data in the contact form that is visible to viewers. Like sending an email, everything you have to do is type your message right into the webpage rather than using an email program in Figure. 7, users are able to upload videos and pictures.

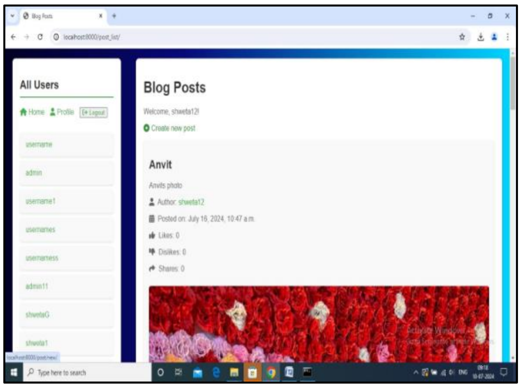


Figure 7: Blog Posts

If someone downloads a specific post, it notifies them with a download alert as shown in Figure. 8.

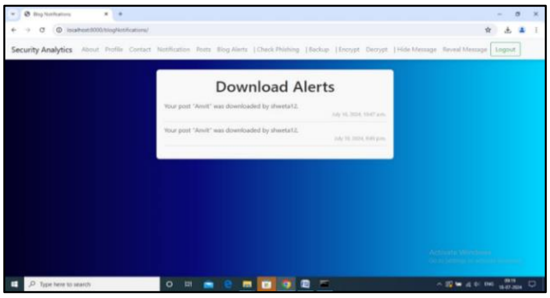


Figure. 8: Download alert

To recognize and prevent URL phishing attempts. If the provided URL is secure, then it indicates that it is phishing else it will show that the provided URL is not phishing as shown in Figure. 9.

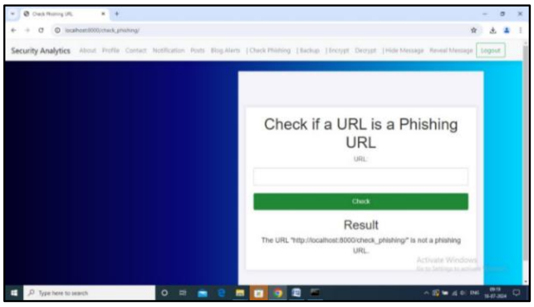


Figure. 9: URL Phishing

To ensure confidentiality, encrypt user data using a secure algorithm as shown in the Figure. 10.

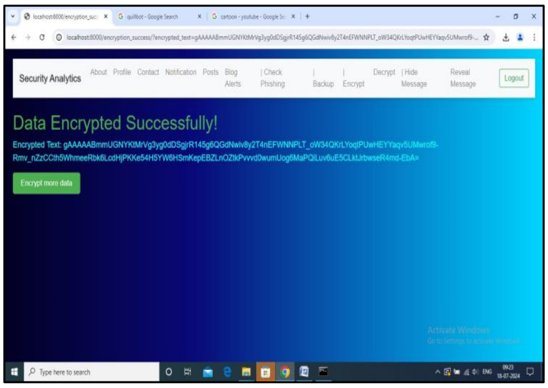


Figure 10: Encryption

When necessary, decrypt encrypted data to restore it to its original plain text format as shown in Figure. 11.

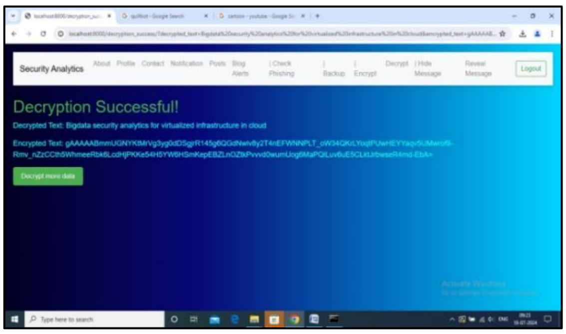


Figure 11: Decryption

A specific image's and a specific content can be hidden by the user as shown in Figure. 12.

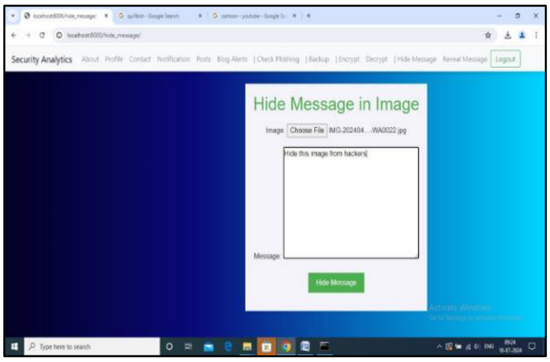


Figure 12: Hide Message and Image

A specific image and a specific content can be revealed as shown in Figure. 13.

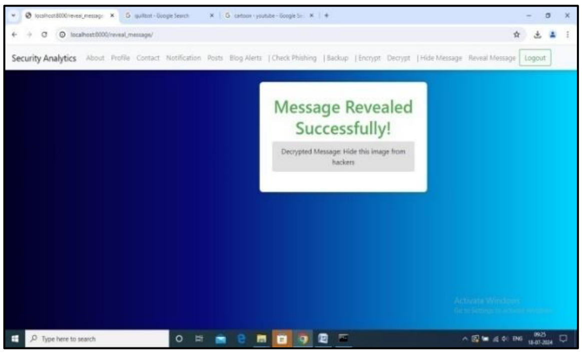


Figure 13: Reveal Message

V. Conclusion:

The proposed model achieves its objectives of creating a secure, user-friendly content management system using Django. The work demonstrates the effective use of Django and other technologies to develop a robust web application. Model incorporates critical tools for managing user roles, moderating content, and overseeing system operations. These features are crucial for maintaining the platform's integrity, ensuring compliance with guidelines, and providing a secure environment for all users. The inclusion of advanced security measures, such as data encryption, phishing detection, and session security management, highlights the work commitment to safeguarding user data and enhancing overall system security. Overall, the model successfully addresses the needs of both users and administrators, providing a well-rounded and secure web application. The detailed use case diagram ensures a clear understanding of system operations, facilitating efficient implementation and management of the platform. Future work could focus on improving the user interface, adding more advanced security features, and integrating real-time notifications.

References:

1. Ashima Narang and Dr. Deepali Gupta, "A Review on Different Security Issues and Challenges in Cloud Computing", International Conference on Computing, Power and Communication Technologies (GUCON) Galgotias University,2018.
2. Priyanka, Deepika Goyal "Protecting Virtualized Infrastructure in Cloud using Amplified Security Techniques*",* International Journal of Innovative Research in Computer and Communication Engineering, 2018.
3. Isaac Odun-Ayo, Olasupo Ajayi and Chinonso Okereke," Virtualization in Cloud Computing: Developments and Trends ", International Conference on Next Generation Computing and Information Systems (ICNGCIS), 2017.
4. Ripon Patgiri, and Umakanta Majhi," Big Data Security Analytics: Key Challenges", Int'l Conf. Data Science, ICDATA'18, 2018.
5. Raghav Toshniwal, Kanishka Ghosh Dastidar and Asoke Nath," Big Data Security Issues and Challenges*",* International Journal of Innovative Research in Advanced Engineering (IJIRAE), 2015.
6. Amira Hosni, "Cloud Computing Virtualization Challenges", International Journal of Advanced Research (IJAR),2017.
7. Isaac Odun-Ayo, Olasupo Ajayi and Chinonso Okereke, "Virtualization in Cloud Computing: Developments and Trends ", International Conference on Next Generation Computing and Information Systems (ICNGCIS), 2017.
8. R.K. Monika and K. Ravikumar, "Protecting Virtualized Infrastructures In Cloud Computing Based On Big Data Security Analytics*",* ICTACT JOURNAL ON SOFT COMPUTING, 2021.
9. Thu Yein Win, Huaglory Tianfield, and Quentin Mair, "Big Data Based Security Analytics for Protecting Virtualized Infrastructures in Cloud Computing", *IEEE*, 2016.
10. Deshpande Chandrika, Dr. M. Sreedhar Reddy "Protecting Virtualized Infrastructures in Cloud Computing Based on Big Data Security Analytics", International Journal of Advanced Information Science and Technology (IJAIST), 2017.
11. Meryeme ALOUANE and Hanan EL BAKKALI "Virtualization in Cloud Computing: Existing solutions and new approach", Information Security Research Team - ISeRT, IEEE, 2016.
12. N L Udaya Kumar, Siddaramaiah S R and Dr M Siddappa "Security Issues and Solutions for Virtualization in Cloud Computing Service", International Journal of Engineering Research & Technology (IJERT),2015.
13. Darshan Tank, Akshai Aggarwal & Nirbhay Chaubey, "Virtualization vulnerabilities, security issues, and solutions: a critical study and comparison*",* International Journal of Information Technology, 2014.
14. J.Yamuna , Francis Xavier "Big Data Based Security Analytics for Protecting Virtualized Infrastructures in Cloud Computing", International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) ,2018,
15. José Moura, Carlos Serrão," Security and Privacy Issues of Big Data ", ISCTE-*IUL*, Instituto Universitário de Lisboa, 2021.
16. Amir DJENNA, Mohamed BATOUCHE, "Security Problems in Cloud Infrastructure", *IEEE*, 2014.
17. Dr. K. Butchi Raju, Dayakar Suddala, "Big Data based Security Analytics to Protect the Virtualized Infrastructure", International Journal of Engineering Development and Research, 2018.

18. Dr. S. S. Manikandasaran, S. Raja and K Bhalaji, "Infrastructure Virtualization Security Architecture Specification for Private Cloud*", international* journal of computer sciences and engineering, 2018.
19. D. P. Acharjya, Kauser Ahmed P," A Survey on Big Data Analytics: Challenges, Open Research Issues and Tool", *(IJACSA)* International Journal of Advanced Computer Science and Applications, 2016.
20. Muhammad Kazim, Rahat Masood, Muhammad Awais Shibli, and Abdul Ghafoor Abbasi," Security Aspects of Virtualization in Cloud Computing ", National University of Sciences and Technology, 2017.

ISSN: 2582-8118 ***Volume 4, Issue 5; Sept 2024***

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYSIS IN SCIENCE AND ENGINEERING

Web: https://www.iarj.in/index.php/ijrase/index

92. Design and Development of Novel Machine Vision-based Non-Destructive Assessment Techniques for Post-Harvest Fruit Quality Control

Mr. Manjunath Y. Kammar, Mr. Virupaksha Gouda

Research Scholar, Dept. Of CSE-DS,

"Ballari Institute of Technology and Management", Bellary,

Visvesvaraya Technological University, Belagavi, Karnataka, India.

Dr. Aradhana D.

Professor, "Ballari Institute of Technology and Management", Bellary,

Visvesvaraya Technological University, Belagavi, Karnataka, India.

Abstract:

India ranks first in production of Bananas (25.7%), Papayas (43.6%) and Mangoes (40.4%) etc. India's fruit processing sector is now very decentralized. Many units are in the cottage, home-scale, and small-scale sectors, with capacities as low as 250 tons annually, despite the fact that major Indian and international corporations have capabilities of roughly 30 to 50 tons per hour. After China, it comes in second place in the world for fruit output. According to the National Horticulture Database published by NHB, 90.2 million metric tons of fruits were produced in India in 2015–16. India is one of the world's top producers of all deciduous fruits, primarily grown in the Indian states of Jammu and Kashmir, Nagaland, Himachal Pradesh, and Maharashtra. These fruits include apples and pear, peach and plum, berry, and cherry. Non-destructive analysis to examine fruits at their surface without using any invasive methods that compromise the quality and aspect of the food. Nowadays, a number of fruits are used commercially for the inspection, sorting, and grading of their surface qualities, non-destructive evaluation of internal quality, like texture and flavor. However, presents significant technical obstacles. the consistency of form and size, the lack of flaws, and the color of the skin and flesh. Fruit quality considerations include the following: texture-related characteristics including turgidity, hardness, and softness; maturity; firmness; uniformity of size and shape; and the absence of faults in skin and flesh color. The goal of the current multidisciplinary research project is to create a unique non-destructive quality control method for fruits by combining spectral analysis of internal quality attributes with analysis of the fruit photos' external appearance attributes.

Keywords:

Non-invasive quality assessment, local agricultural commodities, physical and chemical properties. Machine Learning.

ABSTRACT:

India is the world's leading producer of fruits. India ranks first in production of Bananas (25.7%), Papayas (43.6%) and Mangoes (40.4%) etc. India's fruit processing sector is now very decentralized. Many units are in the cottage, home-scale, and small-scale sectors, with capacities as low as 250 tons annually, despite the fact that major Indian and international corporations have capabilities of roughly 30 to 50 tons per hour. After China, it comes in second place in the world for fruit output. According to the National Horticulture Database published by NHB, 90.2 million metric tons of fruits were produced in India in 2015–16. India is one of the world's top producers of all deciduous fruits, primarily grown in the Indian states of Jammu and Kashmir, Nagaland, Himachal Pradesh, and Maharashtra. These fruits include apple and pear, peach and plum, berry, and cherry. Non-destructive analysis to examine fruits at their surface without using any invasive methods that compromise the quality and aspect of the food. Nowadays, a number of fruits are used commercially for the inspection, sorting, and grading of their surface qualities; non-destructive evaluation of internal quality, like texture and flavor, however presents significant technical obstacles. the consistency of form and size, the lack of flaws, and the color of the skin and flesh. Fruit quality considerations include the following: texture-related characteristics including turgidity, hardness, and softness; maturity; firmness; uniformity of size and shape; and the absence of faults in skin and flesh color. The goal of the current multidisciplinary research project is to create a unique non-destructive quality control method for fruits by combining spectral analysis of internal quality attributes with analysis of the fruit photos' external appearance attributes.

KEYWORDS:

Index Terms — Non-invasive quality assessment, local agricultural commodities, physical and chemical properties. Machine Learning.

I. Introduction:

Automation in agriculture science raises national production, economic prosperity, and quality. Fruit assortment has an impact on the export market and the assessment of quality. One important sensory factor that affects the market value, consumer preference, and choice of fruits is their appearance.

Although humans are capable of sorting and grading, the process is inconsistent, subjective, costly, time-consuming, and easily influenced by external factors.

It also causes delays in the post-harvest processes, which cost extra money. Therefore, a fruit grading system is needed. Recently, a number of academics have developed several sorting and grading algorithms using computer vision.

Using destructive techniques is the most common method for assessing the quality of agricultural products. Fruits are measured for their internal and external characteristics to evaluate their ripeness and maturity. Texture, color, sugar content, and other parameters can be measured using a destructive method.

Destructive methods are more successful in determining the quality of fruits; nonetheless, there are practical issues with cost, effectiveness, and reliability. Therefore, in order to satisfy consumer demand for fresh and healthful fruits, quick, portable, and affordable methods without compromising fruit quality must be developed. When grading or classifying fruits, visual appearance is a crucial way to determine their external characteristics (size, color). Numerous measurement systems that are non-destructive (contact) and non-contact have been introduced for fruit on-plant and lab-level quality inspection. Among the touch and non-contact quality measurement technologies are Near Infrared Spectroscopy (NIRS), Electronic Nose, Machine Vision, Acoustic and Ultrasonic Methods. With these techniques, a variety of fruit physical characteristics can be quickly and non-destructively ascertained. Due to grave worries about high levels of contamination, mostly from non-European fruit flies, five categories of Indian produce will no longer be available on the market.


FreshFruitPortal.com

EU bans Indian mangoes, eggplant

Five categories of produce from India will be banned in the European market following serious concerns about high levels of contamination, mainly from non-European fruit flies.

The ban, set to take place in May, will include mangoes, eggplants, the taro plant and two types of gourd, an EU release announced today.

The "emergency measures" come after a meeting among member states in Brussels just a few hours ago.

"The measures prohibit the import of some fruit and vegetables from India to tackle the significant shortcomings in the phytosanitary certification system of such products exported to the E.U.

"This decision had to be taken due to the high number of such consignments being intercepted at arrival in the E.U. with quarantine pests, mainly insects, like non-European fruit flies," the release said.

"Though the prohibited commodities represents less than 5% of the total fresh fruits and vegetables imported into the E.U. from India, the potential introduction of new pests could pose a threat to E.U. agriculture and production."

Pests were discovered in 207 consignments of fresh produce imported to Europe from India last year.

The ban will last until Dec. 31, 2015 when a review is scheduled to take place.


EU to ban contaminated fruits, vegetables from India

DEUTSCHE PRESSE-AGENTUR

The European Union member states on Wednesday decided to ban the import of five types of fruit and vegetable from India, after several batches were found to be contaminated by pests such as fruit flies, the bloc's executive said.

The prohibition, which goes into effect in May, covers mangoes, aubergines, the taro plant and two types of gourd. These represent less than 5 per cent of the bloc's fresh fruit and vegetable imports from India, according to the European Commission.

The ban is to be reviewed by the end of 2015.

European market. Mangoes, eggplants, taro plants, and two varieties of gourds will all be prohibited as of May 1st, according to an announcement from the EU. The restrictions aim to address the serious deficiencies in the photo sanitary certification system of such items shipped to the European Union by forbidding the import of certain fruits and vegetables from India.

The rapid advancement of information science, image processing, and pattern recognition technologies has led to the evolution of optical sensing technologies as scientific tools for nondestructive fruit quality assessment. By integrating spectroscopy and traditional imaging methods to extract spatial and spectral data from the target, spectral imaging technology is utilized to assess specific food products. The considerable research and development of hyper spectral imaging, which combines spectroscopy and imaging techniques to create a system that can obtain a spatial map of spectral variation, has led to several successful applications in the quality assessment of fruits.

II. Related Work:

1. Chen, I. T., & Lin, H. Y, presented an image-based approach for the yield estimation of cherry tomatoes. The objective is to assist farmers to quickly evaluate the amount of mature tomatoes which are ready to harvest. The proposed technique consists of machine learning based methods for detection, counting, and maturity assessment using multi-spectral images.
2. Blasco, J., Munera, S., Aleixos, N., Cubero, S., & Molto, E. given, the agricultural commodity's component parts vary in terms of size, color, and shape. Moreover, because they are living things, the process of developing precise automatic inspection instruments is complicated by the fact that their properties vary over time.
3. Li, X., Li, R., Wang, M., Liu, Y., Zhang, B., & Zhou, J. offered a technique The hyperspectral imaging approach may capture a sequence of monochromatic images at nearly continuous wavelengths, combining spectroscopic and imaging methods into a single device. Numerous studies that use spectral and/or spatial image processing and analysis have been published, and they suggest using hyper spectral imaging technology to evaluate the quality of fruits and vegetables.
4. Johansen, K., Raharjo, T., & McCabe, M. F. presented unmanned aerial vehicles(UAVs) offer a never-before-seen ability to track the dynamics and development of tree growth and structure across time. Pruning tree crops is widely believed to promote new growth, improve fruiting, facilitate fruit picking, and maybe enhance production by increasing light interception and surface area of the tree crown.
5. Saputro, A. H., & Handayani, W. proposed a method to find the optimal wave length selection in predicting he quality of the banana fruit (Musasp.) was presented. Reduction of the dimension of wavelength was performed in two stages.
6. Islam, M. N., Nielsen, G., Stærke, S., Kjær, A., Jørgensen, B., & Edelenbos, M. created to evaluate the effectiveness of four distinct non-destructive onion quality assessment

techniques; three of these techniques relied on spectral imaging, and one was based on near-infrared spectroscopy. By taking measurements of the onion surface in situ, these techniques use a range of wavelengths, from visible to near-infrared, and various acquisition devices to distinguish between pre-sorted onions.

III. Proposed Methodology:

The goal of the current multidisciplinary research project is to create auniquenon-destructive quality control method for fruits by combining spectral analysis of internal quality attributes with analysis of the fruit photos' external appearance attributes.

Objectives:

To develop Digital Image Processing based algorithms to characterize the external quality attributes appearance characteristics.

- To develop spectral analysis-based algorithms to characterize the internal quality attributes of the fruit.
- To develop a sound knowledge base and employ machine vision techniques to develop Automated Fruit Grading System.
- To develop a comprehensive model for controlling fruit quality and other horticulture crops after harvest.

3.1 Digital Image Processing Techniques for Quality Assessment of External Appearance of Fruits:

Regarding fruit recognition, classifying fruits and vegetables, classifying photographs, and identifying fruit diseases. Classifying fruits and identifying fruit diseases are examples of picture categorization. The majority of research on fruit identification and fruit disease detection used texture and color characteristics to classify fruits. The majority of research on fruit recognition has focused on fruits that are found on trees; nevertheless, we have limited our evaluation to methods that have been suggested for classifying various fruit and vegetable varieties. The majority of fruit disease detection research conducted in the literature is limited to the identification of a specific disease kind.

Issues and Challenges:

Fruits of multiple varieties may be seen in the input photographs in any quantity and at any position. Numerous fruit varieties display notable differences in form, consistency, and hue contingent upon their level of maturity. Oranges, for instance, can be spotty and brown, green, or yellow. Extracting and combining the elements that are helpful for identifying the produce is required because classifying fruits based on a single visual feature might not be enough.

Fruit Recognition and Classification:

Precise defect segmentation is required in fruit disease classification as shown in figure 3 and 4.

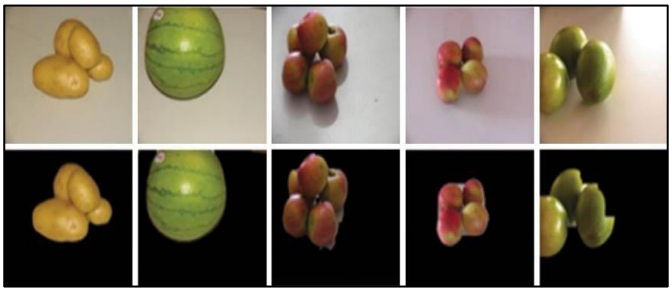


Figure 3: Some Image Segmentation Result on the Fruit Images

If the approach is not exact, the features of the non-infected zone will dominate the features of the infected region. Automatic fruit disease detection is necessary to detect fruit infections as soon as symptoms appear on developing fruits. Fruit diseases can manifest themselves after harvest, resulting in substantial declines in yield and quality.

Figure 4: Some Defect Segmentation Results on the Diseased Fruit

Multispectral imaging-based techniques for internal quality assessment of fruits

Hyperspectral imaging is becoming a more valuable inspection method in the food and agriculture sectors. A hyperspectral imaging system converts a two-dimensional (2-D) spatial matrix of vectors, each of which represents a spectrum, typically in the wavelength range of VIS to NIR, into a three-dimensional (3-D) picture dataset known as a hypercube.

Image Acquisition Modes:

The capture of hyperspectral image data can be done in four general ways: point, line, area, and single shot scanning, as shown in Figure 5. Point scanning moves the item or the detector continually in both spatial dimensions to produce

3-D image cubes while continuously capturing spectral data for each pixel.

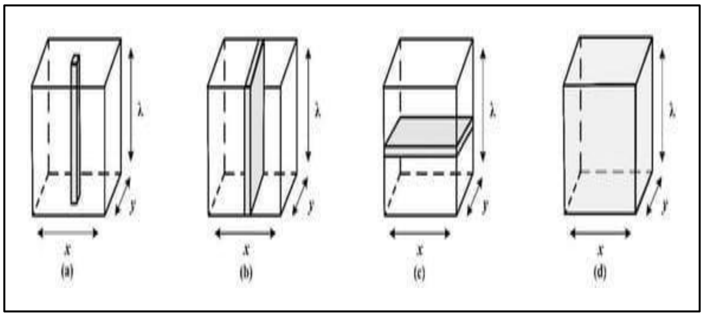


Figure 5: There are four methods for obtaining a hyperspectral picture cube in three dimensions: The four types of scanning are (a) point, (b) line, (c) area, and (d) single shot. The spatial dimensions are represented by x and y, and the wavelength is denoted by.

Sensing Modes:

In hyperspectral imaging, three commonly used sensing modes are depicted in Figure 6: interactance, transmittance, and reflectance. The primary distinction between these sensing modes is how they illuminate the sample during imaging, and each has unique consequences for determining food safety and quality.

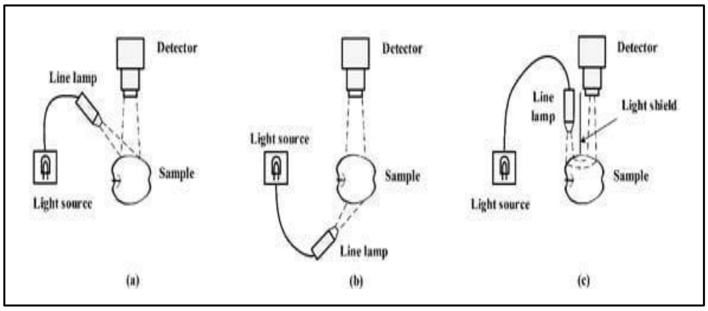


Figure 6: Three sensing modes—reflectance, transmittance, and interactance—are frequently employed to acquire hyperspectral images.

Over the past ten years, hyperspectral imaging has become more and more popular as a nondestructive method of assessing fruit quality in science. By integrating spectroscopic and imaging techniques into a single apparatus, the hyper spectral imaging approach has the potential to record a series of monochromatic images at almost constant wavelengths.

Its characteristics include ripeness, size, weight, form, color, condition, and the existence or lack of defects, stems, or seeds. There are also some internal qualities that are present, including texture, hardness, sweetness, and acidity.

A method of evaluation that is accurate, fast, and unbiased is essential to ensure the quality of fruits and vegetables during processing. Thanks to the quick development of information science, image processing, and pattern recognition technologies, optical sensing technologies have developed into scientific tools for nondestructive fruit quality assessment. Spectral imaging technology is used to evaluate particular food products by combining spectroscopy and conventional imaging techniques to obtain spatial and spectral information from the target. Many fruit quality assessments have benefited from the extensive study and development of hyper spectral imaging, which creates a spatial map of spectral variation by combining spectroscopy and imaging approaches.

The method of taking and analyzing an image at a very wide range of wavelengths is known as hyperspectral imaging. Hyperspectral imaging analyses an image in tens or hundreds of colors, as opposed to multispectral imaging's three or four colors (red, green, blue, and near-infrared (NIR), for example). Similar to other spectral imaging techniques, hyperspectral imaging gathers and analyses data from the whole electromagnetic spectrum. Finding the spectrum for each pixel in an image of a scene is the aim of hyperspectral imaging, which can be used to locate items, identify materials, or spot processes, as illustrated in figure 7.

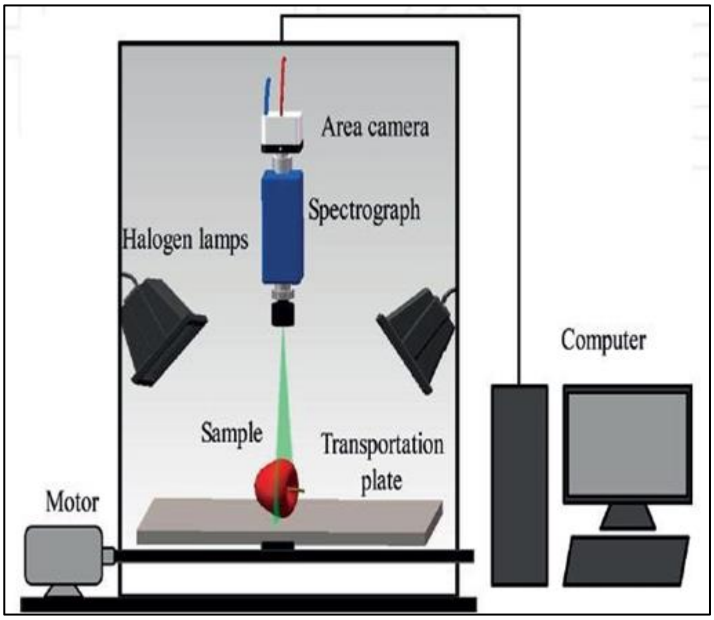


Figure 7: A schematic of the hyperspectral imaging system

The way multispectral cameras operate is by taking pictures at various light wavelengths. After that, these image sets are stitched together to produce mosaics that are accurate to geography and have several layers for every wavelength.

Vegetation indices are produced by adding these layers mathematically.

The surface examination of fruits and vegetables without the use of any invasive methods that compromise the food's quality is known as non-destructive analysis. The non-destructive evaluation techniques provide information on the structure, mechanical, physical, and chemical aspects of food.

3.2 Machine Learning Algorithms for Post-Harvest Quality Control of Fruits:

The application of computer vision and image processing technologies to the food sector and agriculture. Size, color, form, texture, and fault are the most crucial aspects of agricultural products' quality. Computer vision systems, which offer genuine, fair, and non-destructive grading, are utilized to replace manual food inspection. The computer vision-based quality inspection consists of four main processes: acquisition, segmentation, feature extraction, and classification.

Image Acquisition: Among the image acquisition techniques used in food applications are cameras, computed tomography, electrical tomography, magnetic resonance imaging (MRI), and ultrasonography (CT). Charged coupled devices (CCDs) and complementary metal oxide semiconductors (CMOS) image sensors are used to create digital images.

Preprocessing: Multiple noises present in images captured using different procedures degrade the quality of the image. As such, it is unable to provide relevant data for picture processing.

Segmentation: After preprocessing, a digital image must be segmented, or divided into several sections. The primary goal is to remove the background so that the object's significant area may be assessed. For image analysis to advance, good segmentation is essential, and incorrect segmentation will reduce the performance of the classifiers. It is employed in many different fields, including medicine and agriculture. Two popular segmentation methods are clustering and thresholding.

Feature Extraction: Following picture segmentation, features are extracted and estimated for additional analysis. These characteristics, which include useful data for object classification, interpretation, and picture perception, are the fundamental components of a computer vision system. The extracted features in this procedure create feature vectors, which are then classified to identify the input. These feature vectors perfectly and uniquely define the shape of the object. By extracting the features, feature extraction seeks to increase recognition rate. These features provide the explicit data that can be taken into account for quality analysis and assessment in the food business. Fruit defects and maturity are often assessed using morphological, textural, and color characteristics.

Classification: Because classification offers a framework for the artificial imitation of human brain processes that allow individuals to make difficult decisions fast, accurately, and consistently, it is an essential part of the evaluation of food quality. Using image processing techniques, images of fruits and vegetables can be categorized based on a variety of attributes, such as color, size, shape, and texture. These characteristics are utilized to create a training set. A classification method is then employed to extract knowledge that helps decide in cases that are unknown. For the purpose of classifying food in computer vision systems, numerous methods have been developed, such as KNN, SVM, ANN, and CNN (Deep Learning/Convolutional Neural Network).

3.3 Defect Detection of Fruits and Vegetables:

Commercial sorting equipment historically used classic computer vision to analyses the color, texture, size, and shape of fruits visually. Due to the wide variety of defect types, defect detection is still an impossible task. Fruit illness must be automatically detected as soon as it manifests itself. Fruit disease results in yield and quality losses that manifest at harvest. Apple fruits are frequently affected by scab, rot, and blotch diseases. The demand for genuine and impartial equality determination in food items is always growing due to rising standards for excellent quality in food goods. Computer vision systems offer an automated, economical, and non-destructive method for managing these requirements. There are various applications for this image-processing based inspection method when determining the quality of fruits. Because of their outstanding performance, continuously rising costs, usability, and reliable algorithms, image processing and computer vision systems are scientific processes employed in industrial and agricultural settings. These days, a variety of conventional, hyperspectral, and multispectral computer vision technologies are used to evaluate fruit and vegetable quality. A traditional computer vision system (TCVS) inspects common properties such as color, size, shape, texture, and defect. Multispectral and hyperspectral computer vision systems offer dynamic tools to address some flaws that are difficult to detect with TCVS because of the dominance of spectral images. These systems can improve TCVS. Defects can be more accurately detected by overcoming various challenges such as uneven light distribution on the arch surface, strong wavelength selection for diverse applications, stem/calyx recognition, surface assessment, long acquisition and processing times for spectral images, discriminating between different defects, etc. To get better results, other factors should be checked in addition to specific features like color, size, form, and texture while grading. Additionally, by adjusting the weight value, the performance may be enhanced even if all features had the same weight. The quality examination of fruits can benefit from the significant advancements in Terahertz imaging, Raman imaging, and 3D method.

IV. Conclusion and Expected Outcome:

The experimental findings demonstrated the computational efficiency, good extraction rate, and low false alarm rate of the suggested approach. The suggested method outperforms the current one in terms of accuracy of retrieved text boundaries and misses less text information. The system is robust and intolerant to noise, variations in font size and style orientation, complicated backgrounds, and low contrast images, according to observations made with the region growth approach. A 90.94% text extraction accuracy is attained.

The technique can be expanded in the future to eliminate visual blur, high intensity, and shadows. additionally, quicken the process of locating the area and removing text from photos. Word extraction is another possible extension of the approach.

References:

1. Chen, I. T., & Lin, H. Y. (2020). Detection, Counting and Maturity Assessment of Cherry Tomatoes using Multi-Spectral Images and Machine Learning Techniques. In VISIGRAPP (5: VISAPP) (pp. 759-766).

2. Blasco, J., Munera, S., Aleixos, N., Cubero, S., & Molto, E. (2017). Machine vision-based measurement systems for fruit and vegetable quality control in postharvest. Measurement, modeling and automation in advanced food processing, 71-91.
3. Li, X., Li, R., Wang, M., Liu, Y., Zhang, B., & Zhou, J. (2017). Hyperspectral imaging and their applications in the nondestructive quality assessment of fruits and vegetables. In Hyperspectral imaging in agriculture, food and environment. Intech Open.
4. Johansen, K., Raharjo, T., & McCabe, M. F. (2018). Using multi-spectral UAV imagery to extract tree crop structural properties and assess pruning effects. Remote Sensing, 10(6), 854.
5. Saputro, A. H., & Handayani, W. (2017, October). Wavelength selection in hyperspectral imaging for prediction banana fruit quality. In 2017 International Conference on Electrical Engineering and Informatics (ICELTICs) (pp. 226-230). IEEE.
6. Islam, M. N., Nielsen, G., Stærke, S., Kjær, A., Jørgensen, B., & Edelenbos, M. (2018). Novel non-destructive quality assessment techniques of onion bulbs: a comparative study. Journal of food science and technology, 55(8), 3314-3324.
7. Lu, Y., Huang, Y., & Lu, R. (2017). Innovative hyperspectral imaging-based techniques for quality evaluation of fruits and vegetables: A review. Applied Sciences, 7(2), 189.
8. Li, Xiaona, et al. "Hyperspectral imaging and their applications in the nondestructive quality assessment of fruits and vegetables." Hyperspectral imaging in agriculture, food and environment (2018): 27-63.
9. Elmasry, Gamal, et al. "Principles and applications of hyperspectral imaging in quality evaluation of agro-food products: are view. "Critical reviews in food science and nutrition 52.11 (2012): 999-1023.
10. Huang, Hui, Li Liu, and Michael O. Ngadi. "Recent developments in hyperspectral imaging for assessment of food quality and safety." Sensors14.4 (2014): 7248-7276.
11. Wu, Di, and Da-WenSun. "Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review—Part I: Fundamentals. "Innovative Food Science & Emerging Technologies 19 (2013): 1-14.
12. Wu, Di, and Da-WenSun. "Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review—Part II: Applications." Innovative Food Science & Emerging Technologies 19 (2013): 15-28.
13. Lu, Yuzhen, et al. "Hyperspectral imaging technology for quality and safety evaluation of horticultural products: A review and celebration of the past 20-year progress." Postharvest Biology and Technology 170 (2020): 111318.
14. Gowen, Aoife A., et al. "Hyperspectral imaging–an emerging process analytical tool for food quality and safety control." Trends in food science & technology 18.12 (2007): 590-598.

ISSN: 2582-8118 *Volume 4, Issue 5; Sept 2024*

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYSIS IN SCIENCE AND ENGINEERING

Web: https://www.iarj.in/index.php/ijrase/index

93. Use of AI Assistance in Online Proctored Exams

Nitiraj Kulkarni
Department of Artificial Intelligence,
Vishwakarma University, Pune, India.

Mahadev Biradar
Department of Mathematics,
Basaveshwar Engineering College, Bagalkote, India.

Jagadish V. Tawade
Department of Mathematics and Statistics,
Vishwakarma University, Pune, India.

Bhupender Singh
Department of Psychology,
Vishwakarma University, Pune, India.

Abstract:

Online proctored examinations have transformed the landscape of assessment, provided convenience and flexibility while simultaneously prompted concerns regarding integrity and security. This study delves into the susceptibilities inherent in these systems, specifically honing in on potential breaches by examinees. Through a thorough examination of the elements comprising online proctoring - encompassing client interfaces, provider platforms, and server infrastructure - we identify loopholes and vulnerabilities that could be exploited. Furthermore, we explain a systematic approach students use to partake in academic dishonesty, utilizing tools such as virtual machines and counterfeit camera feeds. To combat these issues, we suggest practical remedies, such as the creation of Python scripts for monitoring video feeds for abnormalities and enhancements to website coding to deter exploitation. This research underscores the necessity for robust security protocols and the incorporation of Artificial Intelligence to protect the integrity of online proctored exams in the modern digital age.

Keywords:

Online Proctored Exams, Vulnerabilities, Academic Misconduct, Security Measures, Artificial Intelligence, ChatGPT.

1 Introduction:

As the world progresses, the methods of conducting exams are also evolving. In India, many exams are transitioning from the traditional pen-and-paper format to online proctored systems. However, with this shift comes a set of challenges. Since this technology is relatively new, students may not be fully aware of the potential loopholes and methods to circumvent the system. Exam providers must recognize these vulnerabilities and take proactive steps to address them, thereby enhancing the security and effectiveness of the system.

A software for proctored exams, namely Proctor U, was tested on students. They reported that 88.95% of students were satisfied with using the software. Those who were unsatisfied reported three main difficulties, mainly personal and technical. In the software, they used cameras to detect unauthorized materials present during the exam. Over half of the students indicated that the use of Proctor U would influence their future decision to take another online course, either negatively or positively.

A study conducted by Harmon et al. revealed statistics of the Grade Point Average (GPA) of several students in different exam scenarios. It was observed that the GPA of students in online proctored exams was marginally higher than in non-proctored exams raising concerns about the ethical effects of online exams.

To curb academic malpractices Samuel et al. conducted a series of online tests with randomized question order. Feedback given by the students clearly indicated that the randomized order was effective in stopping academic malpractices. With the randomized question order students weren’t able to copy from each other thus effectively reducing the number of copied answers.

Nigam et al. highlighted significant worries that institutions need to address regarding AI-based proctoring systems, including security, privacy, and ethics. They emphasized the importance of having a reliable, secure, and user-friendly proctoring system to maintain academic integrity and uphold strict security standards.

Hosseini et al. recommended students install proctoring in advance such that students can configure all the required parameters in the software beforehand. But doing this would pose a risk as currently the introduction of Artificial Intelligence (AI) in proctoring adds another layer of complexity. Upon learning that AI is involved in monitoring exams, students may experience heightened anxiety and perceive the exams with greater seriousness. However, it's important to note that the AI used in these systems may not always be highly advanced and may not effectively detect cheating attempts. Therefore, there is a need to thoroughly examine the shortcomings of online proctored exams and propose practical solutions to mitigate them.

This paper aims to delve into the various flaws present in online proctored exams and offer feasible remedies. By identifying these weaknesses and proposing effective solutions, we strive to contribute to the improvement of online assessment systems, ensuring fairness and integrity in the evaluation process.

2 Understanding Online Proctored Exams:

Online proctored exams represent a novel approach to evaluating students' academic performance, allowing educators to administer assessments remotely. While this method offers convenience in many respects, it also introduces certain risks that can potentially undermine students' overall performance. This section provides an overview of the processes involved in online proctored exams, shedding light on the intricate mechanisms employed in their administration.

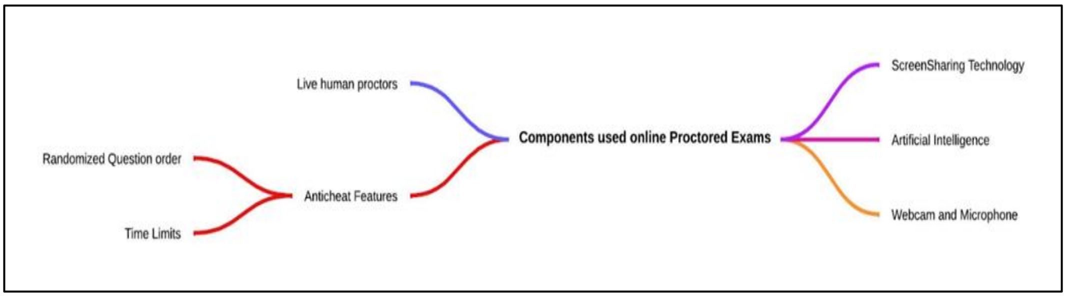


Figure 1: Components used in online Proctored Exams.

1. **Client/Exam Taker's Side:** The Client/Exam Taker's Side of online proctored exams encompasses a multitude of elements crucial for facilitating a seamless testing experience. Essentially, this aspect revolves around utilizing the audio and video capabilities of the client's device to ensure comprehensive monitoring during the examination process. These functionalities serve as essential tools for proctors to maintain the integrity and security of the exam environment.

Fundamentally, one pivotal component of the Client/Exam Taker's Side lies in the setup of audio and video. Typically, students are required to activate their device's microphone and webcam, enabling proctors to monitor their activities in real time throughout the duration of the exam. This setup facilitates continuous observation, ensuring that students adhere to the established guidelines and regulations. Additionally, some exams may necessitate screen-sharing access, allowing proctors to monitor the test-taking environment more effectively. It should be noted that while the screen-sharing feature offers a valuable mechanism for live monitoring, it generally permits observation of only one screen at a time, thus limiting the scope of surveillance.

2. **Provider's Side:** The Provider's Side of online proctored exams is a vital cog in the wheel, comprising the digital infrastructure and mechanisms wielded by exam hosting platforms to fortify the integrity and security of the assessment process. These platforms are armed with sophisticated technologies and protocols aimed at thwarting cheating attempts and preserving fairness. At the heart of the Provider's Side lie the online platforms or websites where exams are hosted and administered. These platforms employ advanced coding and security measures, such as JavaScript, to enforce stringent examination conditions. For instance, JavaScript can compel users to enter full-screen mode during exams, curtailing their access to other tabs or applications.

Moreover, any attempt to exit the exam interface triggers system notifications, potentially leading to disciplinary actions if warranted. The Provider's Side encompasses backend functionalities essential for conducting the exam process seamlessly. This encompasses exam delivery systems, data processing algorithms, and communication protocols, all working harmoniously to ensure the smooth execution of exams from commencement to conclusion. Furthermore, online proctoring platforms implement robust encryption techniques to safeguard the confidentiality and integrity of exam content. By encrypting exam data during transmission and storage, providers mitigate the risk of unauthorized access or tampering, thereby ensuring the security of exam questions and content.

3. **Server/Intermediary Side:** The Server/Intermediary Side plays an important role in the online proctored exam setup, serving as the centre that coordinates every aspect of the assessment process. It's like the brain behind the operation, managing a complex network of code and systems to ensure exams run smoothly and securely. Its responsibilities include managing the distribution of exam materials, maintaining the integrity of the exam environment, and overseeing the execution of security protocols.

3 Vulnerabilities in Online Proctored Exams:

Online proctored exams heavily rely on audio and video feeds to monitor and assess test-takers' behaviour during examinations. However, these essential components are susceptible to manipulation, presenting significant challenges to maintaining exam integrity. This section explores the step-by-step process involved in setting up audio and video manipulation tools, shedding light on the intricate techniques utilized to circumvent proctoring measures and gain unfair advantages in examinations.

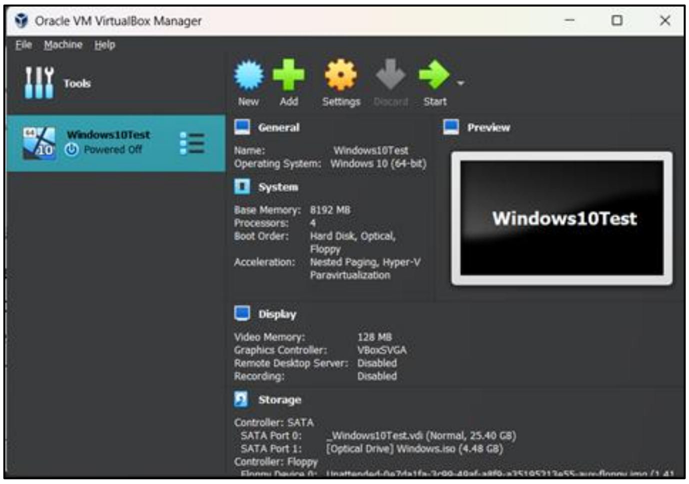


Figure 2: Windows Installed into Oracle's VM Virtual Box

Step 1 Downloading Virtual Machine Provider: The first step in audio and video manipulation involves acquiring a virtual machine provider. Among the various options available, Oracle's VM Virtual Box stands out as a well-known and user-friendly virtual machine simulation software.

By downloading and installing Virtual Box, individuals gain access to a simulated environment where they can run guest operating systems, thus creating a conducive platform for executing manipulation tactics.

Step 2 Installing Windows Operating System: With Virtual Box set up, the next step is to install a Windows operating system within the virtual environment. This entails procuring a licensed copy of Windows and configuring it to run seamlessly within Virtual Box. By emulating a Windows environment, individuals can exploit software compatibility and leverage Windows-exclusive tools for audio and video manipulation test benefits Online proctored tests have transformed assessment, benefiting institutions and students. This section highlights the many benefits of online proctored tests for education.

Step 3 Setting up a fake Virtual Camera: Once the virtual environment is established, the focus shifts to bypassing camera restrictions imposed by proctoring systems. OBS Studio emerges as a software renowned for its versatility and functionality in live streaming. By utilizing OBS Studio, individuals can create a fake camera feed, thus deceiving proctoring systems into believing that the manipulated feed is a genuine representation of the examinee's surroundings.

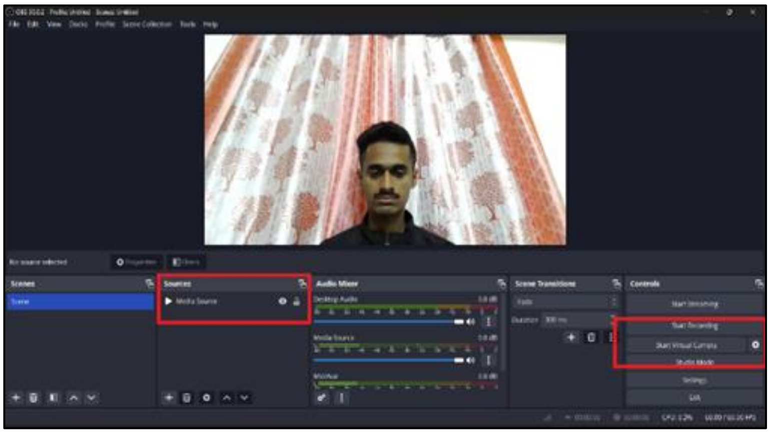


Figure 3: Setting up a virtual camera in OBS Studio

Step 4 Preparing Pre-recorded Video: To further enhance the deception, individuals are advised to pre-record a video where they stare at the screen for 10-15 seconds, mimicking the behaviour of a genuine examinee engrossed in the examination. Careful attention should be paid to minimizing eye movement and maintaining a semblance of authenticity. Once recorded, the video is loaded into OBS Studio, where it serves as the basis for generating the fake camera feed.

Step 5 Live Comparison and Deception Tactics: The culmination of these efforts is a live comparison between the fake camera feed generated by OBS Studio and the genuine webcam feed. As depicted in the comparison, the manipulated feed may exhibit subtle discrepancies such as graininess and lag, simulating the characteristics of a genuine laptop camera. This deceptive tactic enables individuals to trick proctoring systems into accepting the manipulated feed as authentic, thereby evading detection and scrutiny

Step 6 Exploiting AI for Questioning: With the audio and video manipulation infrastructure established, individuals can employ artificial intelligence (AI) to present questions during the examination. As exam takers may divert their attention from the exam window, they can freely utilize their home PC without arousing suspicion. Students have access to numerous AI chatbots online to seek answers to their queries. Despite the convenience offered by AI technologies, it's crucial to recognize their inherent limitations in terms of accuracy. Users should exercise caution when depending on AI for questioning, as inaccuracies could compromise the reliability of their responses.

3.2 Countermeasures and Solutions:

Observing the vulnerabilities identified on the client side, it becomes necessary to devise effective countermeasures and solutions to mitigate the risks posed by audio and video manipulation tactics. Proactive measures must be implemented to fortify the integrity of online proctored examinations and uphold academic standards. This section suggests comprehensive strategies and technological interventions aimed at addressing these challenges and safeguarding the integrity of the examination process

a) **Advanced AI Proctoring:** Advanced AI proctoring represents a cutting-edge approach to ensuring the integrity of online examinations by leveraging artificial intelligence (AI) algorithms for real-time monitoring and analysis. By harnessing the power of machine learning and computer vision, AI proctoring solutions can detect and deter cheating behaviours with unprecedented accuracy and efficiency. We have developed an advanced AI detector specifically designed to enhance the integrity of proctored exams by detecting any repetitions in the video feed. Unlike current methods, which rely on manual verification or random AI checks through random snapshots, this system offers a continuous and proactive approach to monitoring. By connecting directly to the live video feed, our AI detector conducts real-time analysis to identify repeated eye and mouth movements, providing a more comprehensive and reliable means of ensuring academic integrity during exams. This solution aims to address the limitations of existing proctoring methods and uphold the standards of fairness and accountability in academic assessments.

b) **Enhancing JavaScript and Backend Security:** To bolster the security of online proctoring systems, it is imperative to reinforce JavaScript and backend functionalities. This involves implementing robust measures to prevent unauthorized activities and ensure the integrity of exam processes. By strengthening the security of both the frontend and backend components, educational institutions can effectively mitigate the risk of cheating and maintain the credibility of online assessments. One key strategy is to enhance JavaScript code to impose strict restrictions on user interactions during exams. This includes measures such as prohibiting the opening of new browser tabs or windows, preventing the use of keyboard shortcuts to navigate away from the exam interface, and disabling right-click functionality to prevent copying and pasting of exam content. JavaScript can be leveraged to force the browser into full-screen mode during the duration of the exam, thereby minimizing distractions and reducing the likelihood of external interference. Backend security measures play a crucial role in safeguarding exam integrity. By implementing robust authentication and authorization protocols, educational institutions can ensure that only authorized users have access to exam materials and resources. This may involve implementing multi-factor authentication mechanisms, encrypting sensitive data transmitted between the client and server, and implementing access controls to prevent unauthorized users from tampering with exam settings or content.

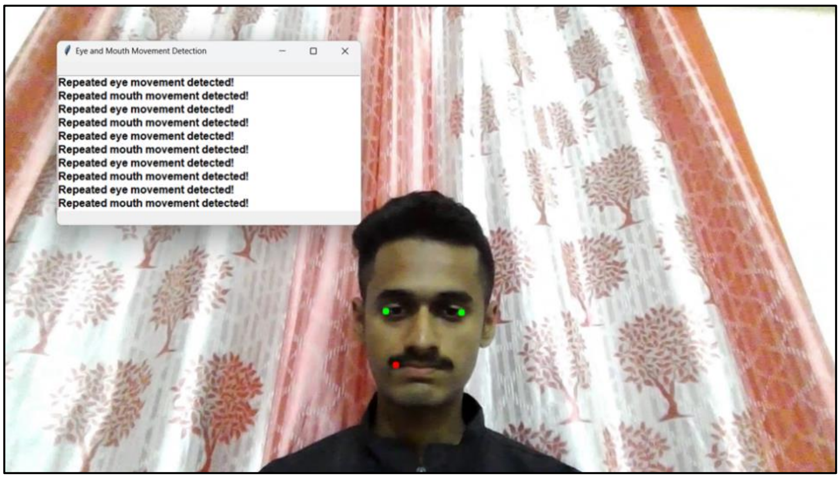


Figure 4: Detection of Repeated Frames in Pre-Recorded Video using Python.

4. Conclusion:

In conclusion, the evolution of online proctored exams has brought about both convenience and complexity in the assessment landscape. While these systems offer flexibility and accessibility, they also pose significant challenges in terms of maintaining integrity and security. Through a comprehensive examination of the vulnerabilities in online proctoring, this study has documented the potential exploits and weaknesses that threaten the credibility of academic assessments. From manipulating audio and video feeds to exploiting loopholes in client-side and backend security measures, test-takers have demonstrated a capacity to avoid existing safeguards and engage in academic misconduct. However, the identification of these vulnerabilities also presents an opportunity for proactive intervention and remediation. By proposing practical solutions such as advanced AI proctoring and enhancements to JavaScript and backend security, educational institutions can fortify their online assessment systems and uphold the standards of academic integrity. Advanced AI proctoring offers a sophisticated approach to real-time monitoring and analysis, leveraging machine learning algorithms to detect and deter cheating behaviours with unparalleled accuracy. Meanwhile, strengthening JavaScript and backend security measures ensures that exam processes are safeguarded against unauthorized activities and data breaches, thereby preserving the integrity of online assessments.

As we navigate the digital era of education, it is mandatory that we continuously innovate and adapt to address emerging challenges and threats to academic integrity. By embracing technological advancements and implementing robust security measures, we can create a more secure and trustworthy environment for online learning and assessment. Ultimately, the goal is to ensure fairness, equity, and credibility in the evaluation of students' academic performance, thereby upholding the integrity of educational institutions and the value of academic qualifications in the digital age.

References:

1. Milone, Anna S., et al. "The impact of proctored online exams on the educational experience." *Currents in Pharmacy Teaching and Learning* 9.1 (2017): 108-114.
2. Harmon, Oskar R., and James Lambrinos. "Are online exams an invitation to cheat?" *The Journal of Economic Education* 39.2 (2008): 116-125.
3. S. S. Chua, J. B. Bondad, Z. R. Lumapas, and J. D. L. Garcia, ‘‘Online examination system with cheating prevention using question bank randomination and tab locking,’’ in Proc. 4th Int. Conf. Inf. Technol. (InCIT), Oct. 2019, pp. 126–131.
4. Nigam, A., Pasricha, R., Singh, T. *et al.* A Systematic Review on AI-based Proctoring Systems: Past, Present and Future. *Educ Inf Technol* **26**, 6421–6445 (2021).
5. Mozes-Carmel, Abraham, and Stuart S. Gold. "A Comparison of Online vs. Proctored Final Exams in Online Classes." *Journal of Educational Technology* 6.1 (2009): 76-81.
6. Hosseini, M. M., Egodawatte, G., & Ruzgar, N. S. (2021). Online assessment in a business department during COVID-19: Challenges and practices. The International Journal of Management Education, 19(3).
7. Ardid, Miguel, et al. "Online exams for blended assessment. Study of different application methodologies." *Computers & Education* 81 (2015): 296-303.
8. Wahid, Rizwana, and Oveesa Farooq. "Online exams in the time of COVID-19: quality parameters." International Journal of Social Sciences & Educational Studies 7.4 (2020):
9. Kaur, Paramjit, et al. "Facial-recognition algorithms: A literature review." *Medicine, Science and the Law* 60.2 (2020): 131-139.
10. Nicola-Richmond, Kelli, Phillip Dawson, and Helen Partridge. "Online proctored exams: rhetoric vs reality." *Higher Education Research & Development* 43.2 (2024): 392-405.
11. Hamadah Alsadoon, Prof. "Using Artificial Intelligence for Proctoring Remote Exams: Advantages and Challenges." *International Journal for Research in Education* 48.1 (2024): 12-27.
12. Ngo, Dinh An, et al. "Examining Monitoring System: Detecting Abnormal Behavior In Online Examinations." *arXiv preprint arXiv:2402.12179* (2024).
13. Harmon, Oskar R., James Lambrinos, and Judy Buffolino. "Assessment design and cheating risk in online instruction." *Online Journal of Distance Learning Administration* 13.3 (2010): 23-33.
14. Vayre, Emilie, and Anne-Marie Vonthron. "Relational and psychological factors affecting exam participation and student achievement in online college courses." *The Internet and Higher Education* 43 (2019): 100671.
15. Sukmandhani, Arief Agus, and Indrajani Sutedja. "Face recognition method for online exams." *2019 International Conference on Information Management and Technology (ICIMTech)*. Vol. 1. IEEE, 2019.
16. Clark, Ted M., et al. "Testing in the time of COVID-19: A sudden transition to unproctored online exams." *Journal of chemical education* 97.9 (2020): 3413-3417.
17. M. Ghizlane, B. Hicham and F. H. Reda, "A New Model of Automatic and Continuous Online Exam Monitoring," *2019 International Conference on Systems of Collaboration Big Data, Internet of Things & Security (SysCoBIoTS)*, Casablanca, Morocco, 2019, pp. 1-5, doi: 10.1109/SysCoBIoTS48768.2019.9028027.
18. Cluskey Jr, G. R., Craig R. Ehlen, and Mitchell H. Raiborn. "Thwarting online exam cheating without proctor supervision." *Journal of Academic and Business Ethics* 4 (2011): 1.

19. Prathish, Swathi, and Kamal Bijlani. "An intelligent system for online exam monitoring." *2016 International conference on information science (ICIS)*. IEEE, 2016.
20. Frankl, Gabriele, and Sofie Bitter. "Online exams: practical implications and future directions." *Proceedings of the European Conference on e-Learning*. 2012.
21. Sarrayrih, Mohammad A., and Mohammed Ilyas. "Challenges of online exam, performances and problems for online university exam." *IJCSI International Journal of Computer Science Issues* 10.1 (2013): 1694-0784.
22. Li, Peng. "Selecting and using virtualization solutions: our experiences with VMware and VirtualBox." *Journal of Computing Sciences in Colleges* 25.3 (2010): 11-17.
23. Rizana, Dani, and Miftahul Huda. "Training on making learning videos using OBS studio." *Community Empowerment* 6.5 (2021): 815-821.
24. Tawade, J., & Kulkarni, N. (2024). Code: Python code for Repeated Face Movement Detection (Version 1). Zenodo. https://doi.org/10.5281/zenodo.10929793

ISSN: 2582-8118 *Volume 4, Issue 5; Sept 2024*

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYSIS IN SCIENCE AND ENGINEERING

Web: https://www.iarj.in/index.php/ijrase/index

Sponsors for SSET 2024

- **Platinum Sponsor (Rs. 50,000=00)**

Mr. Vijay S. Alur

Partner

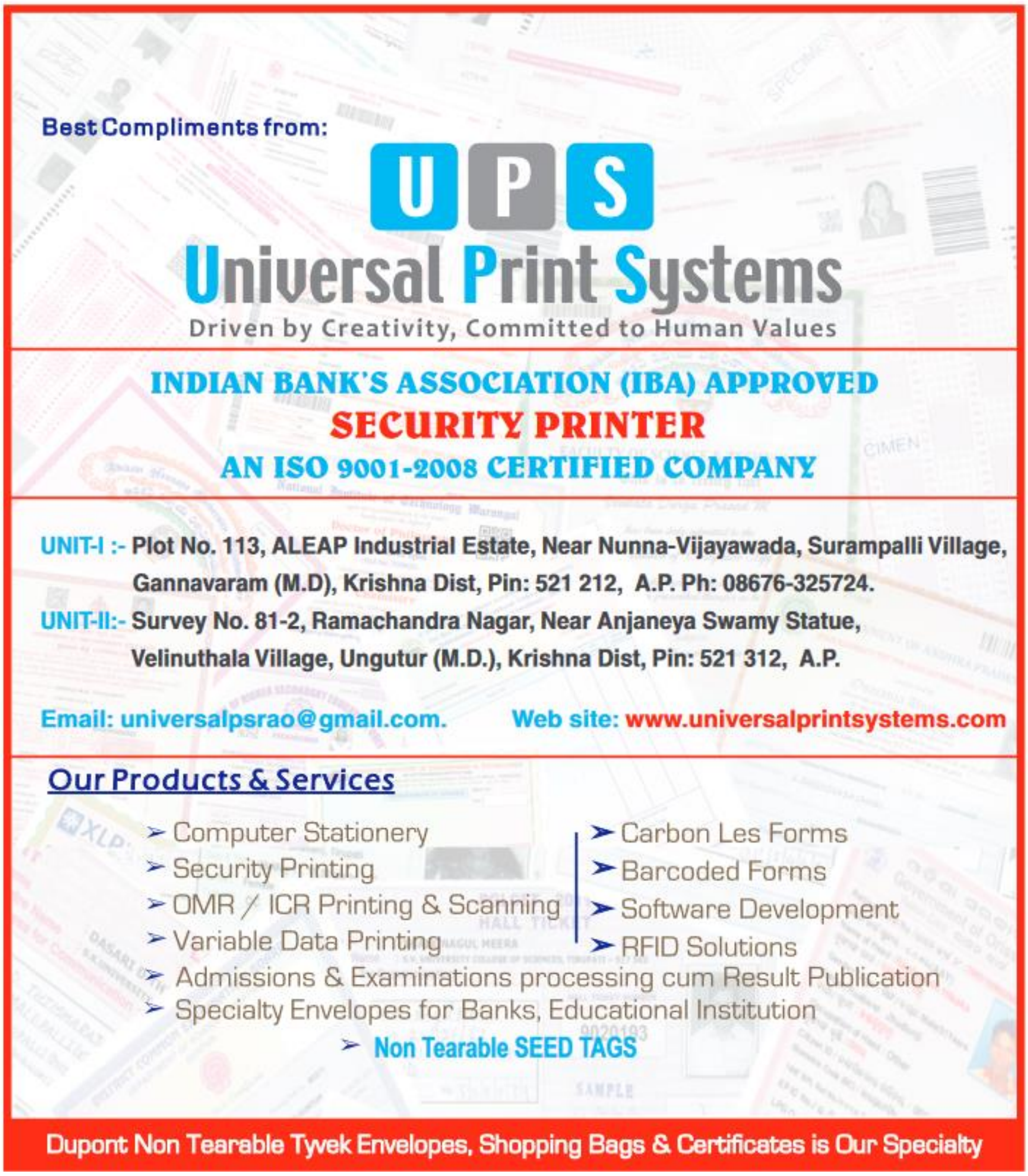
M/S S. S Alur Constructions, 2nd Floor, Vaishanvi Heights, Opposite to KSFC, Vivek Nagar East Vijayapur 586 109, Karnataka.

- **Gold Sponsor (Rs. 25,000=00)**

Universal Print Systems

Partner

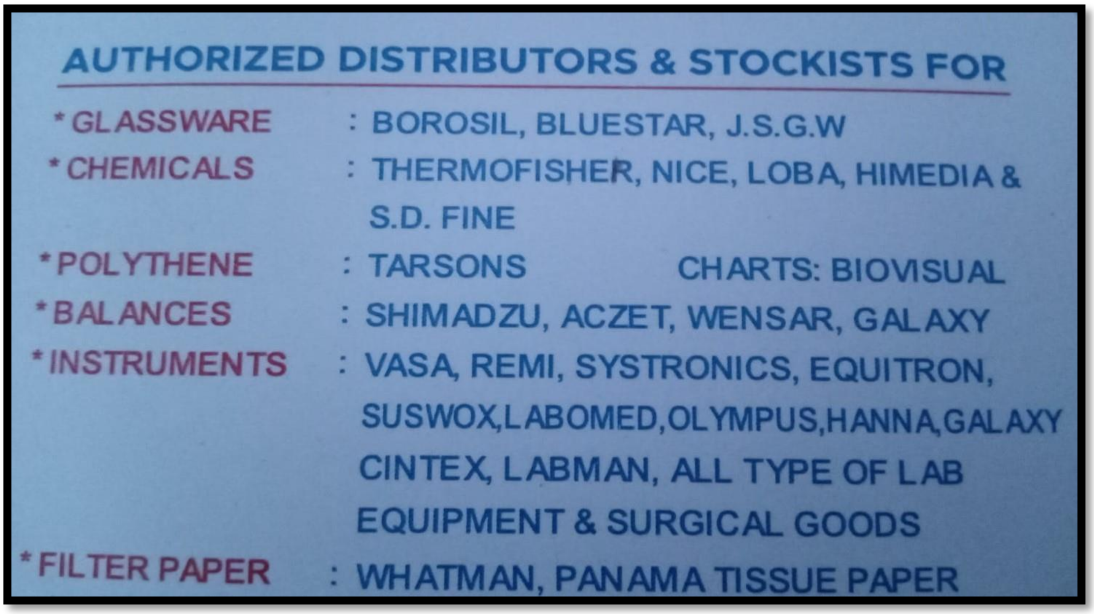
Plot No. 113, ALEAP Industrial Estate, Surampalli(V),
Gannavaram 521212, Krishna District, Andhra Pradesh.



➢ **Silver Sponsor (Rs. 10,000=00)**

Vasa Scienctic Co

95/1, 11th Cross, 4th Main, Malleshwaram
Bangaluru 560 003.



➢ **Silver Sponsor (Rs. 10,000=00)**

Kutub Jalihal Skillzam

1st Floor, Pearl Plaza, Plot #271, Shivabasava Nagar, Nehru Nagar,
Belagavi, Karnataka.

➢ **Silver Sponsor (Rs. 10,000=00)**

Eyesec Cyber Security

➢ **Silver Sponsor (Rs. 10,000=00)**

Angel Constructions

Bagalkote.

ABOUT THE CONFERENCE

The International Conference on Sustainable Solutions in Engineering and Technology (SSET-2024) is scheduled on August 29th & 30th, 2024 at Basaveshwar Engineering College, Bagalkote, Karnataka, India. The conference focuses on engineering and technological solutions for sustainable development to bring together professionals, researchers, policy makers and stake holders from various fields to discuss and address pressing global challenges. This conference serves as a platform for sharing knowledge, exchanging ideas, and fostering collaborations to develop innovative and practical solutions to promote sustainability across different sectors.

The aim of the conference is to identify sustainable engineering practices and to explore their implementation on a broader scale, considering economic, social, and environmental impacts. This international conference on sustainable solutions in engineering and technology promotes the transfer of sustainable technologies, methodologies, and practices from research institutions to industries fostering innovation and driving the adoption of sustainable solutions.

B.V. V. Sangha's
Basaveshwar Engineering College
Approved by AICTE, Affiliated to VTU, Belagavi, Accredited by NAAC with "A"
S. Nijalingappa Vidyanagar, Bagalkote 587102, Karnataka, India

KRIPA DRISHTI PUBLICATIONS

Kripa-Drishti Publications
A-503 Poorva Heights, Pashan-Sus Road, Near Sai Chowk,
Pune – 411021, Maharashtra, India.
Mob: +91 8007068686
Email: editor@kdpublications.in
Web: https://www.kdpublications.in