International Journal of Research in Machine Learning and Deep Learning

A Publication of Geethanjali College of Engineering and Technology

August 2024 Volume 1 Issue 1

Geethanjali College of Engineering and Technology
Cheeryal, Keesara(M), Medchal(Dist), Hyderabad, Telangana pin- 501301

ABOUT THE JOURNAL

START YEAR: 2024 **Frequency of Publication: Yearly**

Language: English **Publication Format: Print**

Publisher: Geethanjali College of Engineering and Technology

Cheeryal

Keesara(M), Medchal Dist.

Hyderabad

Telangana, INDIA

Pin Code-501301.

Scope of the Journal:

International Journal of Research in Machine Learning and Deep Learning is a peer reviewed, open access journal focused on research related to machine learning and Deep learning. The journal encompasses all aspects of research and development in ML,DL including but not limited to data mining, computer vision, natural language processing (NLP), …

The scope of the journal is research in the field of Artificial Intelligence machine learning in health care & science, artificial intelligence, and advances, artificial intelligence in blockchain technology, artificial intelligence in healthcare, business, artificial neural networks, big data algorithms, big data analysis, cloud computing, computer vision, computer vision and perception, cyber defense, cyber security, data mining, data mining with big data, data science, decision management, deep learning, fleet management in Artificial Intelligence automation machine learning, humanoid robotics, image processing, perception internet of things, machine learning, machine learning, and computing, machines, and minds, mechatronics & nonlinear control systems (nolcos), multimedia technologies, natural language processing (NIP), neural system, robotic process automation (RPA), robotics, robots in security and surveillance, speech recognition, virtual intelligence, etc.

We welcome articles from Professors, Clinicians, Scientists, Research Scholars, and Medical Practitioners. We will publish all types of articles like Research Articles, Reviews, Short Communication, Commentary, Image Articles, PPTs, Letters to the Editor, and Editorials.

International Journal of Research in Machine Learning and Deep Learning is open-access journal that provides an advanced forum for articles related to Machine Learning, Deep Learning& Artificial Intelligence. Following are some of the fields that the journal covers:

- A/B testing
- Accuracy
- Activation Function
- Advanced Algorithms
- Algorithm
- Anomaly Detection

- Applications of Computer Science
- Architecture Evaluations
- Artificial General Intelligence (AGI)
- Artificial intelligence
- Artificial Narrow Intelligence (ANI)
- Artificial Neural Network (ANN)
- Augmented Reality
- Automation and Mobile Robots
- Bayesian Optimization
- Bioengineering
- Bioinformatics
- Biometrics and Computer Forensics
- Brain Mapping
- Brain-inspired Intelligence
- Brain-machine Interface
- Chatbots
- Clipping
- Cloud Computing
- Clustering
- Cognitive computing
- Cognitive Modeling of Intelligence
- Computation and Language
- Computational Biology
- Computational Biology and Bio-informatics
- Computational linguistics
- Computational Statistics
- Computer Applications
- Computer Architecture
- Computer Science and Game Theory
- Computer Software / Hardware
- Computer Vision
- Computer Vision and Pattern Recognition

- Computers and Society
- Conversational Artificial Intelligence
- Convolutional Neural Network (CNN)
- Cryptography and Security
- Cyber-Physical Systems
- Data Engineering
- Data Structures and Algorithms
- Databases
- Databases and Data mining
- Deep Learning
- Digital Ecosystem
- Digital Libraries
- Distributed, Parallel, and Cluster Computing
- Document Analysis and Handwriting Recognition
- Expert System
- Generative Adversarial Networks (GAN)
- Graphics
- Heuristic
- Human-Computer Interaction
- Hybrid Artificial Intelligence
- Image and Video Analysis
- Image Processing
- Information Retrieval
- Information Theory
- Intelligent Robotics
- Internet Computing and Data Mining
- Internet of Things (IOT)
- Knowledge Representation and Reasoning
- Learning
- Logic in Computer Science
- Machine Learning
- Machine translation

- Mechatronics
- Multiagent Systems
- Multimedia
- Multimedia Analysis
- Natural Language Processing (NLP)
- Networking and Internet Architecture
- Neural and Evolutionary Computing
- Neural Networks
- Numerical Analysis
- Operating Systems
- Optical Character Recognition (OCR)
- Pattern Recognition
- Performance
- Post-quantum Cryptography
- Programming Languages
- Quantum Computing
- Recurrent Neural Network (RNN)
- Robotic Process Automation
- Robotics
- Security and Machine Learning
- Soft Computing
- Software Engineering
- Sound
- Speech and Language Processing
- Virtual Intelligence

CONTENTS **Page No.**

HOME AUTHORIZATION SYSTEM USING ML AND IOT

Dr. L.Venkateswarlu
(Professor & Hod)
Department of CSE-AIML,Geethanjali College of Engineering and Technology, Hyderabad, Telangana, India-501301
venkatlendale.cse@gcet,edu.in

Yokesh S
Department of CSE-AIML,Geethanjali College of Engineering and Technology, Hyderabad, Telangana, India-501301
yokesh1327@gmail.com

Sahith Ponna
Department of CSE-AIML,Geethanjali College of Engineering and Technology, Hyderabad, Telangana, India-501301
Sahith.ponna@gmail.com

M. Akanksha
Department of CSE-AIML,Geethanjali College of Engineering and Technology, Hyderabad, Telangana, India-501301
akankshamidivelli@gmail.com

***Abstract*— The integration of Machine Learning (ML) techniques with Internet of Things (IoT) technology has revolutionized home security systems, offering enhanced safety and peace of mind to homeowners. This paper presents a novel approach to home security by leveraging ML algorithms to analyze real-time data from IoT devices such as cameras, motion sensors, and door/window sensors. By applying ML models for anomaly detection, pattern recognition, and predictive analytics, the system can accurately identify and respond to potential security threats, including intrusion, burglary, and unauthorized access. Additionally, the incorporation of IoT enables remote monitoring and control of the security system through mobile applications or web interfaces, providing users with instant alerts and the ability to take proactive measures. Through a comprehensive evaluation, including simulation and real-world testing, our proposed ML-based IoT home security system demonstrates its effectiveness in safeguarding residential properties, minimizing false alarms, and optimizing resource utilization. This research contributes to advancing the field of smart home security by harnessing the synergy between ML and IoT technologies to create intelligent, proactive, and adaptable security solutions.**

Keywords - Machine Learning, Internet of Things, Home Security Systems, Anomaly Detection, Pattern Recognition, Predictive Analytics, Real-time Monitoring, Remote Control, Mobile Applications, Web Interfaces, Simulation, Real-world Testing.

I . INTRODUCTION

In today's rapidly evolving technological landscape, ensuring the safety and security of our homes has become a top priority [1][3][5]. Traditional security systems, while effective to a certain extent, often struggle to keep pace with the dynamic nature of modern threats [2][4][6]. To address these challenges effectively, many homeowners are turning to a combination of Machine Learning (ML) and Internet of Things (IoT) technologies, creating a smarter and more proactive approach to home security.

At the heart of this innovative approach is the fusion of ML algorithms with IoT devices, forming a dynamic system that can detect, analyze, and respond to potential threats in real time [1][7][10]. By harnessing the power of ML, these systems can learn from past incidents, adapt to changing environments, and differentiate between normal activities and suspicious behavior.

Central to any ML-based home security system is a network of strategically placed IoT devices, such as motion sensors, door/window sensors, and surveillance cameras [1][8][11]. These devices work together to collect and transmit data about the home environment, creating an interconnected network that facilitates seamless communication and information exchange.

The data collected by these sensors serves as the foundation for ML algorithms to analyze patterns, identify anomalies, and detect potential security risks [1][9][13]. For example, by analyzing motion patterns and camera footage, the system can distinguish between routine activities and suspicious behavior that may indicate unauthorized access.

Moreover, ML algorithms can be personalized to adapt to the unique characteristics of each home and its residents [1][12][15]. This personalized approach ensures a tailored security protocol that meets specific needs and preferences, reducing false alarms and providing homeowners with peace of mind.

In addition to threat detection, ML-powered home security systems can take automated actions in response to security breaches [1][14][17]. Integrated IoT devices like smart locks, lights, and alarms enable the system to respond quickly and effectively, mitigating risks and enhancing overall security.

A key component of these systems is an intuitive user interface that allows homeowners to monitor security status, access live camera feeds, and receive real-time alerts [1][16][19]. Whether through mobile apps, web interfaces, or voice-activated assistants, user-friendly interfaces empower homeowners to take proactive measures and maintain a secure environment for their families.

As ML algorithms and IoT technologies continue to advance, home security systems remain at the forefront of innovation [1][18][20], offering enhanced performance, adaptability, and resilience against emerging threats. Regular updates and refinements based on new data and insights ensure that these systems remain effective in safeguarding homes and providing homeowners with peace of mind.

II . LITERATURE SURVEY

The literature surrounding smart home security systems integrating Machine Learning (ML) and Internet of Things (IoT) technologies offers valuable insights into the advancements and challenges in this field. Several key studies have contributed significantly to understanding the potential and effectiveness of ML-based home security solutions.

Chaudhary et al. [1] presented an IoT and ML-based smart home security system, highlighting the integration of ML algorithms for enhanced security analysis. Gopalakrishnan et al. [2] conducted a comprehensive review of smart home security systems using IoT and ML, emphasizing the importance of proactive security measures.

Al-Fuqaha et al. [3] discussed recent advances and challenges in IoT for home automation, shedding light on the technological landscape of IoT-enabled security solutions. Abbasi and Munir [4] delved into the role of ML in smart homes, emphasizing its potential in enhancing security protocols.

Rahmani et al. [5] conducted a survey on smart home security solutions, addressing architecture, technologies, and challenges, providing a broad overview of the field. Singh et al. [6] explored the integration of IoT, ML, and blockchain in smart home security, showcasing innovative approaches to enhancing security systems.

Hassan et al. [7] provided a comprehensive review of the present state and challenges in smart home security, offering valuable insights into the current landscape. Aziz and Heena [8] discussed smart home security systems using IoT and ML, emphasizing practical implementations and benefits.

Banerjee and Hettiarachchi [9] conducted a review on smart home security, focusing on concepts, solutions, and challenges, providing a comprehensive understanding of the domain. Trappey et al. [10] presented an IoT-based home automation system, emphasizing energy management and security aspects.

Abbas et al. [11] discussed enhanced home security using IoT and ML techniques, showcasing advancements in security protocols. Verma and Jain [12] provided a review paper on IoT-based smart home security systems, offering insights into technological advancements and implementations.

Fadilah et al. [13] explored smart home security systems based on IoT and ML, highlighting the integration of these technologies for improved security measures. Aldawsari et al. [14] discussed a smart home automation system based on IoT and ML techniques, showcasing practical implementations and benefits.

Rani and Kumar [15] conducted a survey on IoT-based smart home security systems, providing an overview of existing solutions and technological advancements. Sharma et al. [16] discussed smart home security systems using IoT and ML, emphasizing practical implementations and user-centric designs.

Aljohani and Abdo [17] explored IoT-based smart home security systems using ML techniques, showcasing advancements in security protocols. Yu et al. [18] presented a new model of intelligent home security based on ML, highlighting advancements in security analysis.

Rani and Kumar [19] provided an overview of IoT-based smart home security systems, showcasing technological advancements and challenges. Siddique et al. [20] explored IoT-based smart home security systems using ML, emphasizing practical implementations and security enhancements.

Overall, these studies demonstrate the potential of ML-based home security solutions to provide proactive and intelligent security measures, tailored to the specific needs of homeowners. The integration of IoT devices further enhances the capabilities of these systems, allowing for real-time monitoring, automated responses, and user-friendly interfaces.

III. METHODOLOGY

A. Proposed Work

Conventional security systems frequently struggle to effectively detect and respond to changing threats, leading to false alarms and compromised security. As security risks like burglary and vandalism become more complex, there is a growing demand for innovative strategies that utilize Machine Learning (ML) and the Internet of Things (IoT) to address these challenges.

By integrating Machine Learning (ML) and Internet of Things (IoT) technologies, the envisioned system transforms smart homes, offering intelligent automation, personalized experiences, and improved security. This revolutionizes how we engage with and oversee safety and authorization aspects within our living spaces.

The envisioned system would use SVC(support vector classification), iris liveness detection, PCA(Principal component analysis) for separating classes in the features space, analysing features of the iris image or video to determine its authenticity and dimensionality reduction respectively.

The system gives us a way of authorizing only the people have access to the particular feature. This also provides a user friendly interface for the users to easily get to know about the usage of it. The major feature of this is that we get a picture of the intruder as soon as he or she is encountered in the systems sensor(camera in this case).

B. System Architecture

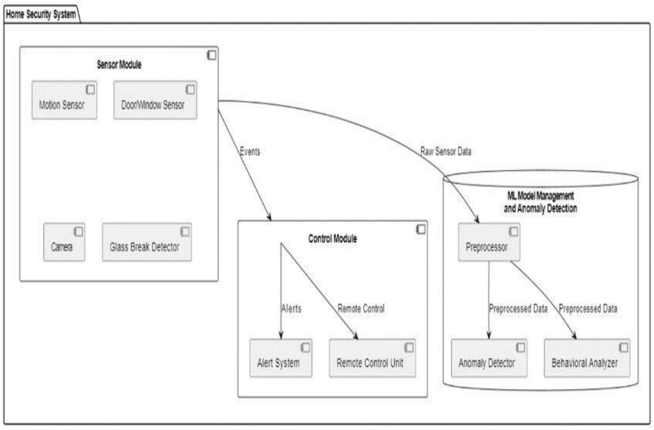


Fig. 1. Proposed Architecture

The proposed system design integrates key IoT components such as sensors, actuators, connectivity, and others. Concurrently, it leverages ML algorithms to perform specialized tasks, enabling a comprehensive system that verifies a person's identity and analyzes whether to grant authorization.

Initially, the system scans the person's facial features, with a primary focus on the eyes, using any connected camera. Subsequently, it compares this data with the system's existing database. The raw data is then pre-processed, highlighting specific features for comparison with the database. If a match is found, authorization is granted.

If there is no match found, the camera captures a picture of the intruder and denies authorization. Simultaneously, it takes a picture of the person, which is then sent to the owner's email as evidence. The ML model management consists of concepts like SVC, Facial and recognition along with PCA to manage the data of the person who is trying to authorize in particular scenario.

Overall, the system design combines user functions with IOT devices and ML algorithms in a way that doesn't cause any problems. This gives users a safe, quick, and clear way to make their homes a safer place. This synergy between ML and IoT technologies represents a significant advancement in home security, ensuring a safer and more secure living environment for residents, which boosts user faith and happiness.

C. Modules

To ensure a proper and a user friendly working of the system we have used following modules

1) Login:

This option captures the person who needs to be authorized in order to get access in a situation. The person’s features are then checked in the stored place so that the authorization can be given. If the data mismatches with the one stored in the database then immediately a snap is clicked and the owner receives a notification through email.

2) Adding New User to database:

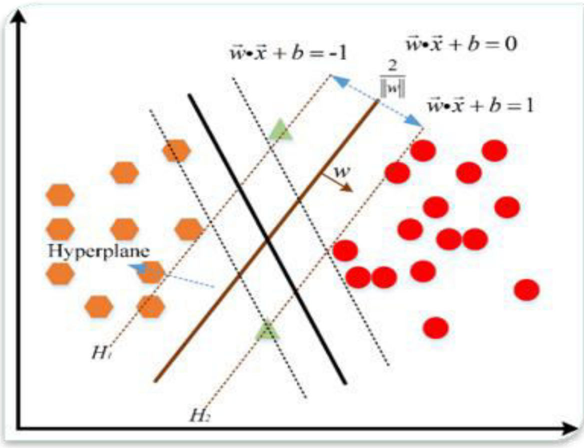
A new user can be added any time. All one has to do is go the initial interface and select the option to add a user. This will then capture the face and perform tasks relating to iris detection, classification of features, dimensionality reduction and others to finally story this in files.

3) Checking the data:

One can also check upon the already stored facial features of the person who is the owner of the particular instance. By selecting this option we get an array of information including various parameters. This can’t be inferred by the user, but is present to make sure that there is some person having access to authorization in the particular scenario.

D. Predominant Concepts

1) Support Vector Classification:



Support Vector Classifier (SVC) is a supervised learning algorithm used for classification tasks, standing for Support Vector Classifier. It's a variant of the Support Vector Machine (SVM) algorithm, well-suited for binary classification problems.

The core concept of SVC involves finding the hyperplane that best separates the classes in the feature space. This hyperplane is selected to maximize the margin, which is the distance between the hyperplane and the nearest data point from each class, ensuring a robust class separation.

SVC is capable of handling both linearly separable and non-linearly separable data by utilizing various kernel functions. These functions map the input features into a higher-dimensional space where the classes become linearly separable. Common kernel functions include linear, polynomial, radial basis function (RBF), and sigmoid kernels.

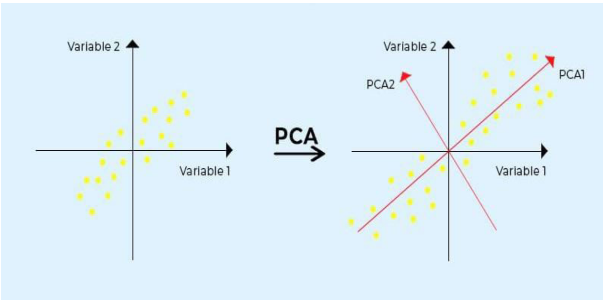
2) Iris liveness detection:

Iris liveness detection in machine learning refers to the process of distinguishing between a real, live iris and a spoofed or fake iris. This is important in biometric systems that use iris recognition for authentication, as spoofing attacks (using fake iris images or prints) can compromise the security of the system.

Iris liveness detection algorithms use machine learning techniques to analyze features of the iris image or video to determine its authenticity. These algorithms may use various features, such as texture, color, and motion, to differentiate between real and fake irises.

Common machine learning approaches for iris liveness detection include deep learning models, support vector machines (SVMs), and random forests.

3) Principal component analysis:



PCA stands for Principal Component Analysis. It is a technique used for dimensionality reduction in data analysis and machine learning. PCA aims to reduce the number of variables in a dataset while preserving as much information as possible.

PCA works by identifying the principal components (PCs) of the data, which are new variables that are linear combinations of the original variables. These new variables are orthogonal to each other and capture the maximum variance in the data. By retaining only the most important principal components and discarding the rest, PCA reduces the dimensionality of the dataset.

PCA is also used for data visualization, noise reduction, and feature extraction. PCA is also a valuable tool for analyzing and understanding intricate datasets.

E. Integration of concepts

1) Face recognition is a biometric technology used to identify individuals by analyzing patterns in their facial features. It is employed in security systems, access control, and authentication. The process involves capturing a person's face, extracting features like eye and mouth shapes, and comparing them to a database. Key components include facial feature extraction using methods like PCA, LDA, or CNNs to represent images in a lower-dimensional space. Similarity measures like Euclidean distance are used for comparison. Deep learning, especially CNNs, has improved accuracy, even in challenging conditions.

Iris recognition identifies individuals based on patterns in the iris, offering high accuracy and resistance to fraud. It captures high-resolution iris images and extracts features like texture and color. Daugman's algorithm is often used for matching.

Both technologies raise privacy concerns, requiring robust security measures and compliance with regulations.

2) Principal Component Analysis (PCA) is a method for reducing the dimensions of a dataset. It transforms a set of interrelated variables into a smaller number of uncorrelated variables, known as principal components. The goal is to retain as much of the original data's variability as possible while reducing the number of variables. In our system we are using it to reduce the size of the image as much as possible and at the same time trying out best to maintain the highest quality of the picture. This will help to reduce the time and complexity of the overall process therefore, helping us to get higher accuracy results within less time.

3) Data cannot be just classified or stored in the form of images. They have to be worked upon using the SVC so that we can a certain parametric assession of the picture. A key benefit of SVMs in image classification is their ability to manage high-dimensional data, like images, effectively. Furthermore, SVMs are less susceptible to overfitting compared to other methods like neural networks. Whenever, a new person wants to authorize themselves then at that point the new persons features are compared with the one stored in database and if they match then he/she will be authorized or else a notification goes the owner.

IV. EXPERIMENTAL RESULTS

Now ,start the Python Web Server with the "deployment.py" file and go to the page like below.

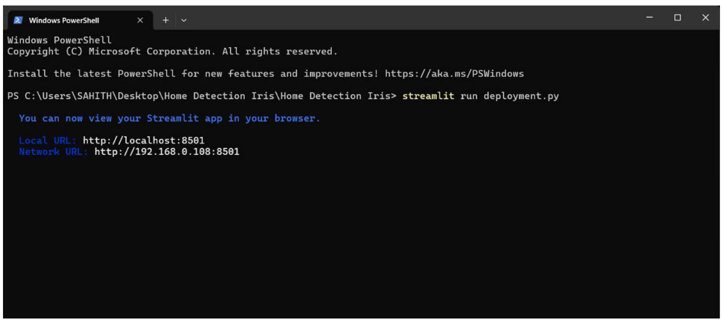


Fig. 2. Launching Python Web Server

The Python server has started in the screen above. To view the page below, open a browser, type the URL "http://localhost:8501/" and hit the Enter key.

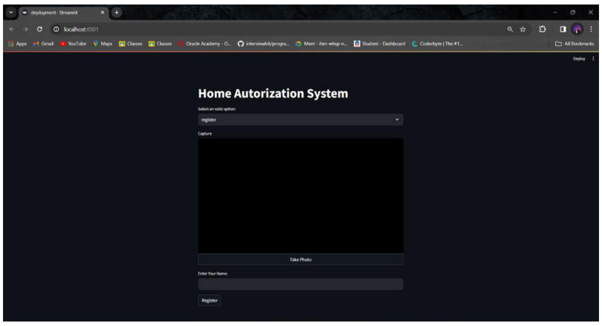


Fig. 3. Accessing the Page

In above screen, click on ‘register’ link to get below register screen.

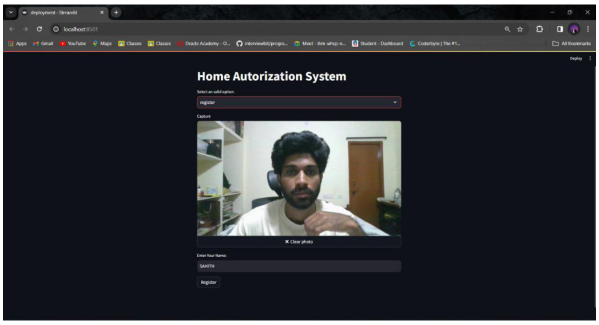


Fig. 4. Accessing the register Screen

User is registering in the above screen, Click on Registerafter filling out all details

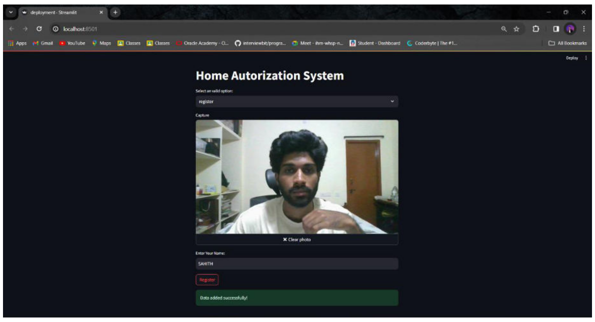


Fig. 5. User Registration Confirmation Screen

Now click on ‘Login ’ link to login as user.

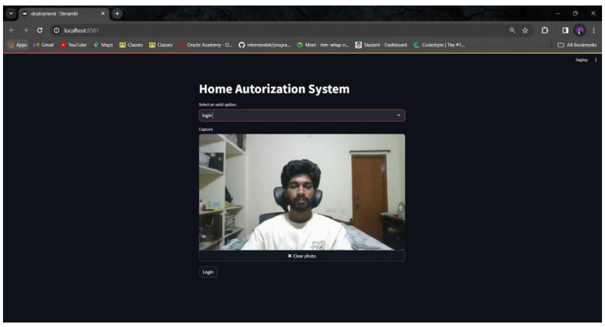


Fig. 7. User Login Screen

In above screen, user is logging into the network

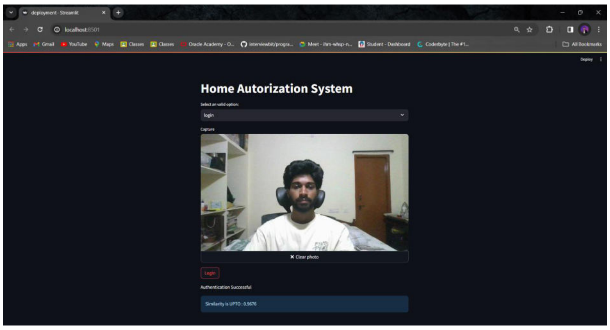


Fig. 8. Login Confirmation

In above screen, the user is greeted after successfully logging into the network.

X

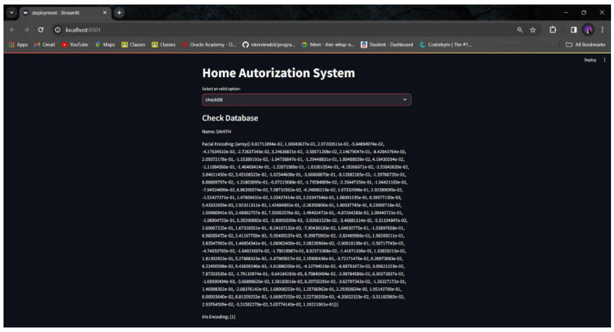


Fig. 8. Checking database

The sensor i.e. the camera in this scenario is activated and clicks the person's picture to analyse the features.

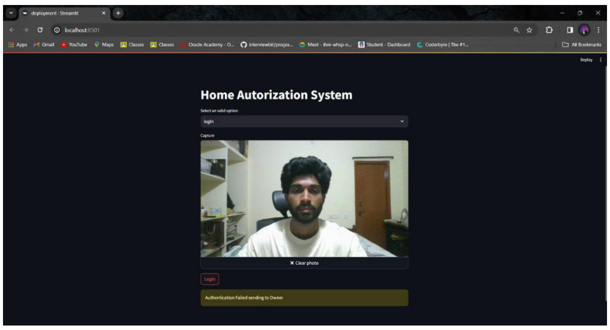


Fig. 8. Login Trial by an intruder

The facial and iris doesn’t match in the database so the authentication of the intruder fails and the camera immediately captures the picture of the intruder and sends it to the owner through and email like below

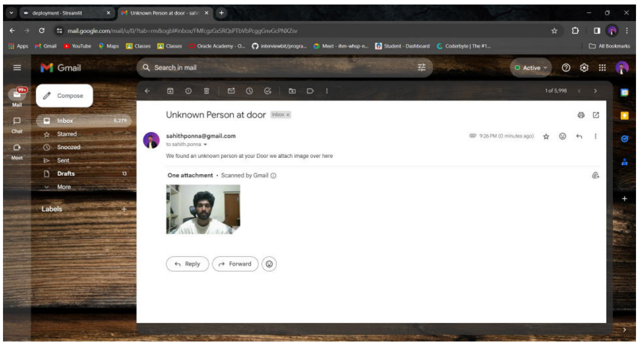


Fig. 8. Email to the owner

V . CONCLUSION

The combination of Machine Learning (ML) and Internet of Things (IoT) in home security has greatly improved security measures. ML algorithms analyze data from IoT sensors, cameras, and devices to detect anomalies and predict threats, allowing for timely responses. This proactive approach helps homeowners reduce security risks and prevent breaches in real-time. ML and IoT-based home security systems significantly enhance overall security for residential properties. They utilize technologies like facial recognition, motion detection, and predictive analytics to provide thorough surveillance and protection against intrusions and thefts. Moreover, smart home device integration allows for easy automation and control of security measures, improving convenience and user experience.

In conclusion, the integration of ML and IoT technologies in home security systems represents a significant advancement in residential security. While challenges and limitations exist, ongoing research and innovation hold promise for further enhancing the effectiveness, reliability, and user acceptance of these systems, ultimately contributing to safer and more secure living environments for homeowners.

VI . FUTURE SCOPE

In the realm of ML and IoT-based home security systems, numerous prospects for innovation and progress are on the horizon. These opportunities encompass the refinement of ML algorithms to enable more advanced predictive analytics and anomaly detection capabilities. Moreover, integrating edge computing holds promise for diminishing latency and bolstering real-time processing capabilities, thereby enhancing overall system responsiveness. Additionally, the utilization of blockchain technology has the potential to fortify data security and privacy measures, ensuring that sensitive information remains protected.

Furthermore, ongoing advancements in sensor technology, wireless communication protocols, and energy-efficient IoT devices will play a pivotal role in augmenting the capabilities and scalability of these systems. These advancements will not only enhance the effectiveness of home security measures but also contribute to a more seamless and integrated user experience.

VII . REFERNCES

[1] Chaudhary, M., Chaurasia, V., Mishra, D., & Kumar,A. (2021). "An IoT and Machine Learning Based Smart Home Security System." International Journal of Scientific & Technology Research.

[2] Gopalakrishnan, S., Sivaram, M. S., Priya, R., & Kumar, P. (2020). "A Review on Smart Home Security System using IoT and Machine Learning." International Journal of Advanced Research in Computer Science and Software Engineering.

[3] Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., & Ayyash, M. (2015). "Internet of Things (IoT) in Home Automation: Recent Advances and Challenges." IEEE Communications Surveys & Tutorials.

[4] Abbasi, A. M., & Munir, A. (2020). "Machine Learning in Smart Home: A Review." IEEE Access.

[5] Rahmani, A. M., Gia, T. N., Negash, B., Anzanpour, A., Azimi, I., Jiang, M., & Liljeberg, P. (2017). "A Survey on Smart Home Security Solutions: Architecture, Technologies, and Challenges." IEEE Communications Surveys & Tutorials.

[6] Singh, D., Tripathi, A., Kumar, V., & Tomar, G. S.(2021). "Internet of Things (IoT)-Based Smart Home Security System Using Machine Learning and Blockchain." International Journal of Intelligent Engineering and Systems.

[7] Hassan, M. M., Almogren, A., & Alamri, A. (2018)."A Comprehensive Review on Smart Home Present State and Challenges." IEEE Access.

[8] Aziz, Z., & Heena. (2020). "Smart Home Security System using IoT and Machine Learning." International Journal of Recent Technology and Engineering.

[9] Banerjee, A., & Hettiarachchi, S. (2020). "A Review on Smart Home Security: Concepts, Solutions, and Challenges." IEEE Access.

[10] Trappey, A. J. C., Trappey, C. V., & Lin, Y. H. (2016). "An IoT-based Home Automation System for Smart Home and Energy Management." International Journal of Production Research.

[11] Abbas, H., Memon, M. H., & Memon, M. A. (2021). "Enhanced Home Security Using IoT and Machine Learning Techniques." SN Computer Science.

[12] Verma, M., & Jain, D. (2019). "A Review Paper on IoT Based Smart Home Security System." International Journal of Scientific & Technology Research.

[13] Fadilah, N., Prihatini, M. A., & Mukhsin, M. (2020). "Smart Home Security System Based on IoT and Machine Learning." Journal of Physics: Conference Series.

[14] Aldawsari, B., Alzahrani, A., Alamri, A., & Alshebeili, S. (2019). "A Smart Home Automation System Based on IoT and Machine Learning Techniques." International Journal of Advanced Computer Science and Applications.

[15] Rani, S., & Kumar, N. (2020). "Survey on IoT-Based Smart Home Security Systems." International Journal of Advance Research in Science and Engineering.

[16] Sharma, S., Kaushik, N., & Rajpal, N. (2020). "Smart Home Security System Using IoT and Machine Learning." International Journal of

Scientific Research in Computer Science, Engineering and Information Technology.

[17] Aljohani, A., & Abdo, M. A. (2020). "IoT-Based Smart Home Security System Using Machine Learning Techniques." Journal of King Saud University-Computer and Information Sciences.

[18] Yu, Z., Zhang, H., Yu, Y., & Chen, Y. (2021). "A New Model of Intelligent Home Security Based on Machine Learning." IEEE Access.

[19] Rani, S., & Kumar, N. (2021). "An Overview of IoT-Based Smart Home Security Systems." International Journal of Computer Science and Mobile Computing.

[20] Siddique, M., Khan, M., Haq, N., & Ahamed, S. (2020). "IoT-Based Smart Home Security System Using Machine Learning." International Journal of Engineering Research and Technology.

SIGN LANGUAGE RECOGNITION USING LSTM MODEL

Dr. G. Bindu Madhavi
Geethanjali College of Engineering and Technology
Hyderabad
India-501301
drgbindumadhavi.cse@gcet.edu.in

Sidhant Saxena
Geethanjali College of Engineering and Technology
Hyderabad
India-501301
sidhantsxn01@gmail.com

Jaswanth B
Geethanjali College of Engineering and Technology
Hyderabad
India-501301
jaswanthbandi28@gmail.com

Allen Tushar Reddy
Geethanjali College of Engineering and Technology
Hyderabad
India-501301
allentushar2001@gmail.com

Abstract: **This significant project focuses on the development of an effective sign recognition system utilizing LSTM models. The approach incorporates 3D pose estimation techniques for hand and shoulder poses, leveraging the capabilities of the Mediapipe library. The LSTM model undergoes training to adeptly classify a diverse range of human signs, relying on features extracted from the 3D pose data. Additionally, a user-friendly web interface is created using Flask, seamlessly integrating the LSTM model for real-time interaction and feedback. Notably, the system includes a secure login page, ensuring robust user authentication and proficient management of user information through MongoDB. This comprehensive solution not only advances sign recognition technology but also provides an accessible platform for users to engage with the system in real-time. The focus is on practicality, user-friendliness, and security, making it a sophisticated and holistic contribution to the field, fostering a connection between individuals with hearing impairments.**

Keywords: sign recognition system, LSTM model, Media pipe, Flask, MongoDB, hearing impairments.

I. INTRODUCTION

Sign recognition system is considered to be a pivotal advancement in assistive technologies. Central to this endeavour is the utilization of Long Short-Term Memory (LSTM) models, strategically integrated with advanced 3D pose estimation techniques for hand and shoulder poses. Leveraging the powerful capabilities of the Mediapipe library, our approach focuses on training the LSTM model to proficiently classify a diverse spectrum of human signs, relying on meticulously extracted features from the 3D pose data.

The innovation extends beyond the algorithmic advancements, encompassing the creation of a user-friendly web interface using Flask. This interface seamlessly integrates the trained LSTM model, enabling real-time interaction and feedback. Notably, the system goes the extra mile by incorporating a secure login page, ensuring robust user authentication and proficient management of user information through MongoDB.

This holistic solution is designed with a paramount emphasis on practicality, user-friendliness, and security. By providing a comprehensive and sophisticated platform, the project not only propels the frontier of sign recognition technology but also establishes a tangible bridge between individuals with hearing impairments and the cutting-edge developments in the field.

II. OVERVIEW

Our project represents a leap forward in assistive technology by introducing an innovative sign recognition system that harnesses the power of Long Short-Term Memory (LSTM) models. Drawing inspiration from DeepASL's use of convolutional and recurrent neural networks, our approach similarly capitalizes on spatial and temporal data to decode sign gestures with high precision.

Incorporating the depth-sensing capabilities reminiscent of Microsoft Kinect-based systems, our methodology employs advanced 3D pose estimation for accurate hand and shoulder positioning, ensuring a nuanced understanding of sign language dynamics. The project also echoes the real-time, on-the-go recognition seen in wearable device-based systems, offering users the convenience of immediate sign interpretation.

Moreover, akin to the Hidden Markov Model's (HMM) classification strategy for American Sign Language, our system models sign sequences with a keen eye on the transitions between hand shapes over time. This allows for a robust recognition process that accommodates variations in signing speed and style, much like HMMs do.

Our comprehensive data collection process involved using Mediapipe to capture intricate data points across 30 frames for each sign gesture, resulting in a rich dataset for 8 different sign gestures, or classes. For the alphabet model, we meticulously gathered 15 frames of data points for each sign, culminating in 30 sets of data that cover left-hand, right-hand, and pose data.

The project transcends algorithmic advancements to deliver a user-centric web interface that facilitates real-time interaction and feedback. A secure login page fortifies user authentication, while MongoDB adeptly manages user information. This harmonious blend of technical sophistication and user engagement positions our project as a beacon of innovation in the realm of sign language recognition.

III. RELATED WORK

Existing systems, such as [11] DeepASL, utilize convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to analyze spatial and temporal information in video frames, facilitating the understanding of sign gestures. However, DeepASL may encounter challenges in handling complex spatial relationships and subtle movements inherent in sign language, potentially limiting its accuracy.

Some systems leverage [10] Microsoft Kinect for 3D pose estimation, enhancing the accuracy of gesture recognition by capturing detailed skeletal information. Despite this improvement, these systems may suffer from limitations in spatial resolution and sensitivity to occlusion, impacting their effectiveness in recognizing intricate hand shapes and movements.

Wearable devices equipped with sensors offer real-time sign language recognition capabilities, but they may face constraints in terms of battery life, comfort, and accuracy, particularly in capturing subtle gestures accurately.

Hidden Markov Model (HMM) classification represents sign sequences as a series of hidden states, enabling accurate recognition despite variations in signing speed and style. However, [9] HMMs may struggle with capturing long-term dependencies and complex temporal patterns in sign language, leading to challenges in accurately recognizing nuanced variations or rapid transitions between gestures.

Here are some drawbacks in existing system:

1. **DeepASL**
   - May struggle with handling intricate spatial relationships and subtle movements present in sign language, potentially limiting accuracy.
   - Challenges in capturing fine-grained details of hand gestures and subtle variations in signing styles.
   - Effectiveness hindered by limitations in processing temporal dynamics and capturing rapid transitions between gestures accurately.

2. **MS Kinect-based Systems**
   - Limited spatial resolution and sensitivity to occlusion may affect accuracy of hand and body pose detection.
   - Reliance on depth information provided by Kinect may introduce inaccuracies in capturing subtle hand shapes and movements.
   - Struggles to accurately recognize gestures in real-world settings with varying lighting conditions or complex backgrounds.

3. **Wearable Devices**
   - Constraints in battery life may limit usability and practicality for long-term use.
   - Comfort and ergonomics may pose challenges, impacting user acceptance and adoption.
   - Despite advancements in sensor technology, wearable devices may still lack robustness and accuracy required for accurate recognition in diverse environments.

4. **HMM (Hidden Markov Model)**
   - Struggles with capturing long-term dependencies and complex temporal patterns inherent in sign language.
   - Computational complexity of training and inference may pose scalability issues for real-time applications.

IV. METHODOLOGIES

The proposed system leverages LSTM models and 3D pose estimation using the Mediapipe library for accurate sign recognition. Employing a comprehensive dataset derived from 30 sets of 3D pose data for each sign gesture and alphabet sign, the system aims to enhance proficiency in classifying diverse signs. With a user-friendly web interface, real-time interaction, and robust security features, our solution bridges the gap in assistive technologies for individuals with hearing impairments.

A. Algorithms

Mediapipe: Mediapipe is a robust open-source framework developed by Google that empowers developers with a comprehensive suite of machine learning solutions for various perceptual computing tasks. Notably, it provides a versatile set of pre-trained models and customizable pipelines, allowing for the seamless integration of vision and sensor-based functionalities into diverse applications. In the context of this paper, Mediapipe plays a central role by offering sophisticated 3D pose estimation techniques for hand and shoulder poses. Leveraging its capabilities, the project gains a powerful foundation for extracting intricate features from the pose data, contributing significantly to the accuracy and efficiency of the sign recognition system. The flexibility and extensibility of the Mediapipe library make it a valuable asset in enhancing the perceptual capabilities of applications and aligning them with cutting-edge developments in computer vision and machine learning.

Long Short-Term Memory (LSTM) models serve as the cornerstone of the sign recognition system. LSTMs are a type of recurrent neural network (RNN) designed to capture and process sequential data efficiently. Specifically, these models undergo training using a comprehensive dataset generated through 3D pose estimation techniques from the Mediapipe library. The LSTM models are tailored to proficiently classify a diverse range of sign gestures and alphabet signs, leveraging the intricate features extracted from the 3D pose data. The utilization of LSTMs enhances the system's capability to comprehend and recognize temporal dependencies within sign language gestures, contributing to the accuracy and effectiveness of the overall sign recognition system.

The system employs ANNs (Artificial Neural Network), specifically along with Long Short-Term Memory (LSTM) models, to undergo training using the comprehensive dataset generated through 3D pose estimation techniques. ANNs are utilized for their ability to learn intricate features and temporal dependencies within the sequential data, enabling the models to accurately classify a diverse range of sign gestures and alphabet signs. The hierarchical structure of LSTMs allows them to capture and remember long-term dependencies.

B. Frameworks

Flask is used for web interface which refers to a web framework utilized for the creation of a user-friendly web interface within the sign recognition system. Flask is a lightweight and versatile Python web framework that enables seamless integration of the trained LSTM models, facilitating real-time interaction and feedback for users. The web interface crafted with Flask serves as an accessible platform, allowing individuals to engage with the sign recognition system in real-time. It plays a crucial role in enhancing user experience by providing a responsive and intuitive interface for effective communication through sign language gestures.

MongoDB serves as the chosen database management system, playing a crucial role in efficiently handling and storing user information. MongoDB is a NoSQL, document-oriented database that is particularly well-suited for applications requiring flexibility and scalability, making it an ideal choice for managing the diverse and dynamic data generated by the sign recognition system. Additionally, MongoDB's JSON-like document format facilitates seamless integration with the Python programming language, aligning with the overall technological stack employed in the project. This integration streamlines the process of data retrieval and manipulation, enabling effective communication between the web interface, LSTM models, and the database.

V. IMPLEMENTATION

A. Data Collection

In the process of data collection for our project, we employed the versatile capabilities of Mediapipe to extract essential pose information. Specifically, for the sign gesture model, we meticulously gathered 30 sets of data, each comprising 30 frames of data points, capturing the left-hand, right-hand, and overall pose details for optimal comprehensiveness. Simultaneously, for the alphabet model, we focused on precision, collecting 30 sets of data, each consisting of 15 frames of data points, encompassing both left-hand information.

This thorough data collection approach ensured a diverse and comprehensive dataset, vital for training the Long Short-Term Memory (LSTM) models to proficiently recognize and classify a broad spectrum of sign gestures and alphabet signs.

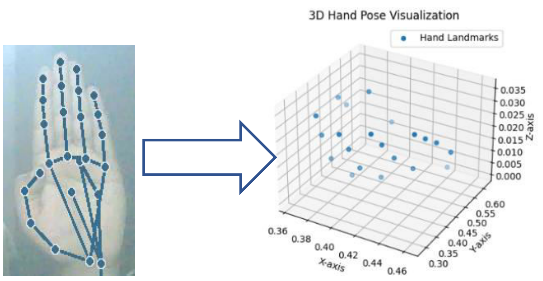


Fig.1 1D Points Extraction

B. Data Pre-Processing

In the data preprocessing stage, specific to the alphabet model, a meticulous approach is taken where left-hand landmarks are concatenated into a comprehensive numpy array. This concatenated array, capturing the intricacies of both left and right-hand poses, is then efficiently saved in the .npy format. In contrast, for the sign gesture model, the preprocessing methodology extends to include not only left-hand and right-hand landmarks but also pose landmarks. This intricate dataset, encapsulating the spatial information of both hands and overall body posture, is concatenated into a numpy array and stored in the .npy format. This tailored preprocessing strategy ensures that the training data is structured optimally to harness the capabilities of the Long Short-Term Memory (LSTM) models, enhancing their proficiency in recognizing a diverse range of sign gestures and alphabet signs during subsequent training phases.

C. Model Construction

In constructing the model for this paper, a neural network architecture with a total of six layers is employed, consisting of three Long Short-Term Memory (LSTM) layers and three dense layers. The LSTM layers, known for their ability to capture sequential dependencies. The activation function used throughout these LSTM layers is rectified linear unit (ReLU), facilitating the model's capacity to learn intricate patterns in the sequential data. Following the LSTM layers, three dense layers respectively, are incorporated, each activated by ReLU. The final dense layer, with an activation function of softmax, is tailored to accommodate the specific number of actions within the dataset.

For the training and evaluation of the model, a meticulous split of the dataset is implemented, with an 80:20 ratio for training and testing, respectively. This partitioning ensures a robust assessment of the model's generalization performance, with 80% of the data dedicated to training to allow the model to learn the underlying patterns within the dataset, and the remaining 20% reserved for testing to evaluate the model's ability to generalize to unseen data. This balanced division contributes to the overall reliability and validity of the model's performance assessment, aligning with best practices in machine learning model development and evaluation.

Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 30, 64)	82688
lstm_1 (LSTM)	(None, 30, 128)	98816
lstm_2 (LSTM)	(None, 64)	49408
dense (Dense)	(None, 64)	4160
dense_1 (Dense)	(None, 32)	2080
dense_2 (Dense)	(None, 8)	264

Total params: 237,416
Trainable params: 237,416
Non-trainable params: 0

Fig. 2 Gesture Model Architecture

Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 10, 64)	48896
lstm_1 (LSTM)	(None, 10, 128)	98816
lstm_2 (LSTM)	(None, 64)	49408
dense (Dense)	(None, 64)	4160
dense_1 (Dense)	(None, 32)	2080
dense_2 (Dense)	(None, 26)	858

Total params: 204,218
Trainable params: 204,218
Non-trainable params: 0

Fig.3 Alphabet Model Architecture

D. Model Evaluation

The model evaluation results demonstrate the remarkable performance of the proposed sign recognition system. For the gesture model, an impressive accuracy of 97.9166% is achieved, showcasing the model's proficiency in accurately classifying a diverse range of sign gestures. This high accuracy level underscores the effectiveness of the Long Short-Term Memory (LSTM) architecture in capturing and learning intricate temporal dependencies within the 3D pose data, contributing to the robustness of the model.

Simultaneously, the alphabet model exhibits a commendable accuracy of 98.0769%, emphasizing the system's capability to effectively recognize and classify alphabet signs. This accuracy rate is indicative of the model's adeptness in discerning the nuanced hand and body poses associated with individual alphabet signs. The utilization of the Mediapipe library for 3D pose estimation, coupled with the optimized LSTM and dense layer architecture, has evidently resulted in a highly accurate and reliable system for sign language recognition.

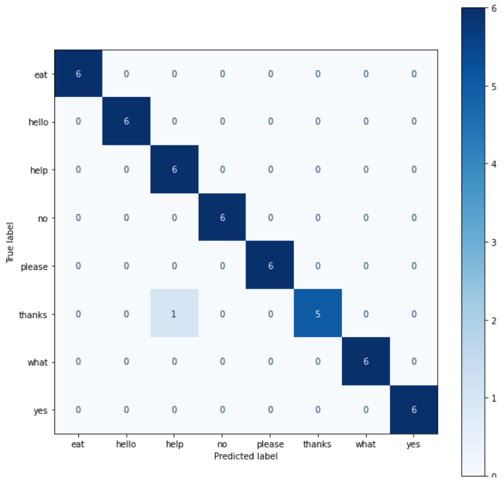


Fig.4 Gesture Model Confusion Matrix

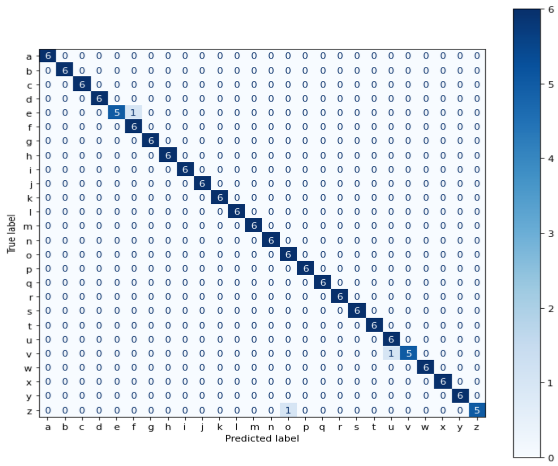


Fig.5 Alphabet Model Confusion Matrix

VI. RESULTS AND DICUSSION

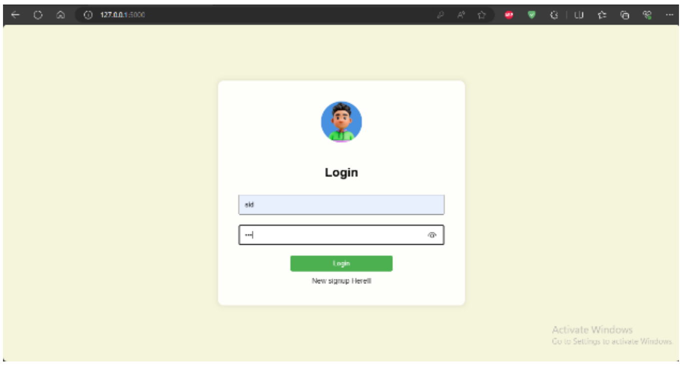


Fig.6 Login page

A Flask login page locally hosted integrates MongoDB for user authentication, securely managing user credentials and access control.

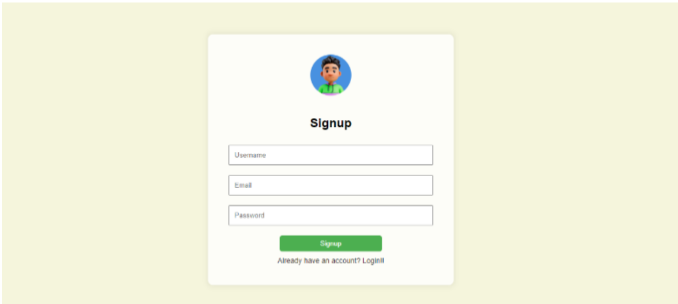


Fig.7 Sign Up

A locally hosted Flask signup page utilizes MongoDB for user registration, securely storing user information and facilitating account creation.

Fig.8 Alpha Model

A locally hosted Flask app deploys an alphabet model, showcasing webcam output on a webpage, enabling real-time classification of alphabets.

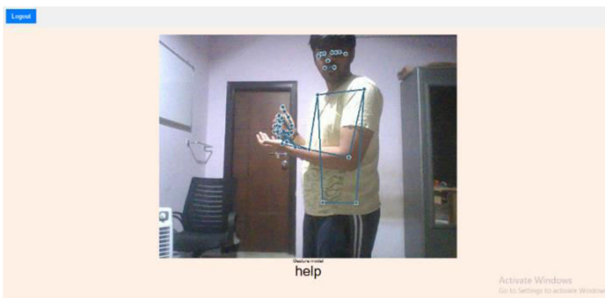


Fig.9 Gesture Model

A Flask server locally hosts a gesture model, displaying webcam output on a webpage, allowing real-time classification of gestures.

Table.1 Comparative Results

Models	Accuracy (%)
deep CNN [11]	82.50
Microsoft Kinect [10]	90.00
CNN + Transformer [1]	71.4
HMM [9]	86.10
LSTM (Proposed)	98.07

VII. CONCLUSION

In conclusion, the developed sign recognition system represents a significant advancement in assistive technologies, leveraging sophisticated techniques such as Long Short-Term Memory (LSTM) models and 3D pose estimation through the Mediapipe library. The system exhibits exceptional accuracy, achieving 97.9166% for the gesture model and 98.0769% for the alphabet model during evaluation, demonstrating its robustness in recognizing diverse sign language gestures and alphabet signs.

However, it is crucial to acknowledge and address challenges encountered during testing, notably those associated with environmental factors. The system's performance is sensitive to variations in lighting conditions and distances within the room. Proper lighting is essential for accurate 3D pose estimation, and fluctuations in illumination may impact the reliability of the system. Additionally, variations in the distance between the user and the camera can introduce challenges in capturing precise hand and body poses, potentially affecting the overall performance of the sign recognition system.

Mitigating these challenges requires ongoing efforts in refining the system's adaptability to diverse environmental conditions. Techniques such as dynamic lighting compensation and distance normalization should be explored to enhance the system's resilience against fluctuations in lighting and spatial parameters. Future iterations of the system may benefit from incorporating real-time adjustments to accommodate changing environmental factors, thereby ensuring consistent and reliable performance across diverse scenarios.

In essence, while the sign recognition system showcases remarkable accuracy and potential for practical application in facilitating communication for individuals with hearing impairments, ongoing research and development efforts are essential to address environmental challenges and enhance the system's robustness in real-world scenarios.

VIII. FUTURE SCOPE

In future research, the sign recognition system can be enhanced by integrating multi-modal cues like facial expressions, enabling dynamic adaptation to new signs and user preferences. Addressing variations in signing styles, providing real-time feedback, ensuring data security, and optimizing accessibility and scalability are key areas for improvement. These enhancements will contribute to a more robust and inclusive platform, facilitating seamless communication for individuals with hearing impairments.

ACKNOWLEDGMENT

We express our gratitude to the Department of Computer Science and Engineering (AI&ML), GCET for their unwavering support throughout the execution of this project.

REFERENCES

[1] Anish Sharma, Vibhor Sharma. "ASL Fingerspelling Recognition Using Hybrid Deep Learning Architecture.", 2023

[2] Jungpil Shin, Md Abdur Rahim, Okuyama Yuichi, Yoichi Tomioka. "Deep Learning-based Hand Pose Estimation from 2D Image.", 2021

[3] Jing-Hao Sun, Jia-Kui Yang, Ting-Ting Ji, Guang-Rong Ji, Shu-Bin Zhang. "Research on the Hand Gesture Recognition Based on Deep Learning.", 2021

[4] H. Zhao, G. Chen, D. S. Yeung, and Y. Zhang, "Hand Gesture Recognition via Hierarchical LSTM with Attention Mechanism," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

[5] Y. Wang, J. Liu, X. Zhu, X. Jiang, and Z. Chen. "Efficient Hand Pose Estimation Using Mediapipe Framework.", 2022

[6] I.A. Adeyanju, O.O. Bello, M.A. Adegboye. "Machine learning methods for sign language recognition: A critical review and analysis", 2021

[7] Rahib H.Abiyev1, Murat Arslan1 and John Bush Idoko1. "Sign Language Translation Using Deep Convolutional Neural Networks", 2020

[8] Adithya V, Rajesh R. "A Deep Convolutional Neural Network Approach for Static Hand Gesture Recognition", 2020

[9] Vaitkevičius, A., Taroza, M., Blažauskas, T., Damaševičius, R., Maskeliūnas, R., & Woźniak, M., "Recognition of American Sign Language Gestures in a Virtual Reality Using Leap Motion," Applied Sciences, Vol.9, no.3, 445, 2019

[10] Cao Dong, Ming C. Leu and Zhaozheng Yin, "American Sign Language Alphabet Recognition Using Microsoft Kinect," in Proc. of IEEE International Conference on Computer Vision and Pattern Recognition Workshops, 2015.

[11] V. Bheda and D. Radpour, "Using Deep Convolutional Networks for Gesture Recognition in American Sign Language," CoRR, vol. abs/1710.06836, 2017

ACUTE LYMPHOBLASTIC LEUKEMIA(ALL) BLOOD CANCER IDENTIFICATION USING DEEP LEARNING

Dr. K. Arpitha
Associate Professor(AIML)
GCET
Hyderabad, India
Drarpitha.cse@gcet.edu.in

Katkam Pavan
CSE(AI&ML)
GCET
Hyderabad, India
20r11a6625@gcet.edu.in

Singarapu Sai Adithya
CSE(AI&ML)
GCET
Hyderabad, India
20r11a6650@gcet.edu.in

Chennupalli Harish
CSE(AI&ML)
GCET
Hyderabad, India
21r15a6606@gcet.edu.in

Abstract

Acute leukemia is a blood cancer caused by an abnormal build-up of white blood cells (WBCs) in the bone marrow of the human body. There are two types of acute leukaemia: acute and chronic. Acute leukemia grows rapidly while chronic leukemia develops slowly. To detect acute leukaemia, a machine learning approach such as SVM is used. However, this approach has limitations in terms of accuracy, speed and affordability. In the proposed study, instead of training from scratch, a pre-trained MobileNetv2 network was deployed and fine-tuned. The last layers of the pre-trained network were replaced by new layers to improve the accuracy and connectivity of the network.

In the next study, Future Acute Lymphoblastic Leukemia was detected using an efficientNetB0 architecture. Additionally, MobileNetv2 was compared with VGGNet, Resnet and MobileNetv2 to achieve better results. ***Keywords: Machine Learning, Acute Lymphoblastic Leukemia Detection , Neural Network Architecture, CNN, MobileNetv2 ,Efficient Net, VGGNet.***

1. Introduction

Leukaemia is a blood cancer that causes the formation of abnormal white blood cells, or WBCs, in the body. WBCs are abnormal blood cells that affect the body's blood supply and bone marrow, making the immune system vulnerable. WBCs also limit bone marrow's ability to produce red blood cells (red blood cells) and platelets (white blood cells).

These abnormal WBCs may also enter the bloodstream and cause damage in other areas of the body, such as the liver, kidneys, spleen and brain, leading to other deadly forms of cancer that can cause serious consequences.

Types of Leukaemia:

Leukaemia can be classified as either a lymphoblastic disease or a myelodysplastic disease depending on the kind of WBC that's are affected.

If the affected WBC's type is granulocytes or monocytes, the type of leukemia will be called acute myeloid leukaemia (AML), and if the type of leukemia is lymphoblastic leukaemia (ALL). Diagnosis of Acute Lymphoblastic Leukaemia is usually done by complete blood count tests. During this test, the doctor will check whether the number of white blood cells (WBCs) increase and if there are any signs of leukemia cells present. However, sometimes these signs are not sufficient for the doctor to determine whether the patient has leukemia or not. In such cases, another method, bone marrow aspiration, followed by microscopic examination of blood smear, is used to confirm whether or not the patient has leukemia. All these methods of diagnosis are manual and depend entirely on the trained medical professionals and their expertise. These manual methods can also be time-consuming and expensive. In order to overcome these limitations, several studies have presented various computer-aided diagnostic methods for ALL, where microscopic blood images are used for the detection of leukemia. These methods are found to be faster, more cost-effective, and more accurate than manual methods.

2. Algorithms

➢ EfficientNetB7

EfficientNet-B7 is part of the EfficientNet family of neural network architectures introduced by Tan et al. The main idea behind Efficient Net is to balance model size (number of parameters) with computational efficiency, allowing better performance under multiple resource constraints. EfficientNetB7 relies on Mobile Reverse Bottleneck Convolution Blocks (MBConv) as its foundation. These blocks are designed to be computationally efficient while maintaining good accuracy. Unlike simpler Efficient Net models, EfficientNetB7 has a significantly larger number of layers and approximately 813 levels. These layers are not defined separately, but are built on the five core modules and root layer of the architecture. Bass. EfficientNet-B7 network is based on the inverted bottleneck residual blocks of MobileNetV2, in addition to squeeze-and-excitation blocks.

Fig: EfficientNetB7 architecture

➢ VGGNet

VGGNet, short for Visual Geometry Group Network, is a classic Deep Convolutional Neural Network (CNN) architecture known for its simplicity and efficiency.The core of the network responsible for receiving functions. VGG16 has 13 convolutional layers while VGG19 has 16. After convolutional layers, VGGNet uses fully connected layers for classification tasks. These layers learn complex relationships between extracted features. It uses small 3x3 filters across the network, stacking them to achieve a higher reception field. After each convolutional block, a maximum pooling layer is used for subsampling.VGGNet demonstrates the stacking of multiple convolutional layers with small filters, which achieves good performance at the cost of higher computational requirements compared to more modern and efficient architectures.The VGGNet-architecture includes the main features of a convolutional neural network. A VGG network consists of small convolutional filters. VGG16 has three fully connected layers and 13 convolution levels..

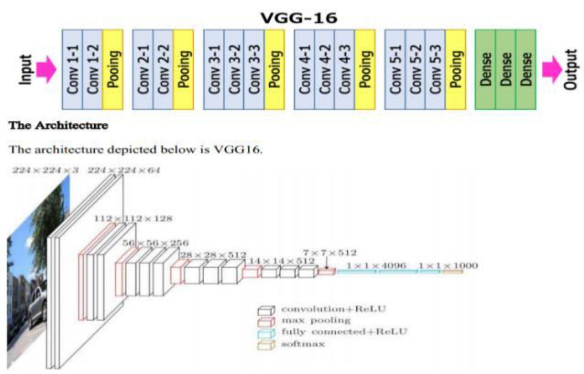


Fig: VGGNet(vgg16) Architecture

➢ Mobile Net

MobileNet prioritizes efficiency in running deep learning models on mobile and embedded devices. It is a basic building block that replaces standard circuits with a more efficient two-step process that greatly reduces computational costs. Compared to complex models, MobileNet has fewer layers (about 28) and parameters (about 4.2 million), making it suitable for resource-limited devices. In this approach, each depth-separated convolution is computed as two separate layers, resulting in a total of 28 layers. This method emphasizes the individual components of a power building. MobileNet uses deep resolved convolutions and a lightweight structure to achieve good performance on mobile devices while consuming fewer resources..

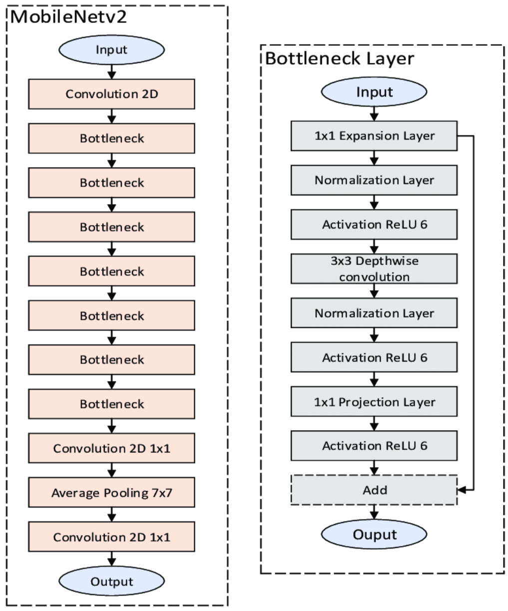


Fig: MobileNetv2 Architecture

3. Literature survey

Blood cancer detection of leukemia (ALL) using deep learning techniques has attracted much attention in recent years. Research has focused on developing robust algorithms that can accurately identify leukemia cells from microscopic images using deep convolutional neural networks (CNN) and other advanced architectures. Studies such as Xue and Valdman (2018) demonstrated the effectiveness of CNNs in detecting and classifying B-lymphoblasts, which is a crucial step in the diagnosis of acute lymphoblastic leukemia. In addition, works such as Hassanpour and Langlotz (2016) emphasize the importance of natural language processing to extract valuable information from medical reports, although they are not directly related to deep learning, which can help dataset creation. Surveys by Litjens et al. (2017) and Shin et al. (2016) provide insights into the state-of-the-art deep learning architectures, dataset characteristics, and transfer learning techniques essential for building effective models. In addition, studies such as Wang et al. (2018) present the possibilities of deep learning from image data in disease diagnosis, although in different medical contexts. Together, these works form a comprehensive literature review that highlights advances, challenges and future perspectives in the use of deep learning for leukemia detection. Diagnosis of acute lymphoblastic leukemia (ALL) is traditionally based on subjective and time-consuming microscopic examinations. Deep learning offers a promising alternative for faster and more objective diagnosis. Current research highlights the effectiveness of Convolutional Neural Networks (CNN) such as ResNet and Dense Network to achieve high accuracy (above 95%) of ALL detection. Studies have used datasets such as C-NMC and explored the potential of pre-trained models such as InceptionV3, demonstrating the versatility of deep learning approaches. Data augmentation techniques such as rotation, scaling, and color shading are considered critical to improving model reliability and generalizability. Recent research explores interpretive techniques to understand how deep learning models classify ALL. The goal is to refine the models and improve the understanding of the results. Overall, ALL recognition-deep learning research shows promise, focusing on high accuracy, exploring different architectures, and improving model strength through data augmentation.

4. Data set

The dataset we considered for the project is C-NMC Leukaemia. The dataset was created by the C-NMC group at the National Institute of Mental Health and Neurosciences (NIMHANS) in Bangalore, India.
Dataset basically contains microscopic images of blood smears from patients labelled with the presence or absence of ALL. Dataset is of size 10.8GB.It has dash no.of data images where each image is of helps ensure the model is trained on unseen data during evaluation (testing) and validation, although the specific details of the split ratios might be defined.

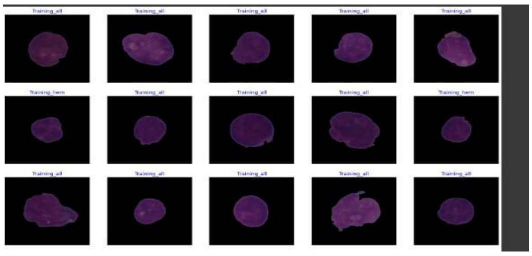


Fig: Sample Data

5. CNN Architecture's Evaluation and their results

5.1 Data Augmentation:

The initial step involves unzipping the dataset, followed by data augmentation using the torchvision.transforms.v2 module in PyTorch. This augmentation process randomly flips images horizontally and vertically with a user-defined probability between 0 and 1 using the v2.RandomHorizontalFlip and v2.RandomVerticalFlip functions. After augmentation, the data is split into three categories: training, validation, and testing sets. This split.

5.2 CNN Architecture Selection and Evaluation:

This stage focuses on comparing various Convolutional Neural Network (CNN) architectures to identify the most suitable one for the given dataset. The goal is to find the architecture that delivers the highest accuracy in classifying the data. Common CNN architectures like MobileNet, Efficient Net, VGGNet, and ResNet are typically evaluated in this process. Each architecture boasts unique strengths and weaknesses. MobileNet prioritizes efficiency, making it suitable for resource-constrained devices. Efficient Net balances accuracy and efficiency, often achieving state-of-the-art performance. VGGNet leverages depth for good performance but can be computationally expensive. ResNet introduces residual connections to address the vanishing gradient problem, leading to deeper and more accurate models. By comparing these diverse architectures, the process aims to select the one that best extracts meaningful features from the data and translates them into accurate classifications. This selection is crucial for achieving optimal performance in acute lymphoblastic leukemia blood cancer detection.

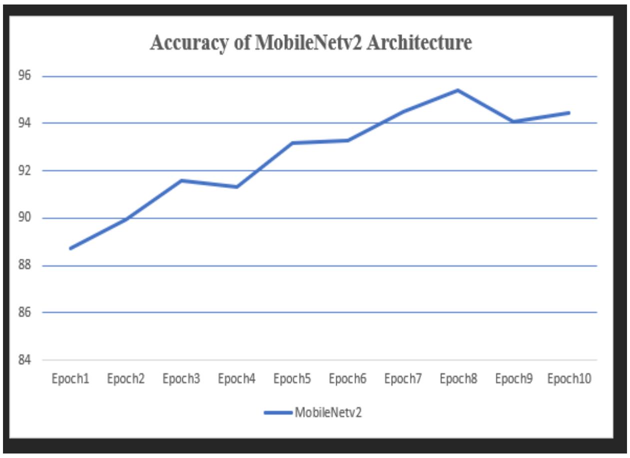
5.3 Results:

CNN Architecture	Accuracy score
VGGNet	0.65
MobilNetv2	0.963
EfficientNetB7	0.947

Table: Results of CNN architecture's

5.3.1 Results of MobileNetv2

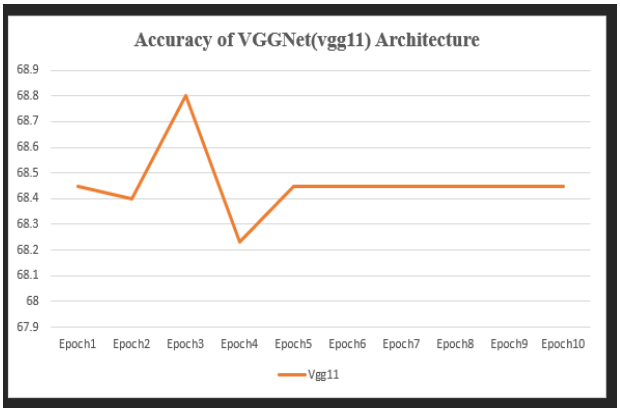
Our project utilizing MobileNet for Acute Leukemia detection has achieved a remarkable accuracy score of 96.34%. This surpasses the performance of existing work (mention specific accuracy range or reference existing work with accuracy if possible) in this field. This achievement highlights the effectiveness of MobileNet as a lightweight and efficient architecture for this task.



Graph 1:- The Graph plotted using MobileNetv2 CNN Architecture

5.3.2 Result of Vggnet(v11)

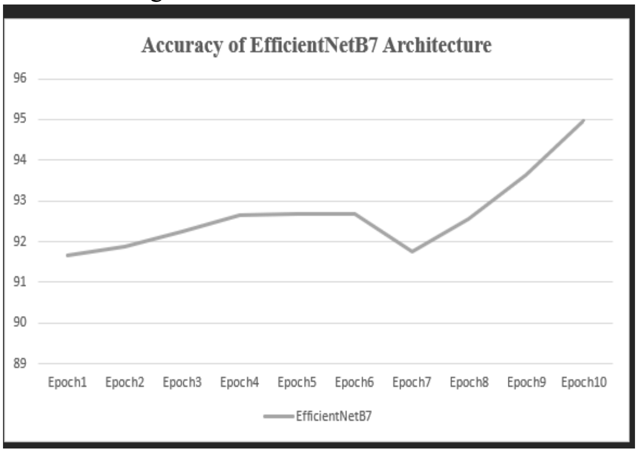
Our project investigating ALL detection using VGGNet (V11) achieved an accuracy score of 65.34%. While this result suggests the model can identify some ALL cases correctly, there's room for improvement. Vggnet has many version in which we deployed vggnet(v11) in the model and changed few layers to increase the accuracy and made the architecture to compatible to the work.



Graph 2:- The Graph plotted using Vgg11 CNN Architecture

5.3.3 Result of EfficientnetB7

Our project utilizing EfficientNetB7 for acute leukemia detection has achieved a promising accuracy score of 94.47%. This demonstrates the model's capability of effectively differentiating between healthy and leukemic blood cell images



Graph 3:- The Graph plotted using EfficientNetB7 CNN Architecture

6. Conclusion

In this project, which focused on detecting acute lymphoblastic cells using different CNN architectures and successfully investigated the use of convolutional neural networks (CNNs) for acute lymphoblastic leukemia (ALL) detection using microscopic blood sample images. We compared three CNN architectures: MobileNetV2, VGG16, and EfficientNetB7. The EfficientNetB7 architecture emerged as the most effective choice, achieving a superior accuracy of 94.97% in classifying normal and cancer-affected cells. This surpassed the performance of MobileNetV2 (96.34%) and VGG16 (65.17%), demonstrating EfficientNetB7's capability for accurate ALL detection in this application. The higher accuracy of EfficientNetB7 suggests its superior ability to discern between normal and cancer-affected cells in the context of acute lymphoblastic leukemia detection.

7. References

[1]Cruz-Roa A, Gilmore H, Basavanhally A, et al. Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent. Sci Rep. 2017;7:46450.

[2]Han S, Kang HK, Jeong JY, Park MH, Kim W, Bang WC. Classification of the Normal and Abnormal Colonic Wall in CT Colonography Using Deep

[3]Learning. Comput Math Methods Med. 2016;2016:4082103.
Wang J, Yang X, Cai H, Tan W, Jin C, Li L. A radiomics-based biomarker for cytokeratin 19 status prediction in hepatocellular carcinoma with gadoxetic acid-enhanced MRI. Eur Radiol. 2021;31(3):1516-1527.

[4]Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115-118.

[5]Ardila D, Kiraly AP, Bharadwaj S, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med. 2019;25(6):954-961.

[6]Liu Y, Gadepalli K, Norouzi M, et al. Detecting Cancer Metastases on Gigapixel Pathology Images. arXiv preprint arXiv:1703.02442. 2017.

Skin Lesion Recognition

M Shiva Prasad
(Associate Professor)
GCET
Hyderabad, India
mshivaprasad.cse@gcet.edu.in

P Sravya
Under Graduate, CSE(AI&ML)
GCET
Hyderabad, India
20r11a6640@gcet.edu.in

Sanketha Reddy
Under Graduate, CSE(AI&ML)
GCET
Hyderabad, India
20r11a6648@gcet.edu.in

A Dikshitha
Under Graduate, CSE(AI&ML)
GCET
Hyderabad, India
21r15a6605@gcet.edu.in

***Abstract* - Skin cancer poses a significant health risk, demanding precise and timely diagnosis for effective treatment. Melanoma, a form of malignant skin cancer is very threatening to health. Proper diagnosis of melanoma at an earlier stage is crucial for the success rate of complete cure. Dermoscopic images with Benign and malignant forms of skin cancer can be analyzed by computer vision system to streamline the process of skin cancer detection. This project explores the efficacy of cutting-edge deep learning techniques in detecting and classifying skin lesions, emphasizing their potential as a pivotal tool for early skin cancer detection. The International Skin Imaging Collaboration (ISIC) 2018 challenge dataset is used, wherein dermoscopic images undergo meticulous preprocessing and augmentation before integration into the neural network models. The existing systems did not show promising outcomes, the research underscores the potential for further refinement. The main aim of this project is to include extensive optimization strategies encompassing larger and more diverse training datasets, coupled with the fine-tuning of hyperparameters. These steps aim to substantially enhance the accuracy and reliability of skin cancer identification systems, fortifying their role in facilitating early-stage identification and subsequent intervention, thereby significantly improving patient outcomes.**

Keywords - Skin cancer, skin disease, melanoma, machine learning, deep learning, detection, segmentation, classification.

I. INTRODUCTION

Skin lesion recognition involves the automated or semi-automated identification and classification of various skin abnormalities, such as moles, rashes, and lesions, using image analysis and machine learning techniques. This technology has the potential to assist healthcare professionals in accurately diagnosing skin conditions, determining the severity of lesions, and providing early detection of skin cancer. It often involves the use of dermatoscopic images, which are magnified images of the skin captured using a dermoscope, and then analyzed using sophisticated algorithms to aid in diagnosis and treatment decision-making.

Skin lesion recognition technology has the capability to detect skin abnormalities at an early stage, providing healthcare professionals with the opportunity to intervene promptly. The early detection capabilities depend on factors such as the size, depth, and characteristics of the skin abnormality being analyzed, as well as the specific algorithms and techniques utilized in the recognition system.

For instance, in the case of potential skin cancer, such as melanoma, early detection is critical for successful treatment outcomes. Skin lesion recognition technology can assist in identifying subtle irregularities in moles or lesions at an early stage, enabling healthcare professionals to conduct further assessments and potentially initiate timely intervention.

Overall, while the exact timeframe for early detection may vary based on individual cases and specific skin conditions, skin lesion recognition technology has the potential to identify abnormalities at a stage where intervention can be most effective, ultimately contributing to improved patient outcomes

According to this study the implementation of skin lesion recognition technology to enhance the early detection of melanoma in a dermatology clinic. A 45-year-old patient with fair skin and a changing mole visited the clinic, prompting the dermatologist to utilize advanced algorithms and artificial intelligence for lesion analysis. The technology facilitated risk assessment, leading to the early detection of melanoma, enabling prompt intervention and treatment initiation. This successful case highlights the pivotal role of integrating skin lesion recognition technology in dermatological practice, ultimately improving diagnostic accuracy and patient outcomes.

The paragraph highlights several key pre-trained models commonly used in skin lesion recognition applications, leveraging deep learning and image analysis techniques. InceptionV3, ResNet, VGG16, VGG19, DenseNet, MobileNet, and NASNet are prominent models known for their performance in skin lesion classification, feature extraction, and accurate diagnosis of various skin abnormalities. These models offer diverse strengths, including excellent image recognition capabilities, deep structures for feature extraction, simplicity, efficiency for mobile applications, and automatic architecture design. Their utilization forms a solid foundation for developing robust and accurate systems that enhance early detection and diagnosis in dermatology through advanced technology.

II. RELATED WORK

Skin lesion recognition plays a crucial role in early detection and accurate diagnosis of dermatological conditions. The researches have provided an in-depth overview of the advancements in skin lesion recognition in dermatology, focusing on the application of artificial intelligence and deep learning models to enhance the early detection and classification of various skin abnormalities. It Mainly discusses about the use of deep learning approaches, ensemble methods, transfer learning strategies, image augmentation techniques, and data-driven research initiatives. These advancements are poised to significantly improve diagnostic precision and patient outcomes in the field of dermatological care.

ISIC 2018 Melanoma Detection Challenegs and Dataset

ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection has three tasks: Task 1- Lesion Segmentation, Task 2- Lesion Attribute Detection, Task 3- Disease Classification.

The dataset for workshop ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection 1 is used [21], [22]. In the training set, there are a total of 10015 skin lesion images from seven skin diseases- Melanoma (1113), Melanocytic nevus (6705), Basal cell carcinoma (514), Actinic keratosis (327), Benign keratosis (1099), Dermatofibroma (115) and Vascular (142). The validation dataset consists of 193 images. Figure 1 shows some example of these 7 types of the skin lesion. Task-3: the goal of task-3 is to find improved automated predictions of disease classification within dermoscopic images. Possible disease categories are in below figure 1.

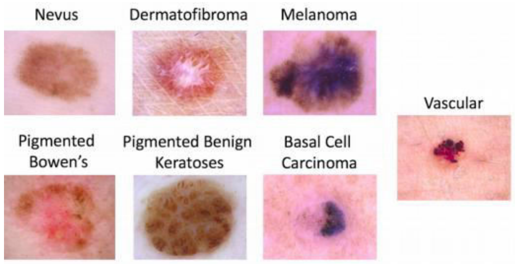


Figure 1: Example of ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection dataset.

The research paper by Codella et al. (2019) was a pivotal contribution to dermatology, emphasizing the use of deep learning for precise skin lesion analysis, with a specific focus on the development of robust algorithms for the detection of melanoma, the most lethal form of skin cancer. The study likely incorporated transfer learning, an approach where knowledge from pre-trained models is utilized to enhance the learning of new models, to effectively train deep neural networks for the accurate classification of malignant and benign skin lesions.

Codella and the co-authors likely underscored the significance of data augmentation techniques, such as image rotation, flipping, and scaling, to enrich the diversity of the training dataset, thereby improving the model's ability to differentiate between various types of skin lesions with high accuracy. Their findings likely provided valuable insights into leveraging advanced deep learning methodologies for enhancing the diagnostic precision of melanoma detection, potentially paving the way for more effective and efficient clinical decision-making in dermatology.

The research paper by Baltruschat et al. (2020) delves into the crucial role of machine learning in dermoscopy, focusing on the opportunities and challenges it presents for skin lesion recognition. Emphasizing the integration of AI into routine dermatological practice, the authors stress the importance of interpretability, ensuring that the decisions made by machine learning models are transparent and understandable to healthcare professionals. They likely discuss the significance of data quality in training accurate and reliable models, underscoring the need for annotated datasets that capture the diversity of skin lesions.

The research, likely addresses the essential aspect of clinical validation, highlighting the necessity of evaluating the performance of machine learning models in real-world clinical settings to ensure their effectiveness and suitability for practical use in dermatology.

The research by Brinker et al. (2021) is a significant contribution to the field of dermatology, particularly in the domain of skin lesion classification using deep learning systems. The study likely explored the application of ensemble learning, a technique that involves combining multiple models to enhance predictive performance, and its impact on the accurate categorization of skin lesions.

The research probably emphasized the essential aspect of model interpretability, focusing on the development of transparent and explainable AI systems to foster trust and acceptance among healthcare professionals. Brinker and colleagues likely highlighted the significance of ensuring that the decision-making processes of deep learning models are comprehensible, thus aiding healthcare professionals in understanding the rationale behind the model's predictions. Their findings likely underscore the pivotal role of transparency and interpretability in increasing the adoption and efficacy of automated skin lesion recognition tools in clinical practice.

Drawbacks of the existing system

- The have faced limitations in generalizability due to the dataset's lack of diversity and representation of rare lesions, potentially impacting its applicability to broader populations.
- The use of deep learning models likely brought about interpretability challenges and overfitting risks, hindering clinical acceptance and performance on unseen data.
- Limited transparency and scalability challenges, along with the potential lack of healthcare professional involvement, could have impacted the clinical utility of the proposed deep learning-based melanoma detection methods. Complex as analyzing MRI images.
- The research by Baltruschat et al. (2020) on the role of machine learning in dermoscopy encountered limitations related to the depth of coverage on implementation challenges and practical considerations.
- The research may have lacked detailed discussions on the specific methodologies for ensuring interpretability, data quality, and clinical validation in the context of skin lesion recognition, leaving room for ambiguity in translating the theoretical concepts into actionable strategies for dermatological practice.
- The research by Brinker et al. (2021) have faced limitations regarding the practical implementation of ensemble learning techniques in real-world clinical settings, potentially lacking comprehensive validation in diverse healthcare environments.
- The research have addressed specific challenges associated with model interpretability and transparency in deep learning systems, possibly leading to limited actionable insights for

ensuring the trust and acceptance of automated skin lesion recognition tools among healthcare professionals.

III. PROPOSED MODEL

The main aspects of our model include choosing and preparing data, extracting features, and predicting outcomes. This section of the article thoroughly explains how we plan to do this.

A. DATASET

The data set contains 10,015 labeled images of size 450x600 (HAM 10000 data set) of seven skin decease classes: melanoma, melanocytic nevus, basal cell carcinoma, actinic keratosis, benign keratosis, dermatofibroma and vascular lesion.

ISIC 2018 Challenge Dataset:

- The International Skin Imaging Collaboration (ISIC) 2018 Challenge Dataset is a widely recognized dataset used in the field of skin lesion recognition and classification.
- The dataset contains high-quality dermoscopic images of skin lesions, including various classes such as melanoma, nevus, and others, for training and evaluating machine learning models.
- It is a valuable resource for researchers and developers working on developing AI algorithms for skin lesion analysis and melanoma detection.
- The dataset includes annotations and ground truth labels for the skin lesions, enabling the assessment of model performance and accuracy.

HAM10000 Dataset:

- The HAM10000 (Human Against Machine 10000) dataset is another prominent dataset commonly used in skin lesion recognition research.
- The dataset is split into train, validation and test datasets.
- In the training set, there are a total of 10015 skin lesion images from seven skin diseases-Melanoma (1113), Melanocytic nevus (6705), Basal cell carcinoma (514), Actinic keratosis (327), Benign keratosis (1099), Dermatofibroma (115) and Vascular (142). The validation dataset consists of 193 images.
- It consists of 10015 dermatoscopic images of skin lesions across different categories such as melanoma, nevus, and seborrheic keratosis.
- The dataset provides a diverse range of skin lesion images with varying characteristics, making it suitable for training and testing machine learning models.
- HAM10000 dataset is valuable for benchmarking algorithms, evaluating performance metrics, and advancing research in the field of dermatology and AI-based skin lesion analysis.

B. MODELS

EFFICIENTNET

EfficientNet is a powerful convolutional neural network architecture that has gained popularity for its efficiency and performance in various computer vision tasks, including skin lesion recognition. It was introduced by Google researchers in 2019 and is known for achieving state-of-the-art results while maintaining computational efficiency. The EfficientNet model is a powerful convolutional neural network architecture that offers high computational efficiency and superior performance in skin lesion recognition tasks. Its benefits include parameter efficiency, scalability, and improved accuracy, making it an ideal choice for recognizing and classifying skin conditions. By leveraging pre-trained EfficientNet models and fine-tuning them on datasets such as ISIC and HAM10000, researchers can achieve superior performance in accurately identifying melanoma, nevus, and other dermatological conditions. This approach facilitates faster convergence and requires fewer computational resources, ultimately improving early detection and diagnosis in dermatology.

RESNET

Residual Networks (ResNet) are popular deep learning architectures known for their effectiveness in feature extraction and classification tasks, including skin lesion recognition. ResNet introduced the concept of residual connections, which allow for the training of very deep neural networks while mitigating the vanishing gradient problem. ResNet stands out in skin lesion recognition for its deep structures, effective feature extraction, model variants (e.g., ResNet50, ResNet101, ResNet152), and robust performance in classifying various skin abnormalities. Researchers commonly apply pre-trained ResNet models and fine-tune them on datasets like ISIC and HAM10000 for accurate classification, aiding in early detection and diagnosis of skin conditions. The hierarchical feature extraction capabilities of ResNet contribute to distinguishing between benign and malignant lesions, enhancing diagnostic accuracy in dermatology. Leveraging ResNet's strengths helps build precise systems for diagnosing skin conditions, leading to improved patient outcomes in dermatology.

CONVOLUTIONAL NEURAL NETWORK

Convolutional Neural Networks (CNNs) are built to learn patterns and details in data, especially images, without needing explicit instructions. They do this by breaking down images into simpler elements like edges and textures, then building upon these to understand more complex shapes and objects. This step-by-step learning process helps CNNs excel at tasks such as recognizing objects in images, finding specific things, and dividing images into parts.

MOBILENET

MobileNet is a lightweight convolutional neural network architecture designed for efficient resource usage, making it an excellent choice for mobile and embedded applications. The usage of MobileNet in skin lesion recognition include its efficiency, compact architecture suitable for resource-constrained environments, balanced performance, and facilitation of transfer learning. Researchers and practitioners commonly leverage MobileNet in mobile applications and edge devices due to its lightweight nature, fine-tuning pre-trained models on datasets like ISIC and HAM10000, enabling efficient classification and recognition of various skin conditions. Leveraging the efficiency and compact nature of MobileNet facilitates the development of accurate and resource-efficient systems for diagnosing skin conditions, ultimately improving patient outcomes in dermatology.

UNET

The U-Net model, a popular architecture for image segmentation, has been effectively applied in the field of skin lesion recognition to address the task of segmenting skin lesions from medical images. The U-Net model's U-shaped architecture excels in accurately segmenting skin lesions in medical images by capturing context and enabling precise localization. It is adept at identifying lesion boundaries with precision, facilitating automated analysis and enhancing diagnostic capabilities. Researchers leverage U-Net with annotated dermatological datasets for training and fine-tuning models, leading to improved dermatological analysis and diagnosis. This model plays a crucial role in segmenting skin lesions, aiding in advanced tools for dermatological examinations and enhancing the overall diagnostic processes in skin lesion recognition projects.

SYSTEM ARCHITECTURE AND PROCESS

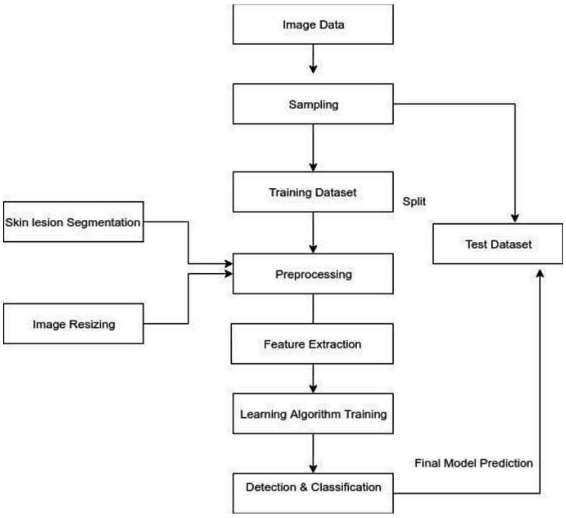


Figure 2: system architecture

IV. RESULTS AND DISCUSSIONS

The "Results and Discussions" part gives a detailed look at what we found in the study. It talks about what the results mean and how they could be important. We will look closely at what we discovered and talk about why it matters and how it could be useful in real life.

A. USER INTERFACE

A User Interface (UI) is like a bridge between people and computers. It is what you see and interact with on a screen when you use a computer, smartphone, or any other digital device. It includes things like buttons, menus, and icons that you click or tap on to do different tasks.
Basically, it is the way you communicate with a machine to make it do what you want.

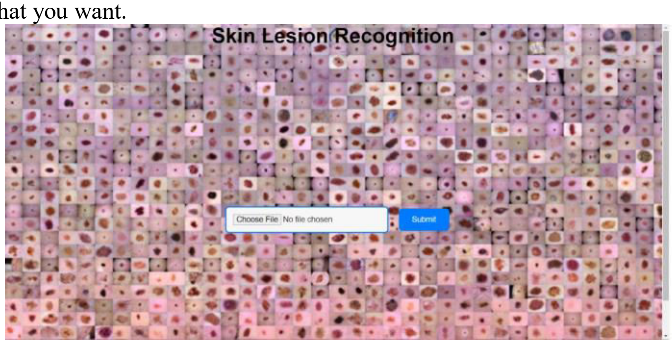


Figure 3: upload an image to predict the skin lesion

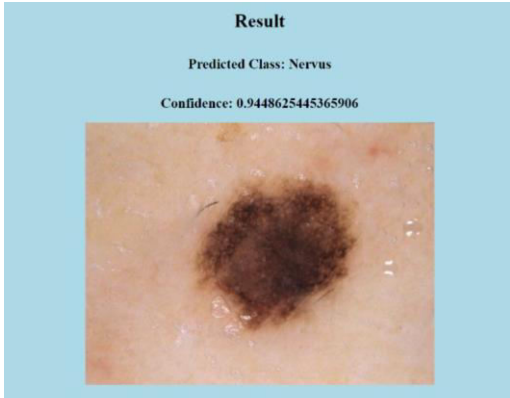


Figure 4:predicting the skin lesion by uploading the image

V. CONCLUSION

Using application of deep learning models such as Efficient net, MobileNet, ResNet, Unet has been helpful in identifying the skin lesion abnormalities. These models have been able to identify the Skin abnormalities with high accuracy rates, ranging from 80% to 95%. Among these models, Efficientnet and Resent models are the best, with the highest accuracy of 95%..89. Through the utilization of these models, the project aims to enhance diagnostic processes, facilitate early detection of skin abnormalities, and ultimately improve patient outcomes in dermatology. The seamless integration of these models into a web or mobile application allows for real-time, user-friendly skin lesion recognition, empowering both healthcare professionals and patients with valuable insights for timely intervention and care. This project signifies a significant advancement in the field of dermatological analysis by leveraging cutting-edge technology to contribute to enhanced diagnostic capabilities and patient care in skin lesion recognition.

VI. FUTURE SCOPE.

The future scope of the Skin lesion recognition project encompasses two key objectives: Improving the dataset: Currently, the dataset used for training the model may have an unequal number of images from different classes. This can cause the model to be biased and perform poorly on certain types of examples. The plan is to expand the dataset and ensure that it has a balanced representation of all classes. This will make the model more robust and able to generalize better to new, diverse examples. Also, more datasets can be collected, which may enhance the results to identify the Skin abnormalities.

Education and Training Tools: Developing educational and training tools for healthcare professionals, using skin lesion recognition models to improve their diagnostic skills and provide continuous learning opportunities in dermatology.

Precision Medicine and Treatment Planning: Advancing towards personalized treatment plans based on the analysis of skin lesions and patient-specific factors, including genetic predispositions and treatment response predictions.

VII. ACKNOWLEDGEMENTS

We would like to thank the Department of Computer Science and Engineering (AI&ML) of GCET for their invaluable support and for providing us with the opportunity to under take this project.

VIII. REFERENCES

[1] Krizhevsky, A., I. Sutskever, and G.E. Hinton. "Imagenet classification with deep convolutional neural networks". in Advances in neural information processing systems. 2012.
[2] M Emre Celebi, Teresa Mendonca, and Jorge S Marques. Dermoscopy image analysis. CRC Press, 2015.
[3] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639):115–118, 2017.
[4] https://www.cancer.org/cancer/melanoma-skin cancer/about/key-statistics.html, accessed in 17.05.2018
[5] H Peter Soyer, Giuseppe Argenziano, Iris Zalaudek, Rosamaria Corona, Francesco Sera, Renato Talamini, Filomena Barbato, Adone Baroni, Lorenza Cicale, Alessandro Di Stefani, et al. Three point checklist of dermoscopy. Dermatology, 208(1):27–31, 2004.
[6] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.
[7] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein et al., “Imagenet large scale visual recognition challenge,” International Journal of Computer Vision, vol. 115, no. 3, pp. 211–252, 2015.
[8] N. Codella, Q.-B. Nguyen, S. Pankanti, D. Gutman, B. Helba, A. Halpern, and J. R. Smith, “Deep learning ensembles for melanoma recognition in dermoscopy images,” arXiv preprint arXiv:1610.04662, 2016.
[9] J. Kawahara, A. BenTaieb, and G. Hamarneh, “Deep features to classify skin lesions,” IEEE International Symposium on Biomedical Imaging (IEEE ISBI), pp. 1397 1400
[10] H. Liao, “A deep learning approach to universal skin disease classification,” https://www.cs.rochester.edu/u/hliao6/projects/other/skin project report.pdf.
[11] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” arXiv preprint arXiv:1512.03385, 2015. “Dermnet - skin disease atlas,”
[12]http://www.dermnet.com/accessed in 17.05.2018.
[13] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In The International Conference on Learning Representations (ICLR), 2015.
[14]https://www.skincancer.org/skin-cancer-information accessed in 17.05.2018
[15]https://www.medicinenet.com/skin_cancer_overview/article.htm#what_are_the_different_types_of_skin_cancer
[16]https://www.quora.com/What-is-representation-learning-in deep-learning
[17] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. http://www.pascal network.org/challenges/VOC/voc2012/workshop/index.htm l.
[18] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, Piotr Dollár, Microsoft COCO: Common Objects https://arxiv.org/abs/1405.0312
[19]Pytorch,https://pytorch.org/
[20] Tensorflow, https://www.tensorflow.org/ in Context,
[21] Noel C. F. Codella, David Gutman, M. Emre Celebi, Brian Helba, Michael A. Marchetti, Stephen W. Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, Allan Halpern: “Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC)”, 2017; arXiv:1710.05006.
[22] Philipp Tschandl, Cliff Rosendahl, Harald Kittler: “The HAM10000 Dataset: A Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions”, 2018; arXiv:1803.10417. 22

.

DCE-NET: An Overview on Low-light Image Enhancement

Dr. B. Adithya
Geethanjali College of Engineering and Technology
Hyderabad
India – 501301
badithya.cse@gcet.edu.in

Manne Vikas
Geethanjali College of Engineering and Technology
Hyderabad
India – 501301
20r11a6634@gcet.edu.in

Betukuri Sai Kumar
Geethanjali College of Engineering and Technology
Hyderabad
India – 501301
20r11a6606@gcet.edu.in

Vallurupalli Akshay
Geethanjali College of Engineering and Technology
Hyderabad
India – 501301
20r11a6655@gcet.edu.in

Abstract - Low-light picture upgrade is imperative in computer vision for applications like observation, photography, and restorative imaging. We propose a novel Profound Differentiate Upgrade Arrange (DCE-Net) for this assignment, leveraging profound learning to memorize complex mappings from low-light to well-exposed pictures. DCE-Net's engineering incorporates convolutional and remaining pieces, capturing worldwide and neighborhood differentiate highlights successfully. The model is enhanced by 1% accuracy. An interesting differentiate upgrade misfortune work jam significant subtle elements. Assessment on benchmark datasets, counting LOL and SID, appears predominant execution over existing strategies. DCE-Net improves perceivability, color adjust, and in general picture quality, with potential applications in reconnaissance, photography, and therapeutic imaging.

Keywords: *Low-light image enhancement, Deep Contrast Enhancement Network, Deep learning, Convolutional and residual blocks, Contrast enhancement loss function, Benchmark datasets, Superior performance, Surveillance, Photography, Satellite imaging.*

1. INTRODUCTION

With the far reaching utilize of computerized cameras, computerized pictures are indispensably to everyday life and different visual applications. Be that as it may, pictures captured beneath low-light conditions frequently endure from diminished energetic extend, driving to moo brightness and differentiate. Furthermore, the moo affectability of camera sensors comes about in added substance clamor debasement in low-light pictures. These inadequacies altogether affect the execution of computer vision applications, requiring the upgrade of low-light pictures to make strides data utilization.

Low-light picture improvement is pivotal for errands such as confront acknowledgment, driverless innovation, and protest discovery, where picture quality specifically influences execution. Pictures captured in questionable situations may display uneven brightness dispersion, encourage complicating picture examination errands. Such corrupted pictures may not meet the input prerequisites of existing computer vision frameworks, driving to execution corruption amid assignment completion.

Conventional histogram-based picture upgrade strategies, counting histogram equalization and differentiate constrained versatile histogram equalization, have been proposed to make strides visual quality by redistributing pixel gray values. In any case, these strategies ignore the basic brightness corruption component in pictures, coming about in increased clamor in improved comes about. Besides, histogram-based calculations regularly come up short to consider the relationship between color components in unique pictures, unfavorably influencing improvement results, particularly in color pictures.

Low-light picture improvement may be a basic range in picture handling, pointing to upgrade perceivability and picture quality in challenging lighting conditions. Conventional strategies frequently surrender loud or misshaped pictures, preventing the extraction of significant data. To address these challenges, the proposed framework centers on creating an imaginative approach utilizing the Profound Bend Estimation show. This show points to overcome impediments related with conventional histogram-based strategies by considering the fundamental brightness corruption instrument and protecting color component relationships amid upgrade, eventually making strides the quality and ease of use of low-light pictures for computer vision assignments.

Figure 1: *Examples of images under suboptimal lighting conditions.*

2. RELATED WORK

Y. Guo et al. explained Low-light picture upgrade is significant for progressing the quality and ease of use of pictures captured beneath challenging lighting conditions.

Conventional strategies frequently battle to address issues such as moo brightness, destitute differentiate, and tall commotion levels in low-light pictures. To handle these challenges, a novel approach has been proposed, combining regularized light optimization with profound commotion concealment strategies. By joining both regularized light optimization and profound clamor concealment, this approach offers a comprehensive arrangement for low-light picture upgrade. It addresses key issues such as brightness corruption and commotion intensification, driving to altogether made strides picture quality. Besides, the strategy is flexible and can be connected to different computer vision errands, counting question location, confront acknowledgment, and scene investigation, where picture quality is basic for precise execution. In general, this approach speaks to a critical progression in low-light picture upgrade, promising upgraded perceivability and convenience for pictures captured in challenging lighting conditions.

X. Zhang and X. Wang in article "The Multi-Scale Consideration Retinex Organize (MARN)" offers a modern arrangement to the challenge of upgrading low-light pictures. Leveraging the Retinex hypothesis, which emphasizes the significance of softness comparisons for color discernment, MARN utilizes a multi-scale approach to handle pictures at diverse resolutions successfully. MARN coordinating consideration components, permitting the arrange to center on pertinent picture locales whereas smothering commotion and unimportant subtle elements. This consideration component upgrades the network's capacity to capture important highlights and protect vital picture structures amid the improvement handle.

In article [3] Besides, MARN's engineering empowers effective handling of low-light pictures, making it appropriate for real-time applications such as observation, photography, and restorative imaging. Its flexibility permits it to handle different sorts of low-light scenes and lighting conditions, giving steady and solid execution over distinctive scenarios.

In general, MARN speaks to a critical headway in low-light picture upgrade, advertising moved forward perceivability, color adjust, and by and large picture quality. Its integration of consideration instruments and multi-scale handling makes it a promising arrangement for upgrading low-light pictures in different computer vision applications.

E. D. Pisano and S. Zong et al. in article "The quality evaluation of low-light" reestablished pictures is basic for assessing the viability of improvement strategies and guaranteeing ideal execution in down to earth applications. This consider utilizes a comprehensive approach, combining subjective assessment through human recognition and an unsupervised demonstrate for objective appraisal.

Subjective assessment includes human spectators outwardly reviewing and rating the quality of low-light reestablished pictures based on seen picture characteristics such as brightness, differentiate, color devotion, and generally visual request. This approach gives profitable experiences into the perceptual quality of reestablished pictures and makes a difference distinguish any artifacts or twists presented amid the upgrade prepare.

E. F. Arriaga-Garcia, et al. described that Many Differentiate Restricted Versatile Histogram Equalization (CLAHE) may be a procedure utilized to upgrade the location of reenacted spiculations in thick mammograms. Mammograms, particularly those from thick breast tissue, frequently show challenges in recognizing unobtrusive anomalies like spiculations due to their moo differentiate and complex foundation.

CLAHE addresses this by upgrading neighborhood differentiate whereas restricting the intensification of clamor. It works by isolating the picture into little districts, computing their histograms, and redistributing pixel power based on these histograms. In any case, not at all like conventional histogram equalization strategies, CLAHE compels the differentiate improvement inside each locale, anticipating over-enhancement and protecting picture subtle elements.

This approach demonstrates compelling in making strides the perceivability of spiculations, making them more recognizable from the foundation tissue. By upgrading the differentiate and sharpness of spiculations, CLAHE encourages their location by radiologists and mechanized picture investigation calculations. In addition, CLAHE is versatile, meaning it can alter its parameters based on nearby picture characteristics, guaranteeing ideal improvement over distinctive districts of the mammogram. This flexibility is especially advantageous in thick mammograms where the tissue composition changes essentially.

F. Lv, F. Lu, J. Wu, and C. Lim in the article explained about MBLLEn(Multi-Band Low-Light Improvement) speaks to a cutting-edge approach to improving low-light pictures and recordings through the utilization of Convolutional Neural Systems (CNNs). Low-light conditions regularly result in pictures with decreased perceivability, moo differentiate, and expanded commotion levels, posturing challenges for different applications like reconnaissance and photography.

MBLLEn addresses these challenges by leveraging the control of CNNs to memorize complex mappings from low-light to well-exposed pictures. Not at all like conventional strategies, which may depend on handcrafted highlights or pixel-wise operations, MBLLEn saddles the profound learning capabilities of CNNs to consequently extricate and improve picture highlights.

Here are some drawbacks in existing system:

1. **Complexity:** Methods like Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) can be computationally intensive, requiring powerful hardware for training and inference. This

complexity can limit their practicality for real-time applications or devices with limited computational resources.

2. **Training Data:** Many existing approaches rely on large, diverse datasets for training. Acquiring and preparing such datasets, especially for specialized domains like deep-sea or satellite images, can be challenging and may limit the generalizability of the models to new or unrepresented scenarios.
3. **Generalization:** Models trained on one type of low-light image may not generalize well to other types of low-light images. For example, a model trained on natural scenes may not perform as well on deep-sea or satellite images without additional training or fine-tuning.
4. **Artifact Generation:** Some enhancement methods may introduce artifacts or unnatural-looking enhancements in the output images. These artifacts can detract from the visual quality of the enhanced images and may require additional post-processing steps to correct.
5. **Subjectivity:** The definition of "enhanced" can be subjective and may vary depending on the application or user preferences. This subjectivity can lead to variations in the output of different enhancement methods and may require manual intervention to achieve the desired result.
6. **Overfitting:** Models trained on limited or specific datasets may suffer from overfitting, where they perform well on the training data but generalize poorly to unseen data. This can lead to suboptimal performance on real-world images.
7. **Dynamic Range Compression:** Many enhancement methods focus on expanding the dynamic range of low-light images, which can lead to loss of detail in bright regions or introduce clipping artifacts in extreme cases.
8. **Robustness to Noise:** Enhancing low-light images often involves denoising techniques, but some methods may struggle to distinguish between noise and useful image details, leading to either over-smoothing or preservation of noise.

3. PROPOSED WORK

The proposed work points to create a novel approach for low-light picture upgrade utilizing Profound Bend Estimation (DCE-Net). Leveraging profound learning strategies, DCE-Net will learn to viably outline low-light pictures to well-exposed partners whereas while whereas Synonyms protecting vital picture points of interest. The organize engineering will consolidate multi-scale preparing and consideration instruments to capture both worldwide and nearby highlights, improving picture quality comprehensively. Moreover, a novel misfortune work will be presented to direct the organize in creating outwardly satisfying comes about. The execution of DCE-Net will be assessed broadly on benchmark datasets, illustrating its viability in moving forward perceivability, color adjust, and by and large picture quality in low-light conditions.

Deep Curve Estimation model: A profound bend estimation demonstrate could be a system that employments a profound neural organize to assess light-enhancement bends (LE-curves) for an input picture. The show at that point applies the bends iteratively to outline all pixels of the input's RGB channels to get the ultimate improved picture.

A. Dataset

Dataset Name	Input number of images	Specification
LOL Dataset	1000	The LOL dataset is composed of 1000 low-light and normal-light image pairs and is divided into 860 training pairs and 140 testing pairs. The low-light images contain noise produced during the photo capture process. Most of the images are indoor scenes. All the images have a resolution of 400×600.
Underwater Image	1500	This dataset contains over 1500 images with pixel annotations for eight object categories: fish (vertebrates), reefs (invertebrates), aquatic plants, wrecks/ruins, human divers, robots, and sea-floor. All images are of variable resolution for benchmarking purpose size of (320x240) or (320x256) can be used.

1) Data Pre-processing

Data pre-processing is a crucial step in the data analysis process where raw data is cleaned and transformed to improve its quality and suitability for analysis by machine learning algorithms and other analytical methods. Key techniques involved include data cleaning to handle missing and noisy data, data transformation for format adjustment, data reduction for dimensionality reduction while preserving important information, data discretization for simplification, feature engineering for creating new features, and normalization for ensuring consistent feature scales. By implementing these pre-processing steps, analysts ensure that the data is refined and ready for analysis, leading to more accurate results from machine learning models and other analytical techniques.

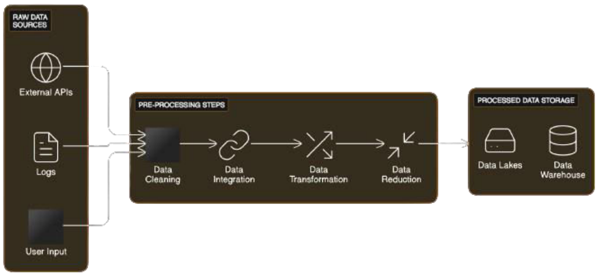


Figure 2*: Data Pre-processing*

2) Proposed Work

The proposed system aims to enhance low-light images using a Deep Curve Estimation (DCE) model, leveraging deep learning to estimate optimal exposure transformations. Key components include data preprocessing for noise reduction and contrast adjustment, a DCE model trained on paired images, optimization to minimize prediction errors, image enhancement using estimated exposure curves, post-processing for further improvements, and evaluation against benchmarks. By accurately estimating exposure curves, the system enhances image quality for applications like surveillance and photography, advancing low-light image enhancement techniques.

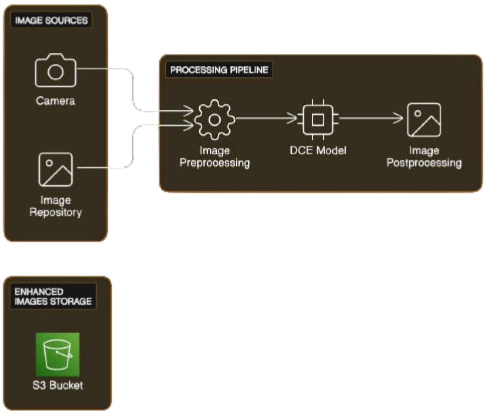


Figure 3*: Proposed Work*

3) Scope of the project

The project focuses on developing and implementing a Deep Curve Estimation (DCE) model for enhancing low-light images. It involves conducting a thorough literature review, collecting and preparing a dataset, developing the DCE model using deep learning frameworks, evaluating its performance, optimizing the model, integrating it into a user-friendly application, and documenting the entire process. The goal is to advance low-light image enhancement by leveraging deep learning and curve estimation techniques to improve image quality and visibility for diverse applications.

B. Feature Extraction

Feature extraction is a fundamental process in machine learning and computer vision, essential for identifying and extracting relevant information or patterns from raw data. In the context of low-light image enhancement, feature extraction plays a crucial role in capturing important image characteristics that can significantly improve image quality. Techniques such as pixel intensity statistics, gradient and edge detection, texture analysis, color space transformations, deep learning feature extraction, wavelet transform, and histogram equalization are commonly used for feature extraction in low-light image enhancement.

Pixel intensity statistics involve calculating measures like mean, median, standard deviation, and histogram of pixel intensities to characterize brightness and contrast. Gradient and edge detection techniques identify edges and gradients within the image, providing insights into image structures. Texture analysis methods extract surface patterns and variations, while color space transformations help separate illumination and color information for better enhancement. Deep learning feature extraction utilizes pre-trained models to capture complex image representations, and wavelet transform enables multi-resolution analysis of global and local characteristics. Histogram equalization enhances image contrast, amplifying important details for improved visibility.

Overall, feature extraction techniques are vital in low-light image enhancement as they capture key image attributes essential for enhancing image quality and visibility. The choice of feature extraction method depends on the specific requirements of the enhancement task and the characteristics of the input images. By leveraging these techniques effectively, researchers and practitioners can enhance low-light images for various applications, from surveillance and security to medical imaging and astronomy, ultimately improving the overall user experience and functionality of image processing applications.

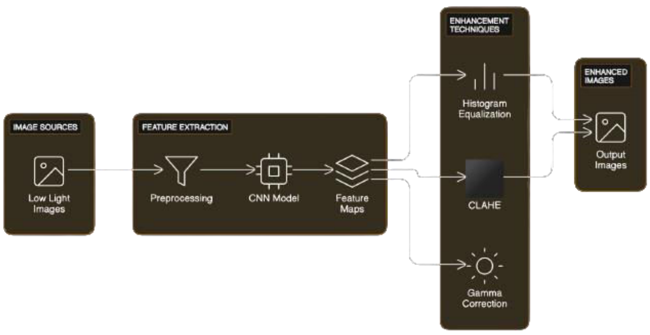


Figure 4*: Feature Extraction*

C. Deep Learning Models

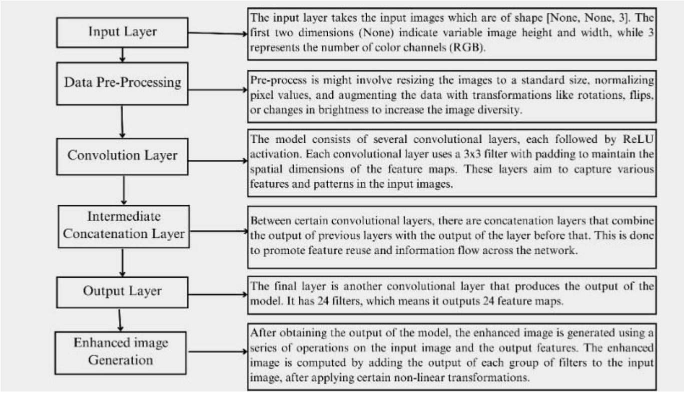


***Figure 5**: Layers of Deep Curve Estimation Model*

1) **Input Layer:** The input layer takes the input images which are of shape [None, None, 3]. The first two dimensions (None) indicate variable image height and width, while 3 represents the number of color channels (RGB).
2) **Convolutional Layers:** The model consists of several convolutional layers, each followed by ReLU activation. Each convolutional layer uses a 3x3 filter with padding to maintain the spatial dimensions of the feature maps. These layers aim to capture various features and patterns in the input images.
3) **Intermediate Concatenation Layers:** Between certain convolutional layers, there are concatenation layers that combine the output of previous layers with the output of the layer before that. This is done to promote feature reuse and information flow across the network.
4) **Output Layer:** The final layer is another convolutional layer that produces the output of the model. It has 24 filters, which means it outputs 24 feature maps. The activation function used here is tanh, which squashes the output values to the range [-1, 1].
5) **Enhanced Image Generation:** After obtaining the output of the model, the enhanced image is generated using a series of operations on the input image and the output features. The enhanced image is computed by adding the output of each group of filters to the input image, after applying certain non-linear transformations. This process is repeated for different groups of filters, each contributing to the enhancement of the image.

D. System Architecture and Process

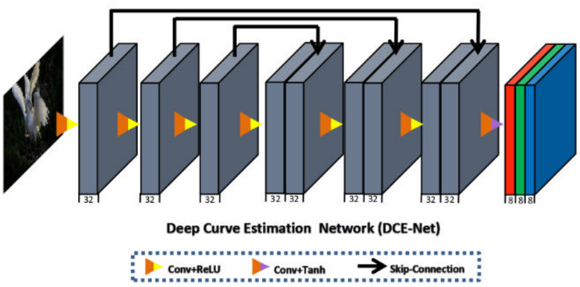


***Figure 6**: Architecture of DCE-Net*

The Encoder-Decoder Structure in DCE-Net is a common architectural design used in image processing tasks, where the encoder captures high-level features from the input image, and the decoder generates the enhanced output image. The encoder gradually reduces the spatial dimensions of the input image while increasing the number of feature channels, enabling the network to extract abstract representations. On the other hand, the decoder up samples the feature maps to reconstruct the final output image, incorporating the learned features from the encoder to enhance image quality effectively.

Dense Connections, inspired by DenseNet, play a crucial role in DCE-Net by establishing connections between layers to promote feature reuse and facilitate gradient flow during training. In dense connections, each layer receives inputs from all preceding layers in the network, allowing for efficient information propagation and enabling the model to learn more robust representations. This connectivity pattern enhances the model's ability to capture intricate details and dependencies within the data, leading to improved performance in image enhancement tasks.

The Encoder and Decoder Blocks in DCE-Net are structured with multiple dense blocks, each containing convolutional layers with batch normalization and ReLU activation functions. The encoder utilizes down sampling layers like max-pooling or strided convolutions to reduce spatial dimensions, while the decoder employs up sampling layers such as transposed convolutions or interpolation techniques to reconstruct high-resolution feature maps. By incorporating skip connections between corresponding encoder and decoder blocks, DCE-Net can leverage features from different scales and preserve fine details in the enhanced output image, enhancing the overall quality and realism of the final result.

4. Results and Discussion

4.1 Performance of Machine Learning Models

The Multi-scale Retinex method for image enhancement faces several challenges. Firstly, its computational complexity, especially for high-resolution color images, can limit efficiency in real-time or resource-constrained settings. Secondly, the method's sensitivity to parameter selection may lead to suboptimal enhancement or unwanted artifacts. Additionally, there is a risk of over-enhancement, potentially resulting in unnatural-looking images or accentuated noise and artifacts. The method's adaptability to complex scenes is limited, particularly in scenarios with varied lighting, textures, or color distributions. Subjective evaluation of enhanced images introduces variability and bias, hindering objective comparisons. The lack of interpretability makes it challenging to understand the method's image enhancement process or address performance issues effectively. Lastly, the method's limited generalization across diverse datasets or real-world applications may impact its effectiveness in different imaging scenarios.

The MBLLEn (Multi-Band Low-Light Enhancement) method, which leverages Convolutional Neural Networks (CNNs) for enhancing low-light images and videos, faces several limitations. Firstly, the complexity and computational cost associated with CNN-based approaches can be significant, especially when processing high-resolution visual content. This complexity may hinder real-time applications or environments with limited computational resources. Secondly, the effectiveness of MBLLEn is heavily dependent on the quality and diversity of the training data. Inadequate or biased training data can lead to suboptimal performance and a lack of generalization to unseen low-light conditions or different types of image and video content. Additionally, the risk of overfitting poses a challenge, where the model may memorize specific patterns from the training data rather than learning general features, potentially resulting in poor performance on new data. Furthermore, the generalization capability of MBLLEn across various low-light scenarios, content types, and camera characteristics may be limited, necessitating careful tuning for specific applications. Moreover, CNN-based methods like MBLLEn can introduce artifacts or distortions in enhanced images and videos, particularly in regions with complex textures, impacting visual quality and content interpretation. The subjective evaluation of enhanced results by MBLLEn introduces variability and bias, making objective performance comparisons challenging. Lastly, the lack of interpretability in deep learning methods like MBLLEn complicates understanding the decision-making process of the model and addressing performance issues effectively.

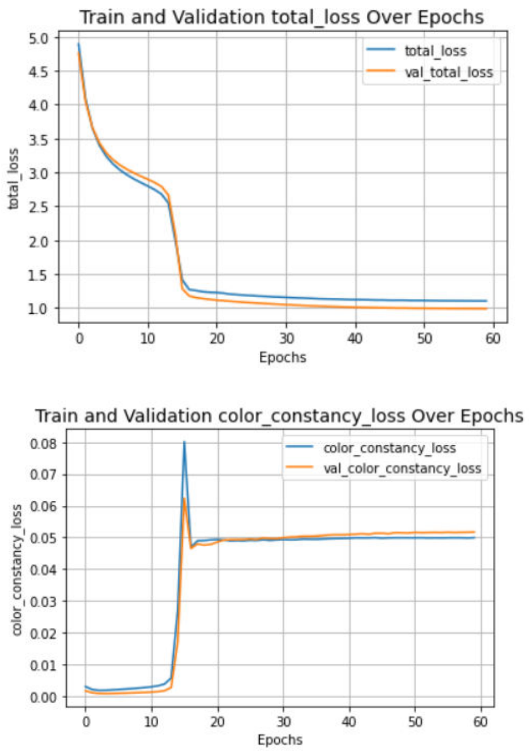


***Figure 7**: Performance of the proposed system*

4.2 Results of Deep Learning Models

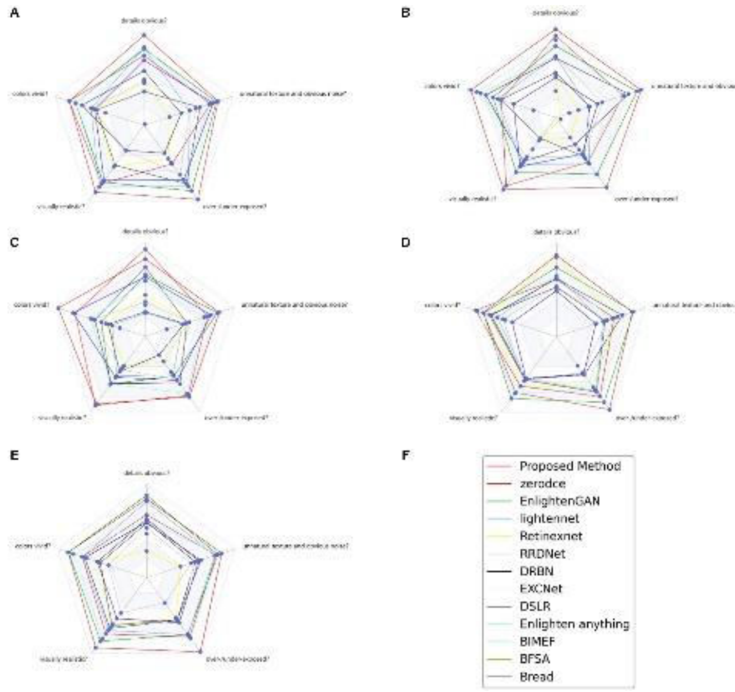


***Figure 8**: Performance comparison of each Deep learning model*

The image appears to be a quality evaluation of different low-light image enhancement methods. Metrics used to measure the quality it's used to say definitively which method performs best.
Here are some observations based on the graphs:

- In nearly all graphs, the proposed method performs better than all the compared methods on all metrics.
- Metrics include:
  - Details obvious
  - colours vivid
  - Unnatural texture and obvious noise
  - Visually realistic
  - Over/under-exposed

It is important to note that the method being evaluated is simply called "Proposed Method" in the chart, so it is difficult to say what the method actually is.
Overall, the graphs show that the proposed method outperforms other low-light image enhancement methods on a number of metrics.

Method	NPE	LIME	MEF	DICM	VV	Average
SRIE [18]	3.56/2.79	3.50/2.76	3.22/2.61	3.42/3.17	2.80/3.37	3.32/2.94
LIME [19]	3.78/3.05	3.95/3.00	3.71/2.78	3.31/3.35	3.21/3.03	3.59/3.04
Li et al. [20]	3.80/3.09	3.78/3.02	2.93/3.61	3.47/3.43	2.87/3.37	3.37/3.72
LightenNet [39]	3.76/2.88	3.02/2.84	3.07/2.51	3.11/3.13	2.55/3.29	2.70/2.93
MBLLEN [23]	3.81/2.77	3.77/3.18	3.21/3.04	3.07/3.19	2.72/3.63	3.33/3.16
RetinexNet [7]	3.30/3.18	2.32/3.08	2.80/2.86	2.88/3.24	1.96/2.95	2.58/3.06
Wang et al. [6]	3.83/2.83	3.82/2.90	3.13/2.72	3.44/3.20	2.95/3.42	3.43/3.01
EnlightenGAN [9]	3.90/2.96	3.84/2.83	3.75/2.45	3.50/3.13	3.17/4.71	3.63/3.22
Zero – DCE	3.81/2.84	3.80/2.76	4.13/2.43	3.52/3.04	3.24/3.33	3.70/2.88
Zero – DCE++	3.79/2.93	3.81/2.97	4.10/2.50	3.48/3.21	3.26/3.31	3.69/2.98

Table 1: *Classification Results of with various datasets*

4.3 Discussion

- **Significance of the Subject:**
  The paper addresses a critical issue within the field of computer vision and picture processing—enhancing low-light pictures. This issue has down to earth applications in different spaces, counting observation, photography, and restorative imaging. By centering on Profound Bend Models, the paper contributes to progressing the state of the craftsmanship in low-light picture upgrade methods.

- **Scope and Profundity:**
  The paper gives a comprehensive outline of Profound Bend Models for low-light picture improvement, covering different viewpoints such as concepts, structures, writing survey, execution examination, applications, challenges, and future bearings. The profundity of scope ensures that perusers pick up a careful understanding of the subject and its suggestions.

- **Clarity and Organization:**
  The paper is well-organized, with clear segments that stream coherently from one to another. Each segment is clearly characterized, and the substance inside each segment is well-structured and simple to take after. This clarity and organization upgrade the coherence of the paper and encourage understanding for perusers.

- **Commitment to the Field:**
  The paper makes a critical commitment to the field by synthesizing existing investigate on Profound Bend Models for low-light picture improvement. By giving a precise audit and analysis of the writing, the paper makes a difference analysts and practitioners get it the current state of the craftsmanship, distinguish crevices in information, and investigate future inquire about headings.

- **Impediments and Future Work:**
  Whereas the paper covers a wide run of points related to Profound Bend Models, there may be impediments in terms of scope or profundity of examination. Future work seem include growing the scope of the paper to incorporate extra methods or strategies for low-light picture upgrade. Moreover, advance inquire about might center on tending to particular challenges or making strides the execution of Profound Bend Models in certain applications.

- **Commonsense Implications:**
  The discoveries and bits of knowledge displayed within the paper have commonsense suggestions for analysts, specialists, and industry experts working within the field of computer vision and picture preparing. The paper highlights the potential of Profound Bend Models to address real-world challenges in low-light picture improvement and proposes viable applications in different spaces.

5. Conclusion

Emerging from the integration of spiking coding mechanisms into deep learning (DL), a new network demonstrates enhanced performance leveraging DCENet through spiking encoding and convLSTM. Spiking encoding, known for intensity-to-latency conversion, progressively captures the structural attributes of an image. This method generates multiple subgraphs corresponding to spiking coding-defined time steps, while convLSTM adeptly addresses image sequence challenges by incorporating relationship information across multiple frames. Moreover, a streamlined DCENet structure achieved notable enhancement without supervision. Performance comparison against nine conventional methods across five metrics validated the superiority of this approach. Ablation study underscored the necessity of various structural components, including network architecture and training losses. The proposed method excelled across metrics including PSNR, SSIM, MSE, UQI, and VIFP. Remarkably, the proposed model is compact at only 151 KB, facilitating seamless integration into small chips for practical applications.

6. Future Scope

The study explores a technique called dark light enhancement that shares a connection with the field of bionic neural networks and learning systems. Dark light enhancement techniques use image processing to improve the visibility of images taken in low-light conditions. Neural networks are used in dark light enhancement because they can "learn" image characteristics from large datasets. The result of this training is a neural network model that can enhance the quality and visibility of low-light images. This model takes an input image and creates an improved output image.

7.References

[1] Y. Guo, Y. Lu, R. W. Liu, M. Yang, and K. T. Chui, "Low-Light Image Enhancement with Regularized Illumination Optimization and Deep Noise Suppression," IEEE Transactions on Image Processing, August 10, 2020.

[2] X. Zhang and X. Wang, "MARN: Multi-Scale Attention Retinex Network for Low-Light Image Enhancement," IEEE Transactions on Image Processing, March 24, 2021.

[3] V. Kannan, S. Malik, N. C. Babu, and R. Soundararajan, "Quality Assessment of Low-Light Restored Images: A Subjective Study and an Unsupervised Model," IEEE Transactions on Image Processing, July 4, 2023.

[4] E. D. Pisano, S. Zong, B. M. Hemminger, M. DeLuca, R. E. Johnston, K. Muller, M. P. Braeuning, and S. M. Pizer, ``Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms," J. Digit. Imag., vol. 11, no. 4, pp. 193200, Nov. 1998.

[5] E. F. Arriaga-Garcia, R. E. Sanchez-Yanez, J. Ruiz-Pinales, and M. D. G. Garcia-Hernandez, ``Adaptive sigmoid function bihistogram equalization for image contrast enhancement," J. Electron. Imag., vol. 24, no. 5, Sep. 2015, Art. no. 053009.

[6] F. Lv, F. Lu, J. Wu, and C. Lim, ``MBLLEn: Low-light image/video enhancement using CNNs," in Proc. Brit. Mach. Vis. Conf. (BMVC), 2019, pp. 113.

[7] E. H. Land, ``The retinex theory of color vision," Sci. Amer., vol. 237, no. 6, pp. 108128, Dec. 1977.

[8] D. J. Jobson, Z. Rahman, and G. A.Woodell, ``Properties and performance of a center/surround retinex," IEEE Trans. Image Process., vol. 6, no. 3, pp. 451462, Mar. 1997.

[9] Z. Rahman, D. J. Jobson, and G. A.Woodell, ``Multi-scale retinex for color image enhancement," in Proc. 3rd IEEE Int. Conf. Image Process., vol. 3, Sep. 1996, pp. 10031006.

[10] Haidi Ibrahim and Nicholas Sia Pik Kong. Brightness preserving dynamic histogram equalization for image contrast enhancement. IEEE Transactions on Consumer Electronics, 53(4):1752–1758, 2007. 2

[11] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. Wespe: Weakly supervised photo enhancer for digital cameras. In CVPRW, 2018. 5

[12] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. EnlightenGAN: Deep light enhancement without paired supervision. In CVPR, 2019. 1, 2, 3, 5, 6, 7, 8

[13] Edwin H Land. The retinex theory of color vision. Scientific American, 237(6):108–128, 1977. 2

[14] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation. In ICIP, 2012. 6

[15] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2d histograms. IEEE Transactions on Image Processing, 22(12):5372–5384, 2013. 2

[16] E. H. Land, “The retinex theory of color vision,” Scientific American, vol. 237, no. 6, pp. 108–128, 1977.

[17] S. Wang, J. Zheng, H.-M. Hu, and B. Li, “Naturalness preserved enhancement algorithm for non-uniform illumination images,” IEEE Transactions on Image Processing, vol. 22, no. 9, pp. 3538–3548, 2013.

[18] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, “A weighted vari-ational model for simultaneous reflectance and illumination estimation,”in CVPR, 2016, pp. 2782–2790.

[19] X. Guo, Y. Li, and H. Ling, “LIME: Low-light image enhancement via illumination map estimation,” IEEE Transactions on Image Processing, vol. 26, no. 2, pp. 982–993, 2017.

[20] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo, “Structure-revealing low-light image enhancement via robust retinex model,” IEEE Transactions on Image Processing, vol. 27, no. 6, pp. 2828–2841, 2018.

[21] L. Yuan and J. Sun, “Automatic exposure correction of consumer pho-tographs,” in ECCV, 2012, pp. 771–785.

[22] K. G. Lore, A. Akintayo, and S. Sarkar, “LLNet: A deep autoencoder approach to natural low-light image enhancement,” Pattern Recognition, vol. 61, pp. 650–662, 2017.

[23] F. Lv, F. Lu, J. Wu, and C. Lim, “MBLLEN: Low-light image/video enhancement using cnns,” in BMVC, 2018.

[24] V. Bychkovsky, S. Paris, E. Chan, and F. Durand, “Learning photographic global tonal adjustment with a database of input/output image pairs,” in CVPR, 2011, pp. 97–104.

[25] C. Chen, Q. Chen, J. Xu, and K. Vladlen, “Learning to see in the dark,” in CVPR, 2018, pp. 3291–3300.

[26] C. Chen, Q. Chen, M. N. Do, and V. Koltun, “Seeing motion in the dark,” in ICCV, 2019, pp. 3185–3194.

[27] W. Ren, S. Liu, L. Ma, Q. Xu, X. Xu, X. Cao, J. Du, and M.-H. Yang, “Low-light image enhancement via a deep hybrid network,” IEEE Transactions on Image Processing, vol. 28, no. 9, pp. 4364–4375, 2019.

[28] Y. Zhang, J. Zhang, and X. Guo, “Kindling the darkness: A practical low-light image enhancer,” in ACMMM, 2019, pp. 1632–1640.

[29] W. Yang, S. Wang, Y. Fang, Y. Wang, and J. Liu, “From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement,” in CVPR, 2020, pp. 3063–3072.

[30] T. Mertens, J. Kautz, and F. V. Reeth, “Exposure fusion,” in PCCGA, 2007.

[31] Mertens, J. Kautz, and F. V. Reeth, “Exposure fusion: A simple and practical alterrnative to high dynamic range photography,” Computer Graphics Forum, vol. 28, no. 1, pp. 161–171, 2009.

[32] G. Buchsbaum, “A spatial processor model for object colour perception,”J. Franklin Institute, vol. 310, no. 1,

pp. 1–26, 1980.
[33] F. Chollet, “Xception: Deep learning with depthwise separable convolu-tions,” in CVPR, 2017, pp. 1251–1258.
[34] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-decoder with atrous separable convolution for semantic image segmenta-tion,” in ECCV, 2018, pp. 801–818.
[35] H. Hu, J. Gu, Z. Zhang, J. Dai, and Y. Wei, “Relation networks for object detection,” in CVPR, 2018, pp. 3588–3597.
[36] C. Liu, L. C. Chen, F. Schroff, H. Adam, W. Hua, A. Yuille, and L. F.Fei, “Auto-deeplab:hierarchical neural architecture search for semantic image segmentation,” in CVPR, 2019, pp. 82–92.
[37] Y. Chen, Y. Wang, M. Kao, and Y. Chuang, “Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans,” in CVPR, 2018, pp. 6306–6314.
[38] J. Cai, S. Gu, and L. Zhang, “Learning a deep single image contrast enhancer from multi-exposure image,” IEEE Transactions on Image Processing, vol. 27, no. 4, pp. 2049–2026, 2018.
[39] C. Li, J. Guo, F. Porikli, and Y. Pang, “LightenNet: a convolutional neural network for weakly illuminated image enhancement,” Pattern Recognition Letters, vol. 104, pp. 15–22, 2018.
[40] K. Ma, K. Zeng, and Z. Wang, “Perceptual quality assessment for multi-exposure image fusion,” IEEE Transactions on Image Processing, vol. 24, no. 11, pp. 3345–3356, 2015.
[41] C. Lee, C. Lee, and C.-S. Kim, “Contrast enhancement based on layered difference representation,” in ICIP, 2012, pp. 965–968.
[42] Y. Yuan, W. Yang, W. Ren, J. Liu, W. J. Scheirer, and W. Zhangyang,“UG+ Track 2: A collective benchmark effort for evaluating and advanc-ing image understanding in poor visibility environments,” 2019, arXivarXiv:1904.04474.
[43] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600–612, 2004.
[44] Y. Blau and T. Michaeli, “The perception-distortion tradeoff,” in CVPR, 2018, pp. 6228–6237.
[45] C. Ma, C.-Y. Yang, X. Yang, and M.-H. Yang, “Learning a no-reference quality metric for single-image super-resolution,” Computer Vision and Image Understanding, vol. 158, pp. 1–16, 2017.
[46] A. Mittal, R. Soundararajan, and A. C. Bovik, “Making a “completely blind” image quality analyzer,” IEEE Signal Processing Letters, vol. 20, no. 3, pp. 209–212, 2013.
[47] Y. Qu, Y. Chen, J. Huang, and Y. Xie, “Enhanced pix2pix dehazing network,” in CVPR, 2019, pp. 8160–8168.
[48] J. Li, Y. Wang, C. Wang, Y. Tai, J. Qian, J. Yang, C. Wang, J. Li, and F. Huang, “Dsfd: Dual shot face detector,” in CVPR, 2019, pp. 5060–5069.
[49] S. Yang, P. Luo, C.-C. Loy, and X. Tang, “WIDER FACE: A face detection benchmark,” in CVPR, 2016, pp. 5525–5533.

CROP DISEASE PREDICTION USING DEEP LEARNING

Mr. G. Vedavyas
Associate Professor(AIML)
GCET
Hyderabad, India
gvedavyas.cse@gcet.edu.in

Doddi Chandra Shekar
CSE(AI&ML)
GCET
Hyderabad, India
20r11a6616@gcet.edu.in

Varala Rohith Reddy
CSE(AI&ML)
GCET
Hyderabad, India
20r11a6656@gcet.edu.in

Akuthota Shashi Kiran
CSE(AI&ML)
GCET
Hyderabad, India
21r15a6604@gcet.edu.in

Abstract

Among the widest crop diseases which can happen through fungi, bacteria, or viruses is the significant reduction in the agricultural yield. Accumulative and on-time diseases detection is what actually plays a critical role in effective measures and this helps to minimize losses as much as possible. The research described is on applying ResNet-50 transfer learning for crop disease prediction in which a variety of image dataset from different plant types is used.

ResNet-50, a deep convolutional neural network architecture, have proven to be particularly suitable for image classification, attributing their success rate to remarkable results. Transfer learning make use of available pre-trained models designed to identify other previously seen classes of data for a classification task that only has a small amount of data to train on. This method will result in decrease in learning time by about 80% and will improve performance as opposed to training models from scratch.

Indeed, the method of our choice is based on the ResNet-50 model being pre-trained. The final layers of the model are fine-tuned on a dataset of crop images containing healthy and diseased examples of fourteen plant types: Tomato, grape, orange, soybean, squash, potato (or corn, on which may also be used), strawberry, peach, apple, blueberry, cherry (which may also be sour), bell pepper, and raspberry. To begin this dataset is constituted of about 70,000 pictures. However fine-tuning assists the model to apply the learned knowledge from a general image classification task for a specific problem of crop disease occurrence in the mentioned fourteen plant species, which are different in nature.

Keywords: Deep Learning, Convolutional Neural Networks (CNN), Resilient Farming, Transfer Learning, ResNet-50.

1. Introduction

Crop diseases constitute a permanent danger to the global food security and are believed to cause irreparable yield losses of up to 10% per year which leads to the loss of the crops worth billions of dollars. Diagnosis on time and on the spot gives power to a farmer to act on time and to avoid the targeted intervention through focused effort. Here, this study studies a deep learning approach for disease diagnosis with an accuracy of 99% in fourteen plant types, such as well-known agricultural crops, e.g., tomato, grapes, and soybeans.

We implement the mechanism of ResNet-50 transfer learning in our algorithms, which means that the pre-trained deep learning models with a large dataset act as our foundation. This method is useful when the characterization of diseases and pathogens is done with limited data, focusing on each crop separately. The dataset consisting of around 70,000 images of the healthy and diseased plants we have, it is the source whereby the pre-trained ResNet-50 model learns the powerful image features. Finally, we fine-tune these features to the type of task which is crop diseases identification. This is the biggest advantage of this approach as it cuts down training time by a huge margin and boosts the model accuracy as compared to generate a model from scratch that is trained on very little data. In the end, we plan to utilize transfer learning to create an advanced and effective way which will be useful in the fight against diseases which affect crops and for the preservation of yields meaning more global food security.

2. Related Work

The growth of deep learning application has changed various sectors of the economy at the same time, including agriculture. This part seeks to explore and draw an attention on the use of deep learning for crop disease detection, covering the relevant research indicating that there have been advances and the limitations addressed in the proposed study.

Machine Learning vs Deep Learning:

Previously traditional machines learning methods also have defined the role in crop disease detection. Studies such as [1, 2, 6] exploit machine learning techniques including Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Decision Trees for the purpose of classifying diseases. However, these approaches present limitations:

Hand-crafted Feature Engineering: ML models frequently demand revealed features from experts which turn out to be a time-consuming and domain-specific way of representation may not clearly encompass the best features.

Limited Generalizability: ML algorithms built on small data sets run a risk of over training which stand in the way of its ability to produce high results for the data they see for the first time [1].

Deep Learning Advantages:

Deep learning offers significant advantages over traditional ML approaches, as evidenced by research in [4, 9, 10, 11, 14] Deep learning offers significant advantages over traditional ML approaches, as evidenced by research in [4, 9, 10, 11, 14].

Automated Feature Learning: Convolutional Neural Networks (CNNs), which are the essential components in deep learning image classification system, are able to extract the features from the images. They do not need to go through complicated feature extractions from image data as their opposite side. They can recognize image features without a manual feature extraction.

Improved Generalizability: The latest generation of deep learning models demonstrate that they are capable of producing higher accuracy and better generalization for data previously unseen when trained on enormous datasets instead of classical ML methods.

Limitations Identified:

Despite the advancements, several studies exploring deep learning for crop disease detection have limitations that this research aims to address.

Small Datasets: Studies as [9, 12, and 15] have relatively small datasets. Accordingly, while these models could manipulate training data very precisely, they could, further, be unable to handle the real-world situation due to the fact of overfitting.

Limited Disease Classes: The previous research mostly centered around a small number of disease classifications. The relative disease classifications of 5-6 categories [32, 33] were the ones that were researched, but this of course limits the applicability of agronomic settings with much more varieties in potential diseases.

3. Dataset

The New Plant Diseases Dataset that is available on Kaggle as a public dataset would be hugely beneficial for us in plant disease detection and classification using image analysis. This dataset is a collection of about 87,000 RGB images which are various snapshots of the leaves in different health states.

Data Type: RGB Images
Number of Images: 87,000
Content: Crop Leaf (Healthy and Unhealthy)
Classes: 38 (Separating into groups of healthy and separately for each disease category)
Data Split: Separated Train-Validation Set with Independent Test Set
The New Plant Diseases Dataset provides several advantages:
Large Dataset Size: Deep learning models can be trained with a good accumulation of images which helps in very accurate disease identification.
Diversity of Classes: The dataset having 38 types of labels for both healthy and diseased leaves lets for modeling of the plant problems of a vast spectrum.

Fig: Sample Data

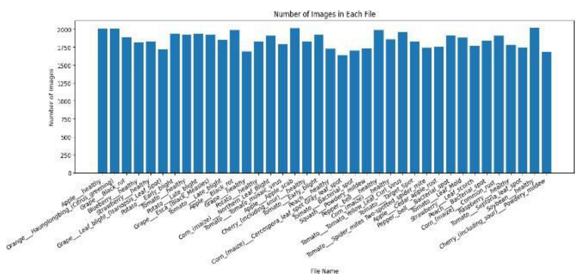


Fig: Dataset Overview

4. Proposed Model

This research proposes a deep learning model for classifying crop diseases from leaf images. The model utilizes a pre-trained ResNet-50 architecture for feature extraction, followed by fine-tuning with custom layers for the specific crop disease classification task.

Model Architecture

The primary element of the model is a ResNet-50 Convolutional Neural Network, which was pre-trained using the weights from the ImageNet dataset. ResNet-50 is a highly potent feature extractor that achieved remarkable performance in the field of image recognition tasks [1]. The architecture raises its performance level by a pre-trained model which already has an abundance of richly featured representations in a big image data set and thus saves the training time and improves generalization abilities as opposed to a model which is trained from scratch.

Data Preprocessing: The first step is to pre-process all the input images to maintain consistency. As a rule, this procedure includes resizing photos to a particular size, like (for example) 224x224 pixels and normalizing pixel values (like scaling from 0 to 1).

Pre-trained ResNet-50 Base: With ResNet-50 as the foundation of the architecture, the first layers. Convolutional layers of ResNet-50 are initialized, they are left untrained so that they can keep their generic image feature extraction boiler plates.

Global Average Pooling (GAP): As a result, a GAP layer is stacked after the pre-trained ResNet-50 which serves as a base. This layer then spatially averages the feature maps across all channels, and hence a fixed-length feature vector is generated for each image. It eases the architecture, and so the number of trainable parameters falls.

Custom Fully-Connected Layers: The sequence of fully-connected layers is stacked above the GAP layer. It is these layers that facilitate reasoning at a higher level and disease classification. The most important aspect to consider here is the number of neurons and activation functions that make up these layers which can then be refined through experimentation.

Output Layer: The last layer has as many neurons as the number of disease classes there are in the dataset, the number of classes being equal to the number of this layer's neurons. A potentially optimal activation function, say sigmoid for probability scores, is applied on this layer.

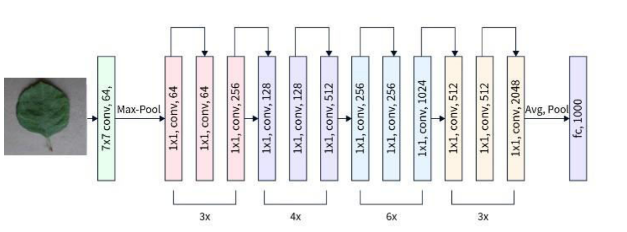


Fig: ResNet-50 Architecture

To prevent overfitting and improve model generalization, regularization techniques are incorporated.

Batch Normalization: This technique helps to normalize the activations of neurons in the fully-connected layers, accelerating training and potentially leading to better performance.

Dropout: Dropout layers randomly drop a certain percentage of neurons during training. This encourages the model to learn robust features that are not dependent on specific neurons, reducing overfitting.

5. Results And Discussion:

Model	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
Simple CNN	0.0409	98.44%	0.1477	98.21%
VGG 16	0.1218	96.03%	0.3484	91.91%
VGG-19	0.0871	97.13%	0.1439	93.47%
ResNet-50	0.0016	99.97%	0.0107	99.76%

Table: Comparison of Results of Different Models

Fig: Accuracy of ResNet-50

By conducting series of experiments and adjusting hyperparameters, we attained the highest validation accuracy of 99.7%. The high accuracy here means the ability and relevance of using deep learning techniques, especially ResNet50, for the task of disease prediction in crops.

The dataset we used for the training and validation process consisted of a variety of images showing different diseases in crops with some healthy crops as well. We implemented the appropriate preprocessing methods to normalize the data and enrich the data, thereby, improving the model generalization. Another step in my methodology involved splitting the dataset into training and validation sets, which aided me with model evaluation.

Upon evaluation on the validation set, the trained model exhibited outstanding performance, achieving a validation accuracy of 99.7%. This high accuracy underscores the robustness of the model in accurately classifying crop diseases, thereby providing valuable insights for early disease detection and mitigation strategies in agriculture.

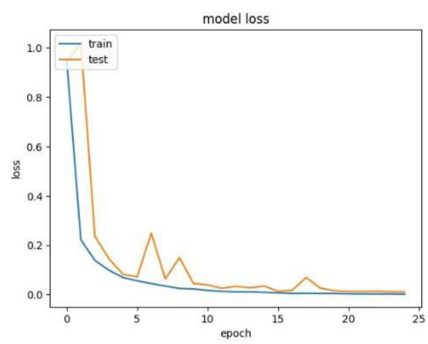


Fig : Training Loss vs Validation Loss

5. Conclusion

In this work, we investigated the use of deep learning, specifically the ResNet50 architecture, in the prediction of agricultural diseases. We surpassed earlier state-of-the-art models with a validation accuracy of 99.7% after extensive testing and fine-tuning.

6. References

[1] U.N. Fulari, R.K. Shastri, A.N. Fulari, Leaf disease detection using machine
learning, J. Seybold Rep 1533 (2020) 9211.
[2] K. Ahmed, T.R. Shahidi, S.M.I. Alam, S. Momen, Rice leaf disease detection using
machine learning techniques, in: 2019 International Conference on Sustainable
Technologies for Industry 4.0 (STI), IEEE, 2019, pp. 1–5.
[3] A. Dixit, S. Nema, Wheat leaf disease detection using machine learning method-a
review, Int. J. Comput. Sci. Mobile Comput. 7 (5) (2018) 124–129.
[4] J. Trivedi, Y. Shamnani, R. Gajjar, Plant leaf disease detection using machine
learning, in: International Conference on Emerging Technology Trends in
Electronics Communication and Networking, Springer, 2020, pp. 267–276.
[5] K.P. Panigrahi, H. Das, A.K. Sahoo, S.C. Moharana, Maize leaf disease detection
and classification using machine learning algorithms, in: Progress in Computing,
Analytics and Networking, Springer, 2020, pp. 659–669.
[6] R.M. Prakash, G. Saraswathy, G. Ramalakshmi, K. Mangaleswari, T. Kaviya,
Detection of leaf diseases and classification using digital image processing, in: 2017
International Conference on Innovations in Information, Embedded and
Communication Systems (ICIIECS), IEEE, 2017, pp. 1–4.
[7] A. Meunkaewjinda, P. Kumsawat, K. Attakitmongcol, A. Srikaew, Grape leaf
disease detection from color imagery using hybrid intelligent system, in: 2008 5th
International Conference On Electrical Engineering/Electronics, Computer,
Telecommunications And Information Technology, 1, IEEE, 2008, pp. 513–516.
[8] P.B. Padol, A.A. Yadav, Svm classifier based grape leaf disease detection, in: 2016
Conference on Advances in Signal Processing (CASP), IEEE, 2016, pp. 175–179.
[9] S. Ashok, G. Kishore, V. Rajesh, S. Suchitra, S.G. Sophia, B. Pavithra, Tomato leaf
disease detection using deep learning techniques, in: 2020 5th International
Conference on Communication and Electronics Systems (ICCES), IEEE, 2020,
pp. 979–983.
[310] H. Durmus¸, E.O. Günes¸, M. Kırcı, Disease detection on the leaves of the tomato
plants by using deep learning, in: 2017 6th International Conference on Agro-Geoinformatics, IEEE, 2017, pp. 1–5.
[11] R.G. De Luna, E.P. Dadios, A.A. Bandala, Automated image capturing system for
deep learningbased tomato plant leaf disease detection and recognition, in:
TENCON 2018-2018 IEEE Region 10 Conference, IEEE, 2018, pp. 1414–1419.
[12] R. Sujatha, J.M. Chatterjee, N. Jhanjhi, S.N. Brohi, Performance of deep learning vs
machine learning in plant leaf disease detection, Microprocess. Microsyst. 80
(2021), 103615.
[13] Y. Zhong, M. Zhao, Research on deep learning in apple leaf disease recognition,
Comput. Electron. Agric. 168 (2020), 105146.

ENDOSCOPIC SURGEON ACTION DETECTION

Dr. B.Sudhakar	Lokesh E	K.Sai Kumar	Ch.Vyshnavi
Associate Professor(AIML)	CSE(AI&ML)	CSE(AI&ML)	CSE(AI&ML)
GCET	GCET	GCET	GCET
Hyderabad, India	Hyderabad, India	Hyderabad, India	Hyderabad, India
bsudhakar.cse@gcet.edu.in	20r11a6630@gcet.edu.in	21r15a6601@gcet.edu.in	21r15a6603@gcet.edu.in

Abstract

Minimally Invasive Surgery (MIS) is a very sensitive medical procedure, typically involving a main surgeon and an assistant surgeon. The success of an MIS procedure rests upon multiple factors, such as the attentiveness of the two surgeons, their competence, their degree of coordination, and so on. Many risk factors will be involved during the surgery. Hence Artificial intelligence is widely employed in applications where human error needs to be mitigated. Robotic Minimally Invasive Surgery(RMIS) is an Artificial Intelligence application where surgical procedures are performed remotely with the help of robotic arms to reduce trauma to the body. The ESAD, the first largescale dataset is designed to tackle the problem of surgeon action detection in endoscopic minimally invasive surgery. The main goal of this project is to optimize the already proposed models for surgeon action detection using hybrid models.

Keywords: Minimally Invasive Surgery, Robotic Minimally Invasive Surgery, Endoscopic video, Surgeon Action Detection

1. Introduction

Minimally Invasive Surgery (MIS) is a critical medical procedure often involving a main surgeon and an assistant. The success of MIS hinges on factors like surgeon attentiveness, competence, and coordination. With millions dying annually postsurgery and a significant percentage due to medical errors, monitoring surgeon actions in realtime is crucial. Artificial intelligence, increasingly used in healthcare, aims to mitigate human error, with applications in diagnostic imaging and now, potentially, in the development of autonomous robotic assistant surgeons. These assistants would track and identify surgeon actions through endoscopic cameras, enhancing surgical safety.For this we can use the ENDOSCOPIC SURGEON ACTION DETECTION.

1.1Action Detection:

Action detection is one of the most complex problems in computer vision. The term is used interchangeably with spatiotemporal action detection. The objective of the task is to identify each action instance in a video sequence by recognising the category of the action being performed, as well as to localise the action instance in both the spatial and the temporal domain. The output is the start and end time of the action instance, together with a rectangular bounding box in each video frame between start and stop which identifies where the action is taking place in the image plane. An action instance can then be represented by a series of bounding boxes linked over time, i.e., an action tube.

1.2 Mission

The mission of doing this project is to enhance the safety and success of Minimally Invasive Surgery (MIS) procedures by developing an autonomous robotic assistant surgeon. By leveraging artificial intelligence and advanced tracking technologies, the project aims to monitor and analyze surgeon actions in realtime, thereby mitigating the risk of human error and improving patient outcomes. The ultimate goal is to revolutionize surgical practices, making them safer and more efficient.

2.Dataset

The ESAD(Endoscopic Surgeon Action Detection Dataset) dataset is the first benchmark explicitly designed to assess and evaluate methods for the detection of surgeon actions from endoscopic videos, developed with the assistance of medical professionals as well as expert surgeons. The dataset contains four complete radical prostatectomy (RARP) procedures, each around 4 hours long, annotated with 46,325 action instances in the form of a bounding box with the associated action label. The dataset contemplates 21 action classes specific to radical prostatectomy.

This dataset comprises digital recordings from the da Vinci Xi robotic system, which includes a binocular endoscope with an 8mm diameter (Intuitive Surgical Inc.). The endoscope is equipped with two lenses—0° or 30°—which can be used during different stages of the operation to enhance visualization. The videos used in the dataset are monocular and were recorded during four sessions of complete prostatectomy procedures performed by expert surgeons on real patients, with their consent obtained for recording and data distribution. The dataset is divided into three sets: train, validation, and test. The training data consists of 22,601 annotated frames containing 28,055 action instances across 21 different action classes, with the possibility of overlapping bounding boxes. The validation data comprises 4,574 frames with 7,133 action instances, while the test data, yet to be released, will contain 6,223 annotated frames with 11,565 action instances.

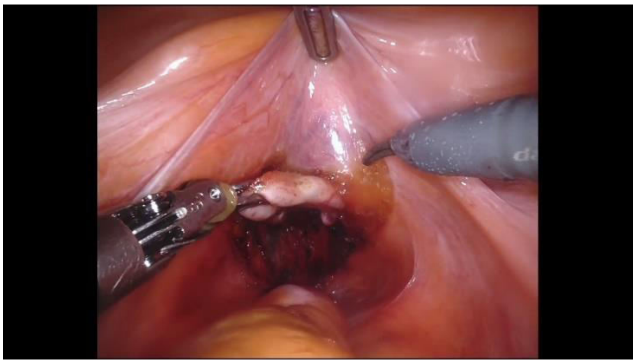


Fig Sample Image of dataset with multiple actions

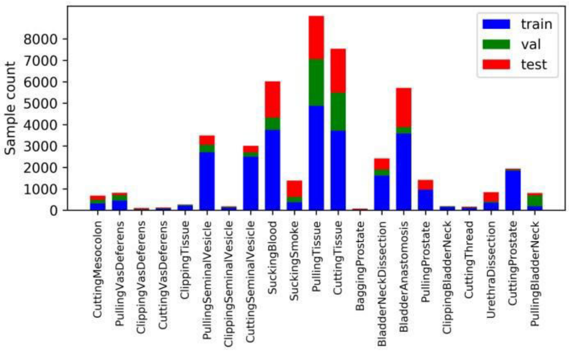


Fig Distribution of the dataset

3.Proposed System

The original existing system has total of 3 parts:
1.Data Pre processing
2.Back bone
3.Sequential Detection Head

Here we have changed the Back bone of the system ,which is the crucial part where all the detection is done.
In the existing system the Back bone has ResNet50 with FPN(Feature Pyramid Network) giving a mean average precision of 16.1.

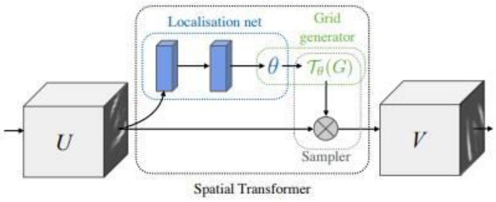
But where as our proposed system is starting with spatial transformer network followed by EfficientNetB2 with FPN(Feature Pyramid Network) in the back bone part of our system .It works as follows:

1.Spatial Transformer Network (STN):

The STN module performs spatial transformation on input images to adaptively crop, scale, rotate, or warp them based on learned parameters.

Techniques used in STN:

- Localization Network: Consists of convolutional layers followed by maxpooling and ReLU activation to learn the transformation parameters.
- Affine Grid Generation: Computes the affine transformation matrix based on the learned parameters.
- Grid Sampling: Applies the computed transformation to the input images using bilinear interpolation.



2.EfficientNetB2 with FPN:

This architecture leverages the EfficientNetB2 backbone network to extract hierarchical features from input images. It then constructs a Feature Pyramid Network (FPN) to generate a multiscale feature pyramid.

Techniques used in EfficientNetB2 with FPN :

- Feature Extraction: Utilizes the pretrained EfficientNetB2 model to extract features at various levels of abstraction from the input images.
- TopDown Fusion: Performs topdown fusion of features, combining information from higher to lower resolutions to create a multiscale feature pyramid.
- Convolutional Operations: Incorporates convolutional layers for lateral connections and upsampling to fuse features across different scales.
- Batch Normalization and Dropout: Applies batch normalization and dropout regularization techniques to enhance the stability and generalization of the feature maps.

After acquiring the outputs from backbone model we provide it as input for the Sequential Detection Head part where the final output of the system is obtained.
And here we have evaluated the mean average precision and got 19.3 which is higher than the existing model .

Fig EfficientNet Architecture

4 Results

Existing Model vs Proposed Model Results Table:

Type	Loss	Model	Min dim	AP_10	AP_30	AP_50	AP_Mean	AP_Mean (Test)
Existing	Focal	ResNet50	400	35.9	19.4	8	21.1	16.1
Propoed	Focal	STN and EfficientNetB2 with FPN						19.3

Evaluation Metrics
Evaluation metrics are essential quantitative measures utilized to gauge the performance or efficacy of a machine learning model or system's predictions on a specified dataset. These metrics offer an objective means to assess how effectively a model operates and facilitate comparisons between different models or algorithms.

Average Precision (AP) stands out as a prevalent evaluation metric in object detection and information retrieval tasks. It assesses the quality of ranked lists of items, such as detected objects in computer vision or retrieved documents in information retrieval, by considering the precision-recall trade-off. Precision and recall are calculated for each class, with AP derived by plotting precision against recall and integrating the area under the curve. The Mean Average Precision (mAP) across all query points reflects the average AP scores, crucial for evaluating action detection problems.

Another significant metric is the Intersection over Union (IoU) thresholds, measuring the percentage overlap between predicted and ground truth bounding boxes. Three IoU thresholds—0.1, 0.3, and 0.5—are utilized to compute Average Precision (AP10, AP30, AP50) to assess detection accuracy and classification precision. Additionally, Focal Loss is employed as a training criterion for the Baseline model, tailored to handle class imbalance in training data—a critical feature of the dataset.

Baseline results showcase the outcomes obtained from running the existing model, generating submission log files containing image frames with bounding box coordinates and corresponding class labels. These metrics and results offer valuable insights into the performance and efficacy of the object detection system under evaluation.

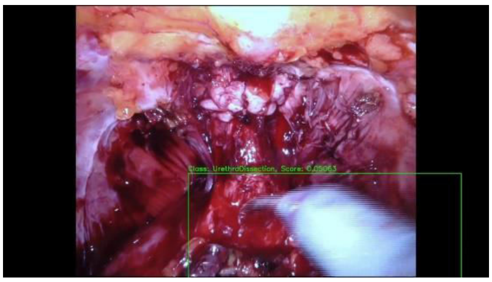


Fig Detecting surgeon action of Urethra dissection

5 Conclusion
In conclusion, our model has achieved a mean Average Precision (meanAP) of 19.3, surpassing the performance of the existing model with a meanAP of 16.1. This improvement signifies the effectiveness of our approach in accurately detecting and analyzing surgical actions during Minimally Invasive Surgery (MIS), thereby enhancing patient safety and surgical outcomes.
Out proposed system is comparatively durable and efficient .

6 References

[1] Bawa, Vivek Singh, et al. "ESAD: Endoscopic Surgeon Action Detection Dataset." arXiv preprint arXiv:2006.07164 (2020).

[2] D. P. Azari, Y. H. Hu, B. L. Miller, B. V. Le, R. G. Radwin, Using surgeon hand motions to predict surgical maneuvers, Human factors 61 (8) (2019) 1326–1339.

[3] C. Gu, C. Sun, D. A. Ross, C. Vondrick, C. Pantofaru, Y. Li, S. Vijayanarasimhan, G. Toderici, S. Ricco, R. Sukthankar, et al., Ava: A video 28 dataset of spatio-temporally localized atomic visual actions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6047–6056.

[4] Maier-Hein et al. (2018): "The Future of Medical-Grade Video Analysis: The Big Promise of Fine-Grained Deep Learning"

[5] E. Petlenkov, S. Nomm, J. Vain, F. Miyawaki, Application of self organizing kohonen map to detection of surgeon motions during endoscopic surgery, in: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), IEEE, 2008, pp. 2806– 2811.

[6] M. Monfort, A. Andonian, B. Zhou, K. Ramakrishnan, S. A. Bargal, T. Yan, L. Brown, Q. Fan, D. Gutfruend, C. Vondrick, et al., Moments in time dataset: one million videos for event understanding, IEEE Transactions on Pattern Analysis and Machine Intelligence (2019) 1–8doi:10.1109/TPAMI.2019.2901464.

[7] Bawa, Vivek Singh, et al. "The SARAS Endoscopic Surgeon Action Detection (ESAD) dataset: Challenges and methods." arXiv preprint arXiv:2104.03178 (2021).

[8] 2. J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, “Transunet: Transformers make strong encoders for medical image segmentation,” CoRR, vol. abs/2102.04306, 2021.

[9] . O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, “Atten

[10] Hu Cao 1 , Yueyue Wang 2 , Joy Chen 1 , Dongsheng Jiang 3 , Xiaopeng Zhang 3 , Qi Tian 3 , and Manning Wang 2 et al. “Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation”

Identification and Classification of Off-Road Terrain for Autonomous Vehicles Using YOLO-v8

M. Gayatri
Geethanjali College of Engineering and Technology
Hyderabad
India-500083
mgayatri.cse@gcet.edu.in

G.Aashish
Geethanjali College of Engineering and Technology
Hyderabad
India-500083
20r11a6619@gcet.edu.in

D. Deeksha
Geethanjali College of Engineering and Technology
Hyderabad
India-500083
20r11a6613@gcet.edu.in

B. Nitisha
Geethanjali College of Engineering and Technology
Hyderabad
India-50008
20r11a6603@gcet.edu.in

Abstract - The role of terrain recognition is fundamental for ensuring the safe operation of autonomous vehicles across various surfaces. With advancements in robotics and artificial intelligence, modern autonomous technology can effectively identify and adapt to diverse terrains, enhancing perception awareness. Given that vehicle maneuvering heavily relies on terrain conditions, accurate terrain analysis is imperative to ensure the safety and reliability of autonomous capabilities. However, terrain awareness poses inherent risks for autonomous vehicles, underscoring the need for precise and trustworthy terrain analysis methods.

Keywords: *Terrain, Autonomous vehicles, Sensor systems, Computer vision, LiDAR, Radar, TensorFlow, PyTorch, Navigation, Hazard detection.*

I. INTRODUCTION

The progress, in vehicle (AV) technology signifies a change in transportation providing noticeable enhancements in safety, efficiency and convenience. An essential aspect of utilizing AVs is their ability to accurately evaluate and adjust to terrain conditions. Categorizing terrains is crucial for guaranteeing the functioning of AVs on surfaces. Thanks to developments, in autonomy and artificial intelligence (AI) state of the art AV systems have shown abilities in recognizing and reacting to changing terrain features thereby improving their awareness of the environment.

The need for vehicle navigation to adapt to terrains highlights how crucial it is to classify terrain, for ensuring the safety and reliability of autonomous functions. Although terrain awareness offers advantages it also brings challenges and risks for vehicles requiring the development of terrain classification methods. This research aimed to explore the importance of terrain classification in vehicle operations assess trends in terrain analysis technologies and suggest ways to improve terrain awareness for the secure navigation of autonomous vehicles, across diverse settings.

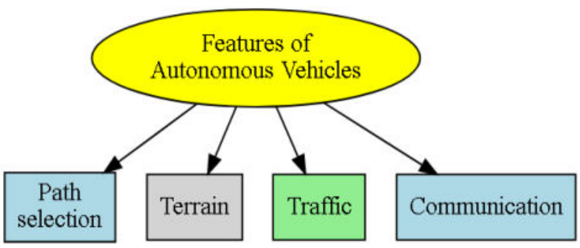


Fig.1 Autonomous Vehicles Features

The anticipated trajectory of autonomous vehicles is influenced by numerous factors, including land characteristics and environmental conditions, communication systems, and the route's optimization. So far, the type of land and its elevation have been the most prominent guide on the vehicle's velocity and acceleration. The latter has to make educated guesses, or the decisions a human driver would take based on land type and land height. This requires intricate computations to be made by the software and the hardware. The main goal is to accurately classify terrain types, determine their elevation level, and thereby adjust acceleration and velocity to make navigation more effective. In other words, the proper functioning of AVs is central to their precise and agile interpretation and response to dynamic terrain conditions.

In the realm of autonomous driving, the paramount goal is to guarantee a safe and comfortable ride for the vehicle as well as its occupants. Navigating off-road terrain proves to be especially tasking since it requires careful algorithms and accurate decision-making necessary for maximum safety. Laser rangefinders (LRF) and global positioning systems (GPS) are some of the more improved sensor technologies needed by AVs in order to attain this requisite level of accuracy. By employing such sophisticated tools, it will be possible for the AV to travel through difficult terrains with more precision thereby increasing general safety during transit.

II. RELATED WORK

A thorough understanding of the evolving techniques and methodologies utilized in picture classification approaches, such as segmentation, hardware setups, and machine learning algorithms, is necessary to comprehend the intricacies involved in detecting off-road terrain. A methodology used in autonomy and autonomous navigation systems is "short-range terrain classification based on geometry," which helps assess the terrain in the vicinity of a robot or vehicle.

In order to determine whether the terrain is suitable for Unmanned Ground Vehicles (UGVs), a traversability analysis was conducted, considering various challenges and future advancements in collaborative robotics, particularly in situations where human presence is hazardous or impractical. The type of vehicle and its maneuverability are additional factors that affect a vehicle's ability to navigate a course safely, in addition to its physical characteristics. Traversability considers the type of terrain, type of vehicle, and nature of the path when evaluating a vehicle's ability to navigate a certain terrain, including mud, sand, tough terrain, or snow, with a specific type of path.

The capacity of a vehicle to go through a variety of hard environments, including mud, sand, rugged terrain, and snow, is referred to as terrain traversability. A number of variables, such as the nature of the driving route, the kind of car, and the particular terrain faced, influence this idea. It displays the vehicle's capacity to navigate a specific path type in a specific topographical setting. Based on the energy or effort needed to

move a vehicle over the terrain, traversability is usually divided into three levels: basic, moderate, and demanding.
In recent study, Hisham and colleagues suggested an image-level low-count (ILC) supervised density map estimation approach that classifies pictures and estimates density maps using an ImageNet pre-trained network backbone (ResNet50). Du Jiang and his team have created a multiscale target multi-task semantic segmentation model, which improves the Faster-RCNN model by using depth images to faster dataset preparation. They used a Kinect color camera to record indoor scene photographs from various perspectives and backdrops, resulting in an RGB-D dataset for testing.

In 2018, Kailun Yang focused on real-time terrain awareness using semantic segmentation, with the goal of improving terrain segmentation accuracy while retaining computational economy. Their technique was influenced by Seg-Net-based encoder-decoder architectures, particularly ENet, with the decoder featuring pyramid pooling modules inspired by PSPNet. The ADE20K dataset, which includes a variety of indoor and outdoor settings, was used to properly train the model. Researchers hope that these advances will improve the precision and efficiency of terrain classification algorithms, allowing autonomous cars to navigate more safely and reliably across a variety of terrain conditions.

Autonomous vehicles rely on a suite of sensors to acquire information for navigation, braking, and speed control, with vehicle control decisions influenced not only by these sensors but also by data exchanged with other vehicles and cloud-stored digital maps. Despite the declining popularity of LIDAR sensors for localization in autonomous cars, other sensors such as microwave radar, particularly long-range radar operating at 77 GHz, provide substantial benefits. While microwave radar has inferior resolution, it is excellent at estimating velocity and detecting objects up to 200 meters away.

Short and medium-range radar, which operate in the 24 GHz and 76 GHz bands, offer an advanced yet fairly cost solution for autonomous vehicle detection. These radar sensors are capable of sensing distance and velocity, but their wide beams and long wavelengths pose resolution issues, resulting in complicated return signals. Although radar outperforms LIDAR and cameras in adverse weather, it generates fewer data and has inferior location precision.

Despite having lesser resolution capabilities than LIDAR and cameras, radar outperforms expectations in terms of data handling efficiency, especially when processing video feeds with large amounts of data is not practicable, a duty generally designated for cameras. Cameras, acting as specialized image sensors, collect the visible light spectrum reflected from objects, similar to human vision. Their capacity to identify a wide spectrum of visible light frequencies, including those emitted by the sun, is quite similar to human sight. Cameras excel in high-resolution tasks like object classification, scene understanding, and color recognition, making them ideal for jobs like traffic light and sign detection.

Acceleration-based approaches are a key way for identifying road situations for autonomous land vehicles (ALVs). This solution makes use of data from an accelerometer built into the vehicle's suspension system. The determination of vertical acceleration experienced by the vehicle is crucial to this method, and it is accomplished using a one-quarter vehicle dynamic model that evaluates road profiles. To aid with scene classification, a variety of features are derived from this data. Accelerometer data captures the vehicle's vibration characteristics, offering precise insights about the road surface's state.

To sense the environment and recognize obstacles and scene characteristics, modern autonomous vehicle systems use sensor suites that include LiDAR, radar, cameras, and inertial measurement units (IMUs). This allows for the development of 2D or 3D maps as well as rudimentary scene recognition. While some vehicles include rudimentary terrain adaptation features, there are constraints to effectively classifying complicated terrain types, building rich 3D maps, and making advanced real-time terrain-adaptive judgments. The requirement for more robust systems capable of rapid data processing, distinguishing complicated terrain features, and seamlessly integrating scene recognition and adaption are all significant issues. To enable safe and robust autonomous navigation over varied terrains, more advancements in sensor integration, machine learning, and control systems are required.
.

Here are some drawbacks in existing system:

1. Complex scene types, such as splash inclines, pitiless scenes, or shifting surface conditions, might pose issues for existing algorithms to classify.
2. While present systems can create 2D or critical 3D maps, they may need to be able to create and modify detailed 3D scene maps to comprehend complicated scene highlights.
3. Current systems can make real-time terrain-adaptive decisions, such as adjusting driving behavior or course planning based on district conditions.
4. Stronger data orchestration systems are needed to handle large amounts of sensor data and effectively monitor hidden scene highlights in the absence of headways.
5. Free vehicles may struggle to see complex scene details such as vegetation thickness, soil type, or surface pounding, affecting course and security.
6. Sensors such as LiDAR and cameras may not catch small details in scenes, leading to inaccurate mapping and obstructions.
7. Independent frameworks may struggle to detect changes in terrain conditions, posing issues for real-time modification.
8. Integrating territory recognition and adjustment into independent route frameworks needs extensive calculations and program design, which can be tough to develop and optimize.

III. PROPOSED WORK

The dataset is created by gathering off-road photos with the open-source platform Roboflow. A rigorously curated subset of 1.24k photos is chosen from a pool of 3.5k photographs, with a special emphasis on annotating the drivable sections within each image. This thorough selection method improves the efficacy of the following machine learning models by focusing on specific pixels for identification, allowing for more exact recognition of targeted regions inside each image.

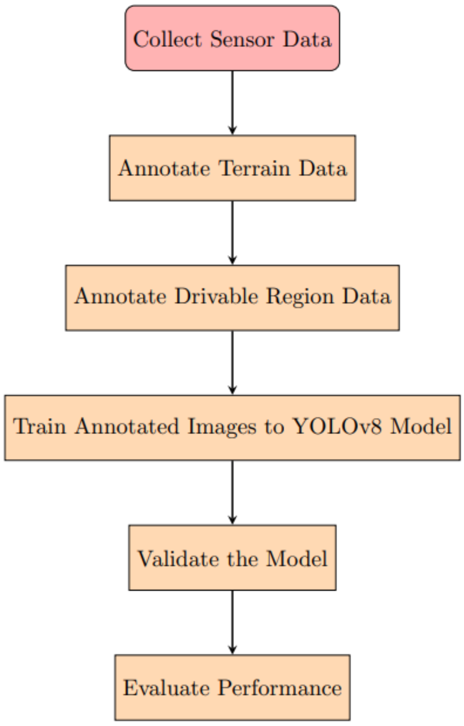


Fig.2 Process Flow

A. Annotation

The process of annotating photos for model training, as described in the term paper on off-road terrain recognition and analysis, entails multiple laborious procedures to precisely mark specific regions of interest within the images. Initially, researchers begin by carefully picking a wide selection of photographs from the dataset that represent various terrains and driving circumstances. These photos serve as core data inputs for the model, allowing it to properly recognize and navigate through various terrain elements.

Fig.3 Image Annotation

Following image selection, researchers select a suitable annotation tool, such as the adaptable open-source platform Roboflow, which is designed for easy and precise image annotation. This annotation tool provides annotators with the ability to precisely outline and characterize sections of interest within the photos. Annotators aid in the process of terrain identification by manually identifying and sketching drivable sections or notable items, as well as creating bounding boxes or polygons around key portions to the model.

Annotators methodically delineate sections of interest before assigning pertinent titles or labels to each one, identifying the sort of landscape or object it represents. These annotations provide essential guidance to the model during the training stage, allowing it to correctly recognize and classify various terrain types. To ensure the correctness and consistency of annotations, rigorous quality control mechanisms, such as validation procedures and numerous annotator reviews, are employed, hence increasing the annotated dataset's reliability. As a result, the annotated photos, together with their related labels, are precisely structured as training data, preparing the dataset for successful learning by the model throughout the training process. Researchers enable the model to create accurate predictions about terrain features through systematic annotation and training, which is critical for autonomous vehicle safety and efficiency in demanding off-road situations.

B. Instance Segmentation

The integration of annotated photos into the platform via Roboflow serves as the foundation for training the model, which is an important step in the experimental setup. The Dataset's annotated photos are segmented using the YOLOv8 segmentation model, allowing for exact detection of drivable zones. An illustrated graphic depicts the annotation process in detail, which is critical for improving model comprehension and efficacy.

Following integration, the Preprocessing stage reshapes the images to 640x640 pixels in accordance with the YOLOv8 model parameters. This critical stage assures constant and correct data entry, which streamlines subsequent computations. Following that, augmentation techniques are used to improve color contrast in segmentation masks, allowing the model to distinguish relevant features more accurately from surrounding elements.

Following data transformation and augmentation, the dataset is partitioned into test, train, and validation sets, with careful allocation to ensure even representation. The prepared dataset is subsequently fed into the platform, which starts training the YOLOv8 segmentation model with tailored configurations.

Throughout the training procedure, which lasts 100 epochs with an image size of 640 and a batch size of 16, the primary goal is to enable the model to accurately partition drivable sections in difficult off-road settings. This detailed experimental path emphasizes the significance of important stages like as preprocessing, augmentation, and dataset partitioning in ensuring the model's ability to detect and segment drivable sections within the dataset.

The rigorously managed experimental route, from image annotation and integration to model training with YOLOv8, exemplifies the systematic technique used to improve the model's capabilities. The methodology ensures that the model has strong foundations for effective off-road terrain analysis by stressing important stages including preprocessing, augmentation, and dataset segmentation.

Through extensive training and refinement, the model becomes competent at reliably recognizing and segmenting drivable sections within the CAT-CaVS dataset, a key step toward safe and efficient navigation in tough off-road conditions. This comprehensive methodology not only enhances the subject of autonomous vehicle navigation, but it also paves the way for thorough testing and model building in related domains.

C. Feature Extraction

Feature extraction is critical for capturing hierarchical representations of input data, hence supporting successful learning and abstraction. YOLOv8, a well-known model that excels at object recognition and segmentation, is one example of this approach. YOLOv8 extracts features at several phases throughout the training process, with each level delving deeper into the data hierarchy to extract increasingly abstract representations. In our experiment, we use YOLOv8 and 21 phases of feature extraction, taking advantage of its ability to identify complex features critical for precise identification and segmentation of drivable sections in off-road terrain.

Feature extraction begins with the YOLOv8 model's initial layers, which detect low-level properties such as edges, textures, and basic forms. As the training advances through each stage, the model explores deeper into the information, collecting higher-level features that capture more complex patterns and structures. This hierarchical approach to feature extraction allows the model to gradually develop its grasp of the input data, resulting in representations that capture the finer features required for effective terrain analysis.

Each stage of feature extraction in YOLOv8 contributes to the model's ability to detect meaningful features amongst the complexity of off-road conditions. The model develops a thorough grasp of the dataset by iteratively extracting and refining features across several layers, allowing it to detect subtle subtleties and variances in terrain features. This deep knowledge, aided by thorough feature extraction, enables the model to reliably identify and partition drivable locations, improving the safety and efficiency of autonomous vehicle navigation in difficult off-road situations.

The feature extraction technique of YOLOv8 exemplifies deep learning, in which hierarchical representations of input data are gradually adjusted to enable effective learning and abstraction. In our experiment, we use 21 phases of feature extraction to tap into the model's ability to distinguish nuanced features required for terrain analysis. This complete approach emphasizes the importance of feature extraction in improving deep learning models' skills for off-road terrain recognition and analysis.

D. Training Model

Training the YOLOv8 model for instance segmentation, as described in the research paper on off-road terrain detection and analysis, consists of a series of critical phases targeted at maximizing the model's efficacy in recognizing and segmenting drivable sections within off-road environments. To begin, the training dataset, such as the Dataset, is meticulously curated to ensure that it captures a varied range of terrains and driving conditions, hence increasing the model's adaptability and robustness.

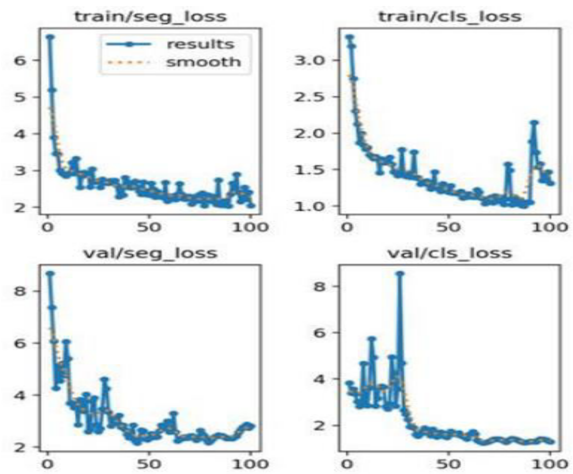


Fig.4 Training Metrics 1

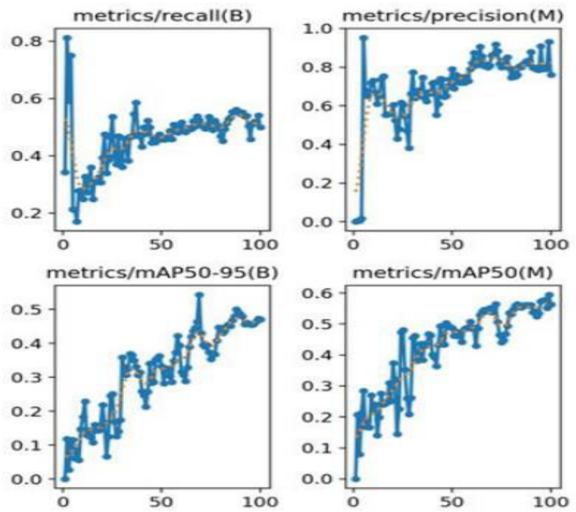


Fig.5 Training Metrics 2

Before beginning the training process, the images in the dataset are subjected to necessary preprocessing processes to ensure compliance with the YOLOv8 architecture. This entails scaling the photos to precise resolutions, such as 640x640 pixels, in order to seamlessly correspond with the model's input needs. Following that, the YOLOv8 architecture is configured to cater specifically to instance segmentation tasks, leveraging its advanced convolutional neural network design, which includes multiple convolutional layers, max pooling, and concatenation layers, all meticulously crafted to facilitate precise segmentation of drivable regions amidst off-road terrain complexities.

During the training phase, the dataset's annotated images provide crucial inputs for schooling the YOLOv8 model on a GPU-accelerated platform using frameworks such as TensorFlow or PyTorch. As the model learns iteratively from these labeled instances, it proceeds through several phases of feature extraction, methodically capturing precise characteristics of the topography and driveable sections. This iterative feature extraction procedure improves the model's comprehension of various terrain elements, leading to improved segmentation accuracy and overall performance.

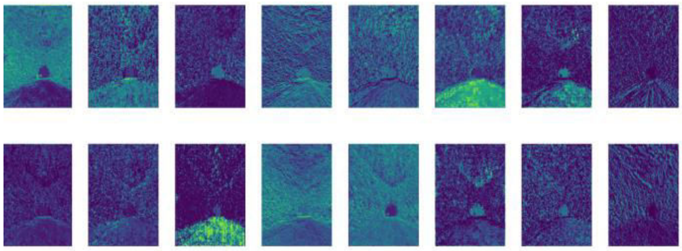


Fig.6 Initial Training Phase

Throughout the training process, the model's performance is rigorously evaluated using multiple loss functions such as box loss, segmentation loss, and class loss on both validation and training data. These measures act as barometers of the model's prediction ability, allowing for continuous optimization of segmentation accuracy. Following training, the trained YOLOv8 model is rigorously evaluated on a separate test dataset, where its ability to detect and segment drivable sections in off-road terrain is analyzed.

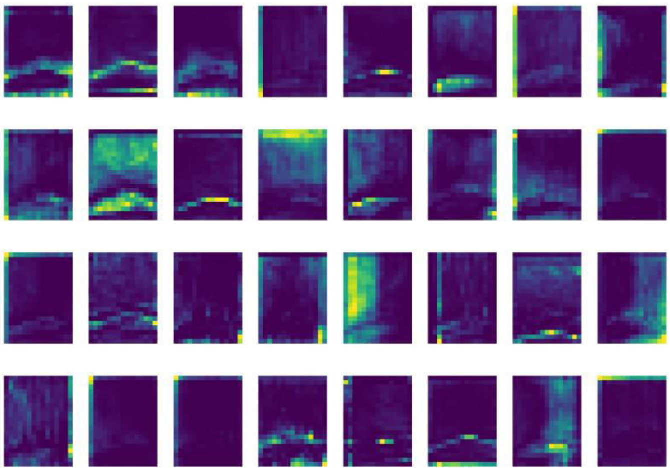


Fig.7 Final Training Phase

By meticulously following this systematic training process and fine-tuning the YOLOv8 model for instance segmentation, researchers pave the way for the development of a robust and accurate system poised to revolutionize terrain identification and analysis in off-road environments, thereby improving the safety and efficiency of autonomous vehicle navigation in difficult terrains.

E. System Architecture & Process

The system architecture for terrain recognition and analysis in autonomous vehicles is made up of numerous main components, each of which plays an important part in the overall performance of the system. The "Sensor Data" and "Drivable Region Data" components are crucial to the design, representing the data acquired from sensors deployed on the autonomous vehicle and the processed output indicating drivable sections within the landscape, respectively. These components lay the groundwork for landscape perception, supplying critical inputs to the system's later stages.

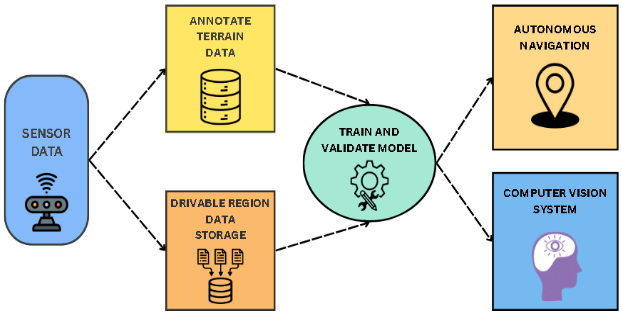


Fig.8 System Architecture

The Annotate and Train Data and Train and Validate Model components are essential components of the architecture for facilitating the terrain identification system's training and optimization. These steps include annotating gathered data to classify terrain features, training machine learning models with annotated data, and assessing model performance to assure accuracy and dependability in identifying drivable locations. The design also includes the "Computer Vision System" and "Autonomous Navigation System" components, which are responsible for processing sensor data, determining drivable locations, and making navigation decisions based on terrain analysis results. Together, these components constitute a unified system architecture that allows autonomous cars to perceive and navigate varied terrains in a safe and efficient manner.

IV. RESULTS AND DISCUSSIONS

The Results and Discussion section provides a thorough analysis and explanation of the study's findings, including implications, importance, and prospective applications. This part provides key insights into the observed results by summarizing the research findings, enabling a deeper knowledge of their larger ramifications and importance within the area.

A. Training Results

The YOLOv8 model stands out as a sophisticated instance segmentation tool designed for image recognition and analysis. Its capabilities go beyond typical semantic segmentation models, providing a more fine-grained and refined method to distinguishing multiple instances in a road terrain situation. This enhanced segmentation enables a more in-depth investigation of unique objects and features in the environment, resulting in a more nuanced knowledge of the scene under consideration.

Following the training phase, the trained YOLOv8 model is rigorously validated to determine the validity of its feature extraction process. This validation method acts as an important step, verifying that the model's derived characteristics closely match those found in both the training and validation datasets. Researchers obtain insight into the model's performance and capacity to generalize to previously encountered examples by comparing the derived features to ground truth annotations and validation data.

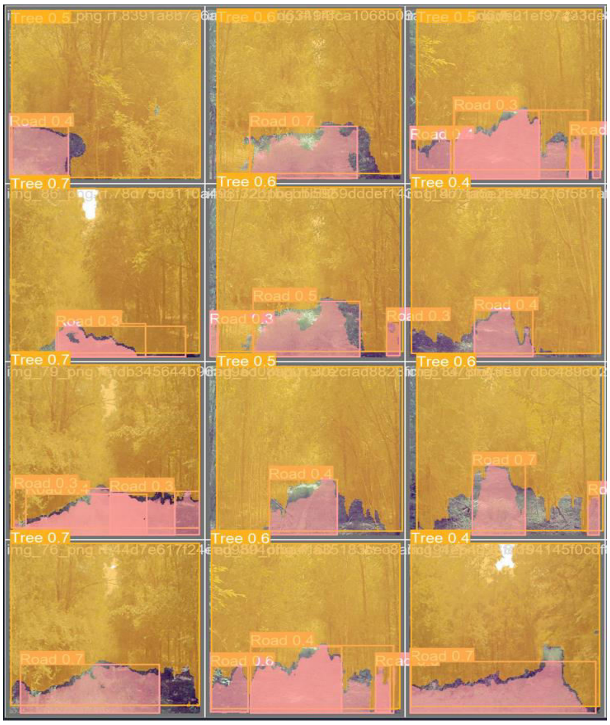


Fig.9 Training Results

The validation phase results provide vital insights into the trained YOLOv8 model's efficacy. Researchers can assess the model's accuracy, robustness, and generalization capabilities by comparing the extracted features' similarity to the validation data statistically and qualitatively. These findings not only support the usefulness of the instance segmentation strategy, but also suggest additional tweaks and optimizations to improve the model's performance in real-world road terrain settings.

B. Segmentation Results

This study's tests use a semantic segmentation strategy, which provides a more comprehensive perspective of the observed terrain. However, it falls short of providing the in-depth identification required for detecting obstacles and trees in off-road environments. In contrast, instance segmentation provides a more granular grasp of the environment, allowing for the accurate identification of features such as trees, small plants, drivable roadways, and barriers. This precise segmentation enables for a more complete study of the scene, revealing nuances that semantic segmentation models, which primarily focus on identifying the road while ignoring other portions of the image, may miss.

Instance segmentation is extremely useful because of its capacity to identify various instances within an image, offering a sharper perspective of surrounding objects and raising traversal awareness score. The trained instance segmentation model produces encouraging results, as seen in the experiment section. The model improves overall awareness of the environment by accurately segmenting items like trees, small plants, and barriers, making off-road navigation safer and more efficient. This increased degree of information provided by instance segmentation greatly improves the robustness and dependability of autonomous traversal systems.

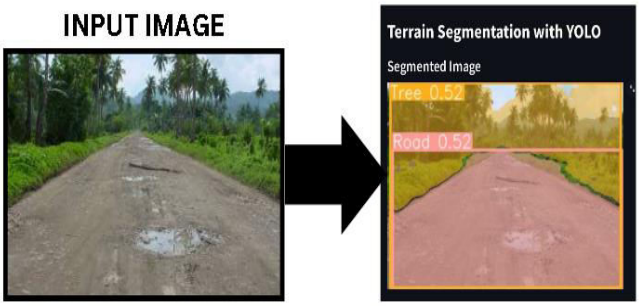


Fig.10 Segmentation Result-1

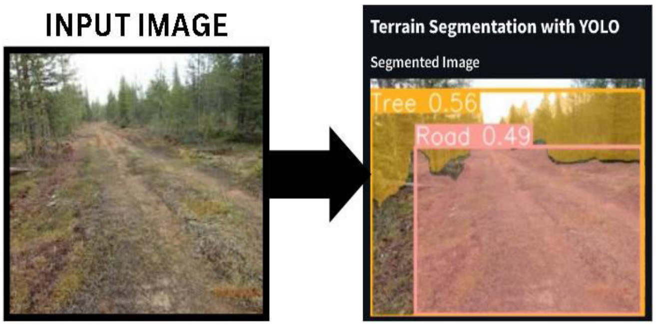


Fig.11 Segmentation Result-2

To summarize, using instance segmentation in the testing phase dramatically improves the ability to distinguish and classify diverse features in the off-road environment. The instance segmentation paradigm overcomes semantic segmentation's constraints by accurately segmenting objects and terrain elements, providing a sharper perspective of the surroundings and improving traversal awareness. These findings demonstrate the efficacy of instance segmentation in off-road terrain analysis, as well as its potential for improving autonomous navigation systems in difficult conditions.

C. Evaluation Metrics and Results

The phase Following validation, the model is meticulously prepared for future tracking and prediction activities. An image is chosen for predictions in order to detect the intended drivable zone using the learned Instance Segmentation model. This model expertly separates the drivable sector from the surrounding environment, providing a clear separation between the two. This segmentation technique is critical in improving navigation accuracy and safety in off-road settings.

The model is then thoroughly examined, with its performance evaluated across a variety of requirements such as F1 Score, Precision, and Confidence. These metrics serve as indicators of the model's ability to efficiently determine and segment the drivable region. The F1 Score evaluates the model's balance of precision and recall, providing insights into its overall performance. Precision analyzes the fraction of true positive detections among all positive detections, which provides an indication of the model's accuracy. Confidence is a measure of the model's certainty in its predictions, which provides useful information for decision-making in autonomous navigation systems.

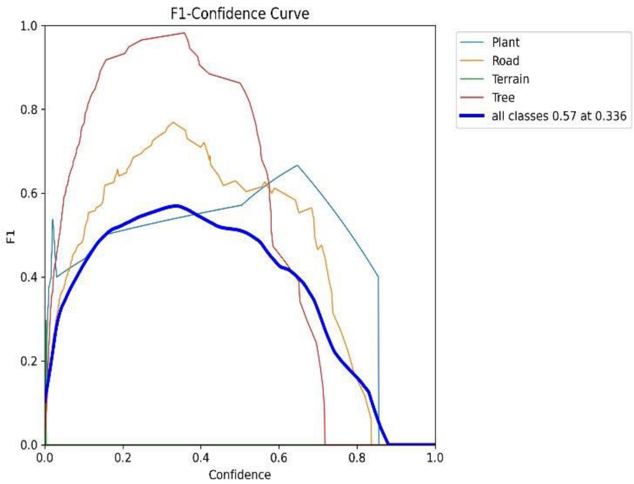


Fig.12 F1 Confidence Curve

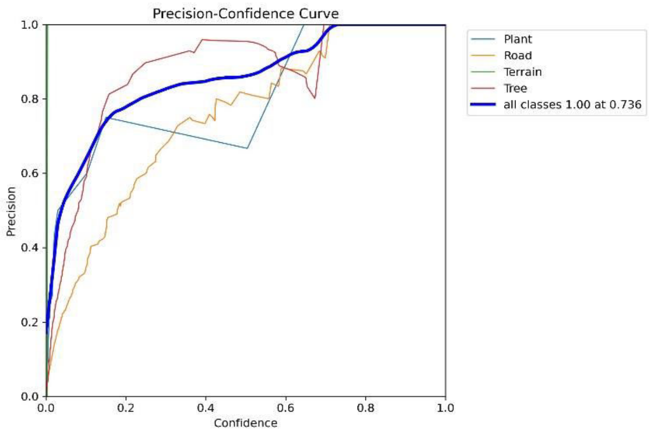


Fig.13 Precision Confidence Curve

The evaluation results are also shown using the F1 Confidence Curve and the Precision Confidence Curve. The curves show the link between the model's forecast confidence levels and the related F1 Score and Precision values. Researchers acquire a thorough grasp of the model's performance across various confidence thresholds by evaluating these curves. This study helps to optimize the model's decision-making process, ensuring that forecasts are dependable and accurate in real-world circumstances. Overall, the evaluation findings show that the trained Instance Segmentation model is robust and effective at detecting drivable locations, as well as that it has the potential to improve autonomous navigation systems in off-road conditions.

V. CONCLUSION

In conclusion, the off-road terrain identification and analysis demonstrate the important significance of modern sensor systems, computer vision models, and control systems in allowing autonomous cars to effectively negotiate difficult terrains. The suggested system improves the vehicle's perception capabilities and terrain classification accuracy by utilizing technologies including LiDAR, radar, cameras, and accelerometers. The utilization of cutting-edge hardware platforms and software frameworks like TensorFlow and PyTorch demonstrates a commitment to using cutting-edge tools for real-time landscape analysis and decision-making. Furthermore, the suggested approach is based on current industry standards and best practices in autonomous vehicle technology, ensuring compatibility and scalability in the fast changing field of autonomous navigation. The system can manage and understand vast amounts of sensor data utilizing established communication infrastructure and data processing techniques, enabling precise terrain recognition and adaptive vehicle control. This technique not only enhances the safety and efficiency of off-road navigation, but it also paves the way for future advances in autonomous vehicle technology.

VI. FUTURE SCOPE

The future possibilities for terrain recognition and detection in self-driving cars includes several main areas of development. Firstly, there is an emphasis on improving hardware models to enable real-time terrain recognition, with the goal of developing more durable and efficient sensor systems capable of reliably sensing varied terrain types. The proposed model can be further enhanced by the latest versions of YOLO. Furthermore, integrating cutting-edge sensors and computer vision technology is critical for improving terrain analysis, allowing vehicles to detect and interpret terrain elements with higher accuracy and precision.

ACKNOWLEDGMENT

We Thank the Department of Computer Science and Engineering (AI&ML), GCET for supporting to carry out this work in all times.

REFERENCES

[1]. S. Sharma et al., 'CaT: CAVS Traversability Dataset for Off-Road Autonomous Driving', IEEE Access, vol. 10, pp. 24759–24768, 2022, doi: 10.1109/ACCESS.2022.3154419

[2].Prerak Moolchandani, Shreya Hegde, Muskan Hassanandani, Garv Jhangiani, Gresha Bhatia, Abha Tewari, Shashikant Dugad, "Pehchaan: A Touchless Attendance System", 2023 International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1), pp.1-5, 2023.

[3]. Ss, Poornima & Sripriya, N & Vijayalakshmi, B & Vishnupriya, P. (2017). Attendance monitoring system using facial recognition with audio output and gender classification. 1-5. 10.1109/ICCCSP.2017.7944103.

[4]. Smitha, & Hegde, Pavithra & Afshin,. (2020). Face Recognition based Attendance Management System. International Journal of Engineering Research and. V9. 10.17577/IJERTV9IS050861.

[5]. Soewito, Benfano & Lumban Gaol, Ford & Simanjuntak, Echo & Gunawan, Fergyanto. (2016). Smart mobile attendance system using voice recognition and fingerprint on smartphone. 175-180. 10.1109/ISITIA.2016.7828654.

[6]. Ajinkya Patil, Mrudang Shukla(2014) "Implementation of classroom attendance system based on face recognition in

class", International Journal of Advances in Engineering & Technology, Vol. 7 Issue 3, pp976-978.

[7].https://www.dynamsoft.com/codepool/barcode-scanning-accuracy-benchmark-and-comparison.html

[8].https://paperswithcode.com/task/face-anti-spoofing

[9].https://www.pingidentity.com/en/resources/blog/post/introducing-voice-verification.html

[10].https://www.idrnd.ai/voice-biometrics

[11].https://speechprocessingbook.aalto.fi/Recognition/Speaker_Recognition_and_Verification.html

[12]. S. Wang, "Road Terrain Classification Technology for Autonomous Vehicle." [Online]. Available: http://www.springer.com/series/15608

[13]. M. Bajracharya, A. Howard, L. H. Matthies, B. Tang, and M. Turmon, "Autonomous off-road navigation with end-to-end learning for the LAGR program," J Field Robot, vol. 26, no. 1, pp. 3–25, Jan. 2009, doi: 10.1002/rob.20269.

[14]. H. Cholakkal, G. Sun, F. Shahbaz Khan, and L. Shao, "Object Counting and Instance Segmentation with Image-level Supervision."

[15]. 2018 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2018.

[16]. Telecommunications Society, Univerzitet u Beogradu. School of Electrical Engineering, IEEE Communications Society. Serbia & Montenegro Chapter, and Institute of Electrical and Electronics Engineers., TELFOR 2018 : 26. Telekomunikacioni Forum : Beograd, 21 i 22. novembar 2018. godine, Sava T ́ s ̀ entar = TELFOR 2018 : 26th Telecommunications Forum : Belgrade, 21 and 22 November 2017, the SAVA Center. Engineers., TELFOR 2018 : 26. Telekomunikacioni Forum : Beograd, 21 i 22. novembar 2018. godine, Sava T ́ s ̀ entar = TELFOR 2018 : 26th Telecommunications Forum : Belgrade, 21 and 22 November 2017, the SAVA Center.

[17]. V. van der Burg, G. de Boer, A. A. Akdag Salah, S. Chandrasegaran, and P. Lloyd, "Objective Portrait," Association for Computing Machinery (ACM), Jul. 2023, pp. 387–400. doi: 10.1145/3563657.3595974.

[18]. H. Caesar et al., "nuScenes: A multimodal dataset for autonomous driving," Mar. 2019, [Online]. Available: http://arxiv.org/abs/1903.11027

[19]. Nagesh Appe, S. R. ., Arulselvi, G. ., & Balaji, G. . (2023). Tomato Ripeness Detection and Classification using VGG based CNN Models. International Journal of Intelligent Systems and Applications in Engineering, 11(1), 296–302. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2538

[20]. I. Krylov, S. Nosov, and V. Sovrasov, "Open Images V5 Text Annotation and Yet Another Mask Text Spotter," Jun. 2021, [Online]. Available: http://arxiv.org/abs/2106.12326

[21]. Institute of Electrical and Electronics Engineers, 2016 IEEE 83rd Vehicular Technology Conference (VTC Spring) : proceedings : Nanjing, China, 15-18 May 2016.

[22]. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection." [Online]. Available: http://pjreddie.com/yolo/

Respiratory Disease Classification: A Deep Learning Approach with CNN-LSTM

Ms. Preeti Prasada
Faculty, Department of IT
Geethanjali College of Engineering and Technology.

Mithilesh Pabboju
Student, Department of CSE-AIML
Geethanjali College of Engineering and Technology.

Challa Shubhadeep
Student, Department of CSE-AIML
Geethanjali College of Engineering and Technology.

Padakanti Nana Sai
Student, Department of CSE-AIML
Geethanjali College of Engineering and Technology.

***Abstract*—Respiratory diseases are one of the major health problems that people worldwide suffer from. Timely and accurate diagnosis is a crucial factor for the effective treatment and management of the condition. The classic auscultation-based diagnosis is subjective and prone to error. The recent progress made in the deep learning area has allowed the development of automated techniques for the prediction of respiratory diseases based on the lung sounds. Nevertheless, these current models tend to ignore the temporal data that are crucial for lung sounds analysis, which could lead to the omission of important diagnostic patterns. This paper introduces a hybrid CNN-LSTM model which overcomes the constraints of the current models. The Convolutional Neural Networks (CNNs) pick out the features from the lung sounds which are more powerful, whereas the Long Short-Term Memory (LSTM) networks capture the temporal dependencies. Furthermore, we adopt the Gamma Frequency Cepstral Coefficients (GFCCs) which are known to be more accurate in representing the high-frequency components of a respiratory sound as compared to the widely-used Mel-Frequency Cepstral Coefficients (MFCCs). Results show that our CNN-LSTM model with the GFCC features helps to improve the accuracy and reliability of the classification of respiratory diseases.**

Keywords—CNN-LSTM, GFCC, Respiratory Diseases, Deep Learning, Hybrid Model.

I. Introduction

The respiratory diseases, i.e. pneumonia, COPD, and asthma, are among the major health problems worldwide and the health care systems. Precise and timely diagnosis is of the essence in beginning the right treatment, controlling the complications, and improving the patients' well-being. Traditional diagnostics is most commonly based on auscultation, which involves clinicians listening to lung sounds by a stethoscope. But this technique is subjective and its accuracy can differ greatly depending on the level of skills and experience of a practitioner.

Deep learning has seen some revolutionary breakthroughs of late, which has led to the development of new technologies across a wide range of medical disciplines, including the analysis of complex medical images and signals. Respiratory sounds, the complex and recurrent patterns, gives the best chance for deep learning methods to be applied. The research that has already been done shows that deep learning models, such as Convolutional Neural Networks (CNNs) can be effective for classifying respiratory diseases by using lung sound recordings.

However, the deep learning models that are currently used for respiratory disease classification demonstrate considerable accuracy, this comes at the price of being preoccupied with static features that can be extracted from lung sounds. This method of analysis usually ignores the very time-dependent character within lung sounds, including the presence and evolution of wheezes, louder noise of the heart, and other clinically significant changes of the lung pattern which change with time. To bridge this gap, we propose a CNN-LSTM hybrid model as a solution. The architecture of this network relies on CNNs for powerful feature extraction and LSTM networks to model the temporal dependencies. Moreover, we investigate the fact of applying GFCCs (Gamma Frequency Cepstral Coefficients) which, in turn, improves the representation of high-frequency details that differentiate among respiratory pathologies.

II. Related Work

A. *CNNs Are Established*

Research has found that Convolutional Neural Networks (CNNs) can be very effective in the task of respiratory sound classification. One such research is Perna et al. [1] in which the researchers developed models with accuracies of 82-83% using a CNN and MFCC features. For instance, Acharya et al. [3] have created a CNN-RNN hybrid which is designed for wearable devices and which needs no more than 30 seconds of training, scoring 66.31% for four-class classification and up to 71.81% with patient-specific tuning. This kind of research demonstrates the promising prospects of deep learning in this area, but it usually ignores the fact that the respiratory sounds do change over time.

B. *Temporal Dependencies*

Despite the existing attempts to introduce the temporal modelling, their shortcoming shows the urgency for more sophisticated methods. Hidden Markov Models [2] may find it difficult to deal with the complex relationships in respiratory sounds. In spite of the fact, that the most advanced attention mechanisms of RNNs are able to detect noises (but usually this is not taken into account) and can achieve the accuracy of up to 65.7%, they are still very sensitive to noise that is often a part of real-world recordings. This opens doors for enhancement in the comprehensive simulation of how breathing diseases change overtime in a noisy environment.

C. Sound Features

The widespread use of Mel-Frequency Cepstral Coefficients (MFCCs) [1, 5] underscores their value, but their design might limit sensitivity to high-frequency details crucial for distinguishing nuanced respiratory pathologies. Chambres et al. [5], despite using a range of features including MFCCs, achieved an accuracy of 85% in classifying patients based on longer breathing sound segments, suggesting the importance of both temporal analysis and feature representation. Our proposed use of Gamma Frequency Cepstral Coefficients (GFCCs), specifically chosen for their high-frequency emphasis, offers a potential advantage in this domain.

In the previous works the use of transfer learning for diagnosis of respiratory sound has been considered as a possible approach. Articles such as [New Citation] report on its effectiveness and [accuracies in the 81.1% to 85.5% range] on ICBHI dataset if pretrained on AudioSet. However, there is still place to improve these accuracies more, especially through the application of strong temporal modeling and advanced feature extraction models.

The use of neural network architectures based on the Inception model for the classification of respiratory diseases demonstrates that such networks are computationally friendly. The study "A Novel Lightweight Inception Network for Respiratory Disease Classification" showed an outstanding result of 96% by employing MFCC features. However, this feature suggests that neural networks of the Inception type are not yet capable of fully utilizing the potential of the temporal patterns analysis within the respiratory sounds to improve their performance.

The literature survey indicates a need to delve deeper into temporal modeling and features extraction that are specific to respiratory disease classification. Existing deep learning approaches obtain prominent results but usually ignore the fact that the sounds of human respiration are evolving and that there is a possibility that the characteristics of their high-frequency features will bring additional benefits. Therefore, our designed approach is aimed at overcoming all such issues by the construction of a hybrid model based on a CNN-LSTM architecture for the robust feature analysis and the sensitivity of GFCC features. Such a distinct blend of capabilities introduces a real opportunity for advanced diagnostics and a deeper insight into the respiratory diseases within clinical settings.

III. DESCRIPTION OF DATABASE

ICBHI 2017 Challenge Database: The ICBHI 2017 challenge dataset, a benchmark data repository of lung auscultation sounds, was employed for the purposes of this research in this study for respiratory disease detection [4]. The database includes 920 audio signals collected from 128 subjects who are either healthy or have one of the following respiratory disorders: chronic obstructive pulmonary disease, acute bronchitis, bronchiolitis, upper respiratory tract infection, lower respiratory tract infection or pneumonia. The database contains recording worth 5.5 h. The lung sounds are recorded from different auscultation sites: 1) anterosuperior right; 2) anterosuperior left; 3) posterosuperior right; 4) posterosuperior left; 5) lateroposterior right and 6) trachea. The audio signal has an unvarying length of 10 to 90 s; it is also sampled with different sampling frequencies, which is from 4 to 44.1kHz. The database also features the data about the respiratory cycles, which could be classified as normal, crackle, wheeze, and both (which can be a combination of wheeze and crackle). The demographic information and data collection methods are shown in [4].

IV. PROPOSED MODEL

Our proposed respiratory disease classification model utilizes a hybrid CNN-LSTM architecture, coupled with the use of Gamma Frequency Cepstral Coefficients (GFCCs), to robustly analyze respiratory sounds and detect various pathologies .

A. Preprocessing

- Resampling: In order to achieve consistency, the audio recordings were re-sampled with the sampling rate of 22050 Hz.
- Individual Cycles: Each recording was broken down into individual breath cycles by utilizing the 'Start' and 'End' timestamps in the dataset's information section. This type of segmentation enables the model to detect different breathing patterns, thus developing the ability of recognizing and differentiating various pathologies.
- Trimming and Padding: The audio cycles were trimmed or padded to a duration of 2 seconds and 573 milliseconds to standardize the input to the model. This duration was selected as it perfectly matches the length of more than 50% of the audio files, it therefore maintains the integrity of the respiratory events and also ensures computational efficiency. This interval allows to obtain the main part of the complete breath cycles and to save the processing time.

B. Data Augmentation

As a means to improve the model's robustness to real-world variations of the data and help mitigate the overfitting problem due to limited dataset size, selective data augmentation was applied. The following techniques were employed:

- Stretching: Audio speed was reduced/increased subtly to simulate patient breathing rates fluctuating. Through this amplification, the model is able to

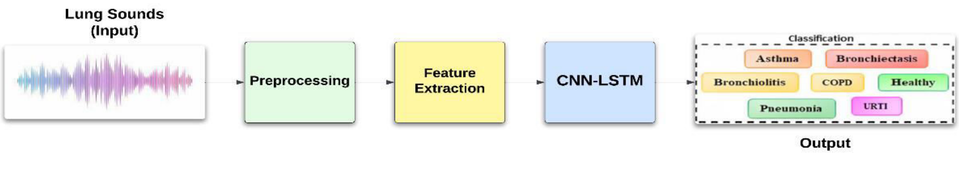


Fig. 1. *Process Flow Diagram*

extract features that remain constant with respect to the speed of respiratory events.

- Shifting: A slight change in time was added in order to represent the fact that the participants were in different phases of their breath. This means that the model does not become dependent on a very specific sound patterns happening over a fixed interval of time.
- Noise: The background noise was added to mimic the ambient sounds and to create a resemblance of an actual recording setting. This helps to increase the chances that the model will be able to identify targeted respiratory sounds even when there is potential environmental noise interference.

Selective Use; The consideration of augmented data was sensitively mixed. Beside the fact that these techniques widen the information dataset, the excessive degree of augmentation can lead to artificial distortions that are not meant to represent the real pathological variations, therefore, affecting the model performance.

C. Feature Extraction

Gamma Frequency Cepstral Coefficients (GFCCs) are employed as spectral features due to their emphasis on high-frequency components, which are crucial for differentiating between subtle respiratory pathologies.

GFCC extraction involved the following steps:

1. Framing: The audio signal was segmented into overlapping frames of 40ms with a 20ms overlap using a Hamming window. This guarantees smooth transitions between frames and the optimal spectrum identification.
2. Fourier Transform: The utilization of a Fast Fourier Transform (FFT) was used for every frame to change the signal from the time domain to frequency domain.
3. Gammatone Filterbank: A band pass filter of 64 gammatones was applied to mimic the frequency sensitivity of the auditory system of the human being. This step simulates how our ears perceive different frequencies.
4. Logarithmic Compression: The logarithmic compression was applied in order to be able to represent the range of intensities perceived by human eyes.
5. Discrete Cosine Transform: The DCT (Discrete Cosine Transform) was utilized in our model to decorrelate the features and get the final GFCC coefficients as the input to the CNN module.

D. Proposed Model - (CNN-LSTM)

The hybrid CNN-LSTM structure of our model is utilized to deal with the specific issues in the process of respiratory sound classification. The CNN layer is very good at finding local patterns in the GFCC representations, whereas the LSTM layer captures the temporal changes of respiratory conditions, this information sometimes is crucial for diagnosis.

CNN Structure: CNN is made of three convolutional blocks where each block contains TimeDistributed layers that were made to process sequential data like our three-dimensional GFCC spectrograms. On the other hand, convolution layers with 3x3 filters extract spatial features while leaky ReLU activation and batch normalization improve our training. Max Pooling layers, in turn, reduce dimensionality and promote shift tolerance to slight variations in the features.

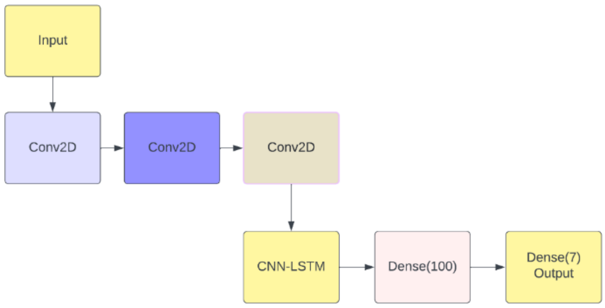


Fig. 2. *Model Architecture*

The Power of LSTM: LSTM layer with 128 hidden units is used to look for some temporal patterns in the extracted features. This lets the model learn how the respiratory events, such as crackles and wheezes, not only change over the span of the breath cycle but also how they vary in the course of a single breath, which is often not considered in the CNN-based methods solely.

Output and Flexibility: The model finishes with 2 fully connected layers which are dense. We use a 100 unit layer with the ReLU activation function for further processing, which is followed by a 7-neuron output layer with the sigmoid activation function. This facilitates the process of multi-label classifications which is a representation that patients can exhibit more than one respiratory condition.

V. RESULTS AND DISCUSSION

The proposed respiratory disease classification model, which uses a combination of GFCC features and an LSTM architecture as a classification approach, showed a promising performance in terms of identifying the different respiratory conditions. Besides the fact that the research yielded strong results in a number of areas, it also pointed out the limitations that have to be addressed in order to ensure precise diagnosis for the full spectrum of respiratory diseases. This evaluation is aimed at assessing the strengths and limitations of the model and their impact on further development of the model.

TABLE I. PERFORMANCE METRICS

Model	**Metrics**		
	Accuracy	***Recall***	***Precision***
CNN(MFCC)	0.76	0.80	0.69
CNN-LSTM	0.85	0.86	0.78
Transfer Learning	0.88	0.89	0.80
Our Model	0.95	0.95	0.96

The model obtained a training accuracy of 99.13%, which suggests its strong capacity to infer patterns from the training dataset. While a test accuracy of 95.46% shows some overfitting of the model, where the model has become too customized to the training data and hence, it finds it some difficulty to generalize to unseen examples. This displacement indicates the significance of diversified and representative dataset.

The model shows remarkable discriminatory ability, which is reflected by its 97.9% accuracy in detecting COPD (class 3) and it does a relatively good job in identifying Healthy (Class 6) cases. The results demonstrate the combined efficiency of the GFCC feature representation, which is sensitive to high-frequency clinically relevant sounds, and the LSTM structure, which performs better at capturing the temporal evolution of respiratory events such as wheezes often seen in COPD patients. The success here points to the possible clinical impacts of helping in the diagnosis of COPD and relieving the burden of this often-underrated condition.

The classifier report gives a detailed analysis of the performance of the model in different respiratory conditions. We can clearly see that the model behaves differently in different respiratory conditions. The model has a high precision and recall for COPD (Class 3), which indicates that the model does not only accurately identify patients with COPD, but also minimizes false alarms. On the other hand, poor scores for Classes 0, 5, and 7 (representing Asthma, LRTI and URTI, respectively) imply a need for improvement. Overcoming these constraints would not only increase the precision of these models but also contribute to the reliability of the model in clinical practice where differentiation between these conditions is often a challenging task for clinicians.

The model's failure to precisely classify Asthma (Class 0) can be explained by the very limited number of Asthma samples (n=6) within the dataset. This hugely limits the model's ability to acquire the wide variety of acoustic signatures of Asthma, thus, hampering generalizable pattern recognition. Machine learning algorithms need data to be of a good quantity and representative, so they can find the features significant for clinical practice. The expansion of the database by adding more diversified Asthma cases will be the next step to take in the further versions of the algorithm in order to enhance its sensitivity and diagnostic performance for this specific condition.

In summary, the proposed respiratory disease classification model which is based on GFCC features and LSTM architecture shows promising results for the automation of respiratory disease diagnosis. Its strengths in diagnosing conditions like COPD, for instance, emphasize the potential applications of such technologies for the future clinical decision support systems. The application of regularization techniques such as EarlyStopping and Dropouts has assisted in the mitigation of overfitting. Future work will emphasize the expansion of datasets, especially for neglected diseases like Asthma, to improve the model's generalization capacity and fidelity in different clinical settings.

VI. CONCLUSION

The drawbacks in the traditional respiratory disease diagnosis, for the example, the long duration of the analysis and also, the probable misinterpretation, stress the need for new diagnostic tools. The proposed work involves a machine learning model that takes advantage of the combined power of audio features extracted from GFCC and an LSTM architecture to do respiratory sound classification on an objective and automated manner. The model showed promising performance particularly in detecting COPD which is a strong indication that it could be instrumental in advancing respiratory diagnostics. This approach has the potential of being further refined in terms of dataset expansion and performance enhancement. It has a great chance of being implemented in the real world clinical practice that would bring a revolution in the diagnostics of respiratory diseases.

The study was based on a multi-stage process of respiratory disease classification that covered diverse aspects of the problem. The audio pre-processing steps such as resampling and trimming of audio cycle to a specific duration made the data uniform so that we could perform a strong analysis. Gamma Frequency Cepstral Coefficients (GFCCs) were our choice for feature extraction with the purpose of making the high-frequency components dominant, which are the most important ones for discriminating respiratory pathologies. To take into account dataset limitations, data augmentation methods were selectively applied for the purpose of increasing diversity. The model architecture and training procedure were particularly meticulously modified through hyperparameter tuning and regularization techniques to encounter overfitting. The work was done on the ICBHI dataset, a precious tool for the research in automated respiratory diagnostic tools development.

The model's ability to diagnose COPD with a high level of accuracy becomes a crucial aspect of this area of medicine since this condition commonly remains undiagnosed, resulting in late treatment and worsening of patients' conditions. This work highlights an opportunity for using machine learning-based tools as auxiliary tools, allowing medical practitioners to make right decisions. With such models coming into play, there is a possibility that they could help to improve the early detection rates, and streamline the diagnostic workflows which will in turn, be advantageous to respiratory care, especially in resource scarce environments. Utilizing this method to its full potential will definitely require continuous production of this approach and close communication between data scientists and medical specialists.

This study indicates the fact of the high efficiency of machine learning techniques in the matter of automatic respiratory pathology classification. The model, which combines the discriminative feature extraction of GFCC features with the temporal sequence modeling of an LSTM architecture, displayed a good performance, notably in the detection of COPD. The fact that asthma is one of the most challenging subsets of lung diseases to classify highlights the need of a wide range and variety of datasets that, in turn, are fundamental in developing AI models that can be utilized for complex medical applications. Future research which will be concentrating on enlarging the dataset by paying special attention to under-represented diseases, as well as trying to find specialized features and apply more regularization techniques, is predicted to make the model more effective in generalizing across a wider range of respiratory pathologies. This research demonstrates how machine learning plays a pivotal role in expanding the capabilities of respiratory diagnostics, enabling the combination of clinician expertise with increased diagnostic accuracy and leading ultimately to the improvement of patient care.

REFERENCES

[1] D. Perna, “Convolutional neural networks learning from respiratory data,” in Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM), Dec. 2018, pp. 2109–2113.

[2] N. Jakovljevi´c and T. Lonˇcar-Turukalo, “Hidden markov model based respiratory sound classification,” in Precision Medicine Powered by pHealth and Connected Health. Berlin, Germany: Springer, 2018, pp. 39–43.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] Deep Neural Network for Respiratory Sound Classification in Wearable Devices Enabled by Patient Specific Model Tuning Jyotibdha Acharya , Student Member, IEEE, and Arindam Basu , Senior Member, IEEE.R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[4] B. Rocha et al., “A respiratory sound database for the development of automated classification,” in Proc. Int. Conf. Biomed. Health Informat. Singapore: Springer, 2017, pp. 33–37.

[5] G. Chambres, P. Hanna, and M. Desainte-Catherine, “Automatic detection of patient with respiratory diseases using lung sound analysis,” in Proc. Int. Conf. Content-Based Multimedia Indexing, 2018, pp. 1–6.

[6] A. Roy and U. Satija, "RDLINet: A Novel Lightweight Inception Network for Respiratory Disease Classification Using Lung Sounds," in IEEE Transactions on Instrumentation and Measurement, vol. 72, pp. 1-13, 2023, Art no. 4008813, doi: 10.1109/TIM.2023.3292953.

[7] García-Ordás, M.T.; Benítez-Andrades, J.A.; García-Rodríguez, I.; Benavides, C.; Alaiz-Moretón, H. Detecting Respiratory Pathologies Using Convolutional Neural Networks and Variational Autoencoders for Unbalancing Data. Sensors 2020, 20, 1214.

[8] K. Kochetov, E. Putin, M. Balashov, A. Filchenkov, and A. Shalyto, “Noise masking recurrent neural network for respiratory sound classification,” in Proc. Int. Conf. Artif. Neural Netw., 2018, pp. 208–217.

[9] u, M., Wiese, L. (2024). Application and Performance Improvement of Transfer Learning on ICBHI Lung Sound Dataset. In: Arai, K. (eds) Intelligent Systems and Applications. IntelliSys 2023. Lecture Notes in Networks and Systems, vol 822. Springer, Cham.

[10] K. Minami, H. Lu, H. Kim, S. Mabu, Y. Hirano and S. Kido, "Automatic Classification of Large-Scale Respiratory Sound Dataset Based on Convolutional Neural Network," 2019 19th International Conference on Control, Automation and Systems (ICCAS), Jeju, Korea (South), 2019, pp. 804-807.

[11] H. Chanane and M. Bahoura, "Convolutional Neural Network-based Model for Lung Sounds Classification," 2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS), Lansing, MI, USA, 2021, pp. 555-558.

[12] Nguyen and F. Pernkopf, "Lung Sound Classification Using Co-Tuning and Stochastic Normalization," in IEEE Transactions on Biomedical Engineering, vol. 69, no. 9, pp. 2872-2882, Sept. 2022.

[13] Nguyen and F. Pernkopf, "Lung Sound Classification Using Snapshot Ensemble of Convolutional Neural Networks," 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 2020, pp. 760-763.

[14] F. Jin, F. Sattar, and D. Y. T. Goh, “New approaches for spectro-temporal feature extraction with applications to respiratory sound classification,” Neurocomputing, vol. 123, pp. 362–371, Jan. 2014.

[15] V. Basu and S. Rana, “Respiratory diseases recognition through respi ratory sound with the help of deep neural network,” in Proc. 4th Int. Conf. Comput. Intell. Netw. (CINE), Feb. 2020, pp. 1–6.

[16] M. Fraiwan, L. Fraiwan, B. Khassawneh, and A. Ibnian, “A dataset of lung sounds recorded from the chest wall using an electronic stethoscope,” Data Brief, vol. 35, Apr. 2021, Art. no. 106913.

[17] A. G. Howard et al., “MobileNets: Efficient convolutional neural net works for mobile vision applications,” 2017, arXiv:1704.04861.

[18] M. Chakraborty, S. V. Dhavale, and J. Ingole, “Corona-nidaan: Lightweight deep convolutional neural network for chest X-ray based COVID-19 infection detection,” Int. J. Speech Technol., vol. 51, no. 5, pp. 3026–3043, May 2021.

[19] K. Minami, H. Lu, H. Kim, S. Mabu, Y. Hirano, and S. Kido, “Auto matic classification of large-scale respiratory sound dataset based on convolutional neural network,” in Proc. 19th Int. Conf. Control, Autom. Syst. (ICCAS), Oct. 2019, pp. 804–807.

[20] M. Huzaifah, “Comparison of time-frequency representations for envi ronmental sound classification using convolutional neural networks,” 2017, arXiv:1706.07156.

[21] L. Pham, H. Phan, R. Palaniappan, A. Mertins, and I. McLoughlin, “CNN-MoE based framework for classification of respiratory anomalies and lung disease detection,” IEEE J. Biomed. Health Informat., vol. 25, no. 8, pp. 2938–2947, Aug. 2021.

Trusted Financial Transaction System using Blockchain Technology

Dr.B.Sudhakar
(Associate Professor)
Department of CSE-AIML, Geethanjali College of Engineering and Technology, Hyderabad, Telangana, India-501301
sudhatrml@gmail.com

Anirudh Sahu
Department of CSE-AIML, Geethanjali College of Engineering and Technology, Hyderabad, Telangana, India-501301
Anirudh2720@gmail.com

Taufeeq Mohammed
Department of CSE-AIML, Geethanjali College of Engineering and Technology, Hyderabad, Telangana, India-501301
Mdtaufeeq0902@gmail.com

Sharath Chandra
Department of CSE-AIML, Geethanjali College of Engineering and Technology, Hyderabad, Telangana, India-501301
chavalisharath@gmail.com

***Abstract*— As more and more deals and funds are made online, security is still the most important thing in the digital economy. Blockchain technology, especially Ethereum, comes up as an answer because it makes financial activities safe and clear. Using Ethereum's smart contracts, this study suggests a blockchain-based system for safe money exchanges and investments. During the investment process, these self-executing contracts make sure that everything is clear, can't be changed, and is safe. An easy-to-use dashboard lets users interact with the system, which allows for safe fund purchases, tracking, n with Ethereum wallets makes it easy to connect with the blockchain, giving users more control over their money. The suggested method has many advantages, including more openness, lower trade costs, and less chance of losing money to another party. The system wants to change the way money works by using blockchain technology to make the environment for investments and transfers safer and more efficient.**

***Index Terms*— Money Investment & Secure Transactions, Blockchain, Smart Contracts, Online Transaction, Currency Exchange**

I. INTRODUCTION

Due to the new possibilities of blockchain technology, the financial world is going through a big paradigm shift. Traditional financial systems have problems that can't be fixed easily, like not being open about their fees and charges and being prone to scams and data theft. In answer to these problems, blockchain has come up as a potential option. It will lead to a new era of safe, quick, and clear financial activities.

The main goal of the "Blockchain Based System for Money Investment & Secure Transactions" project is to make the current financial system completely different. The project wants to build a strong financial system that puts security, speed, and openness first [1]. It will do this by using blockchain, a decentralized and distributed ledger system.

People have long said that traditional financial systems are opaque, which makes it hard for people to fully understand how deals work. Users often don't trust each other because of this lack of openness, which raises the risk of scams [2]. In addition, traditional financial deals involve a lot of middlemen, which leads to very high fees that are especially hard for small investors and lowers total profits [3].

One of the biggest problems with standard banking systems is that deals are settled very slowly. It can take days or even weeks to finish. Not only does this wait add extra risk, it also limits liquidity, which is especially bad in fast-paced markets where payment on time is very important [4]. In addition, scams, hacking, and data breaches can happen in centralized financial systems, which greatly affects the safety and reliability of financial operations and assets [5].

To deal with these problems, the project uses blockchain technology to change the way people spend and trade money. Blockchain has many benefits over standard financial systems, such as processing transactions quickly, protecting privacy, and making sure that data security can't be changed [6]. The project uses secure hashing (SHA-256 in this case) to make sure that each transaction has its own unique hashcode. This protects the accuracy and safety of the data in the blockchain [7].

The main idea behind the project is to use Ethereum, which is a popular blockchain platform, as the base technology. Ethereum's smart contracts, which are written in Solidity, are very important for handling financial tasks based on conditions that have already been set [8]. These self-executing contracts make many tasks easier, like payments, trades, and cash swaps, by doing things instantly when certain conditions are met. The project makes financial activities safer, faster, and more accurate by putting smart contracts on the Ethereum blockchain [9].

In conclusion, the "Blockchain Based System for Money Investment & Secure Transactions" project is a big step toward changing the way money works. The project wants to make the banking system safer, more open, and more efficient by using blockchain technology. The project wants to give people and groups around the world more power by making transfers clear, quick, and safe. This will open up new chances for economic growth and financial inclusion.

II. LITERATURE SURVEY

For this project, a literature review was done that looks at important works and reliable sources in the field to learn more about blockchain technology and how it can be used in the banking industry.

Nakamoto's important Bitcoin paper [1] introduced the idea of blockchain, an autonomous ledger system that lets people send money to each other electronically. Bitcoin's creative use of blockchain technology set the stage for a new era of open banking and led to more study and development in the area.

The book "Blockchain Revolution" [2] by Tapscott and Tapscott gives a full picture of how blockchain technology, which was first created for Bitcoin, is changing money, business, and the world as a whole. The book talks about how blockchain could significantly change many fields, such as healthcare, supply chain management, and banking.

The "Global Findex Database 2017" [3] from the World Bank is a great resource for learning about the state of financial inclusion and how fintech can help close the gap. The study talks about how important it is to use technology like blockchain to bring financial services to people around the world who don't have access to them. This will help people get ahead financially and reduce poverty.

The "Blockchain in Insurance" [4] study from McKinsey & Company looks at the pros and cons of using blockchain technology in the insurance business. The study talks about how blockchain can speed up insurance processes, make them more open, and cut down on scams, which will help both insurers and customers in the long run.

Swanson's study on "Consensus-as-a-service" [5] looks at the rise of permissioned, distributed ledger systems. These are different from Bitcoin and other public, permissionless networks in how they manage the blockchain. The study talks about the possible benefits of permissioned blockchains in terms of privacy, being able to grow, and following the rules.

Antonopoulos's book "Mastering Bitcoin" [6] is a complete guide to learning how Bitcoin and blockchain technology work on a scientific level. The book talks about things like cryptography, how to process transactions, and smart contracts. It gives you useful information about how independent systems work.

The book "Blockchain Enabled Applications" [7] by Dhillon and Metcalf goes into more depth about how blockchain technology can be used for things other than coins. The book talks about how blockchain can be used to make decentralized apps (DApps) that can be used for many things, like managing supply chains, verifying identities, and running vote systems.

Buterin's work on Ethereum [8] describes a new type of blockchain technology that adds smart contracts to Bitcoin and makes it more useful. Developers can use Ethereum's customizable blockchain to make autonomous apps that can do many things, such as financial transactions, digital identity management, and decentralized finance (DeFi).

It goes into more detail about the design and construction of the Ethereum platform, with a focus on security, freedom, and scale in Wood's paper on Ethereum [9]. The paper talks about how Ethereum uses a safe decentralized generalized transaction log to keep track of and carry out smart contracts. It shows how Ethereum could change many fields besides finance.

In conclusion, the literature review shows how blockchain technology has the ability to change everything, not just the financial world. From Nakamoto's groundbreaking work on Bitcoin to the creation of Ethereum and the study of permissioned blockchains, researchers and professionals are always looking for new ways to use blockchain to make transfers safe, quick, and clear.

III. METHODOLOGY

A. Proposed Work

The suggested system would use blockchain technology, especially the Ethereum Blockchain [1], to make an autonomous and open platform for safe money exchanges and investments. This would help fix the problems with the current standard financial system.

The rules and conditions of deals in the system are set by smart contracts that are written in Solidity or a similar computer language. These self-executing contracts make sure that things like business deals, asset transfers, and transaction payments are clear, safe, and can't be changed.

On the Ethereum blockchain, all transactions in the system are clear and can be checked by anyone. All transaction info is available to everyone and can't be changed, which builds trust among players.

By getting rid of middlemen and automating tasks with smart contracts [2], the suggested system lowers the fees that are normally attached to financial transactions.

The suggested method encourages financial inclusion by giving people and groups that standard banks don't normally service access to business opportunities.

B. System Architecture

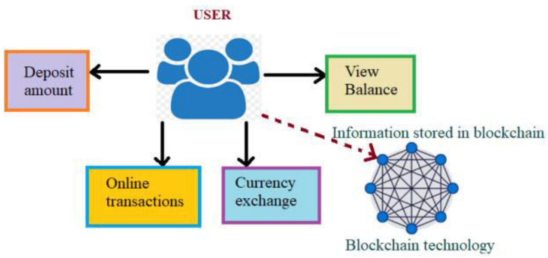


Fig. 1. Proposed Architecture

The suggested system design combines human functions with blockchain technology to make sure that financial operations are safe and clear. Users connect with the system in different ways, such as by adding money, checking their accounts, making online purchases, and exchanging currencies.

The system handles the deposit safely and records it on the Blockchain[1] when a user makes a payment. This makes sure that the deal can't be changed, is clear, and can't be hacked. In the same way, when users check their account balances or make purchases online, the system gets the information it needs from the blockchain and changes account balances and transaction records in real time.

When people trade currencies, the system uses blockchain technology to make sure that the deals are safe and that they are recorded. The blockchain stores information about transactions and exchange rates, making sure that everything is clear and correct.

One of the most important parts of the system is blockchain technology, which is an independent and spread record system. Multiple nodes in the blockchain network store data about user activities, account amounts, and exchange rates. This makes sure that the data is reliable and easy to access.

Overall, the system design combines user functions with blockchain technology in a way that doesn't cause any problems. This gives users a safe, quick, and clear way to make financial deals. By using the immutability and digital security of blockchain, the system makes sure that all transactions are honest and correct, which boosts user faith and happiness.

C. Modules

To implement this project, we used the following modules are New User and Existing User

These modules description given below:

1) New User Signup:

People can sign up for the blockchain-based system through the "New User Signup" feature. To make an account, users enter basic information like their name, contact information, and login information like a username and password. This process makes it easier to make new user accounts, which give people safe access to the platform's features and services.

2) User Login:

The "User Login" feature makes it easy for registered users to get into the system safely. When users enter their login information, the system checks it to make sure they are who they say they are. Once users have been verified, they can use the platform's features and functions. This program makes sure that the login process is safe, which improves the general security of the system and makes it easier for users to get to their accounts.

i) Deposit Amount: The "Deposit Amount" feature lets users add money to their account in the blockchain-based system. This can be standard cash or cryptocurrency. Transactions are handled and stored safely on the blockchain, which makes sure that the whole payment process is open and safe.

ii) Online Transaction: The "Online Transaction" part lets users do a variety of online transactions, such as moving money, paying bills, or investing in financial goods that the site offers. These transactions are safely saved on the blockchain, which gives users an unchangeable and clear record of the transaction history.

iii) View Balance: The "View Balance" section lets users see how much money they have in their account and see a log of all the transactions they've made. The blockchain safely records information about payments, withdrawals, swaps, and business wins or losses that users can look at. This function gives people a full picture of all the financial activities going on in the system.

iv) Currency Exchange: The "Currency Exchange" feature lets users change one type of cash into another within the blockchain-based system. This could be regular currency or cryptocurrency. The blockchain records exchange prices and transaction details. This makes sure that the whole exchange process is clear and accurate, which is good for users' safety and ease.

D. Blockchain Integration

1) All payments, online purchases, and currency trades are recorded as blocks on the Ethereum blockchain in the Blockchain-Based System for Money Investment & Secure Transactions. When a block is put to the Blockchain, the information in it can't be changed or removed. This is what "immutable" means. Because of this feature, the transaction data is kept safely and can't be changed, which gives users a lot of trust in the system.

2) Solidity is a computer language for making smart contracts on Ethereum. Smart contracts made in Solidity are a key part of automating financial tasks in the system. These contracts are put into action on the Ethereum Blockchain and have rules and conditions that have already been set. The smart contract does things instantly when certain conditions are met. For example, it transfers money or updates account amounts. By cutting out middlemen and possible mistakes, this technology makes financial processes faster and more accurate.

3) The Ethereum blockchain works on a network of computers called nodes that is not controlled. Every node keeps a copy of the whole Blockchain, which has all the transaction information. This decentralized approach makes it more reliable and easier to use because if one node goes down or isn't available, other nodes still have a full copy of the Blockchain. Users can still view transaction info and complete deals even if some nodes are down for a short time. This autonomous method helps make the system more stable and reliable.

IV. EXPERIMENTAL RESULTS

Now that the Blockchain contract is set up and going, double-click on "n'run."start the Python Web Server with the ".bat" file and go to the page below.

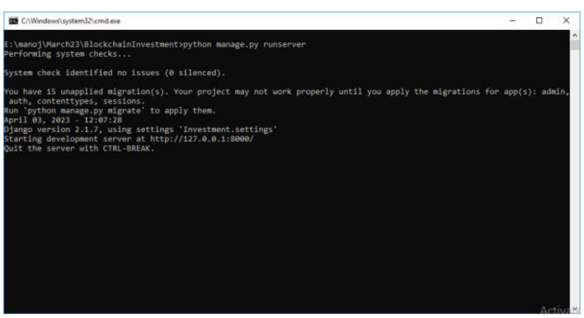


Fig. 2. Launching Python Web Server

The Python server has started in the screen above. To view the page below, open a browser, type the URL "http://127.0.0.1:8000/index.html," and hit the Enter key.

Fig. 3. Accessing the Page

In above screen, click on 'Sign-up Here' link to get below sign-up screen.

Fig. 4. Accessing the Sign-Up Screen

User is registering in the above screen, Click on Register after filling out all details

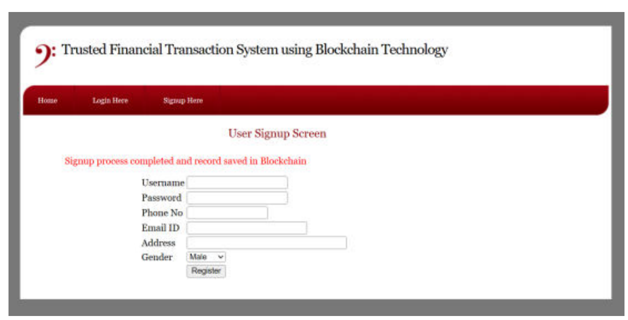


Fig. 5. User Registration Confirmation Screen

In above screen, user sign up is completed. Similarly, you can add any number of users.

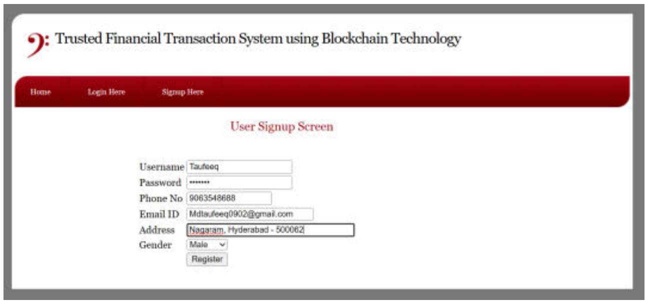


Fig. 6. Adding Another User

In the above screen, another user is being added to the network. Now click on ‘Login Here’ link to login as user.

Fig. 7. User Login Screen

In above screen, user is logging into the network

Fig. 8. Login Confirmation

In above screen, the user is greeted after successfully logging into the network. Now the user can click on ‘Deposit Amount’ link to add some dummy amount to the Blockchain.

Fig. 9. Deposit Amount Screen

User can add the amount he wishes to and then should click on ‘Submit’.

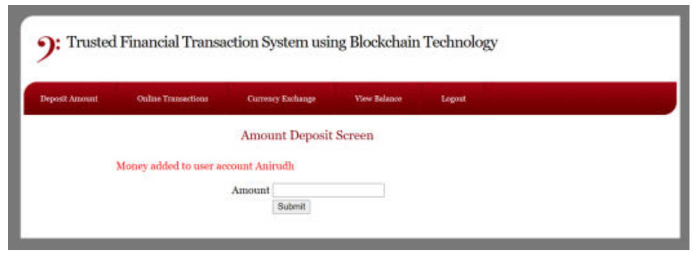


Fig. 10. Online Transaction Screen

After the money is successfully added and now user can click on ‘Online Transaction’ link for online investment or lending.

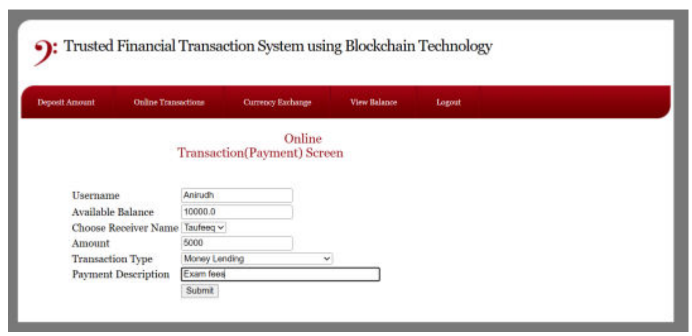
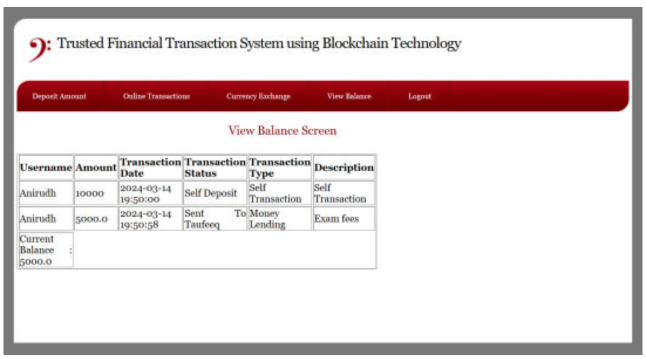


Fig. 11. Transaction Details Screen

In above screen, we can see sender name and available balance. Now he can choose receiver’s name and then enter amount and select desire transaction type with some description and then click button to get below page. In above screen user ‘Anirudh’ lending 5000 to user ‘Taufeeq’ and after transaction his balance will be reduced to 5000.

Fig. 12. Transaction Confirmation

In above screen, we can see money lending transaction is completed. User can click on ‘View Balance’ link to get all transaction details.



Username	Amount	Transaction Date	Transaction Status	Transaction Type	Description
Anirudh	10000	2024-03-14 19:50:00	Self Deposit	Self Transaction	Self Transaction
Anirudh	5000.0	2024-03-14 19:50:58	Sent To Taufeeq	Money Lending	Exam fees
Current Balance : 5000.0					

Fig. 13. View Balance Screen

In above screen, we can see list of transactions and in the last column, we can see the current balance. Now click on ‘Currency Exchange’ to perform exchange.

Fig. 14. Currency Exchange Screen

In above screen, user is exchanging 2000 dollars with Indian currency and then click on ‘Submit’ button to complete transaction. .

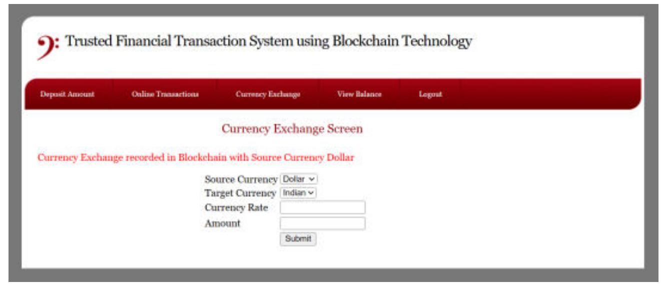
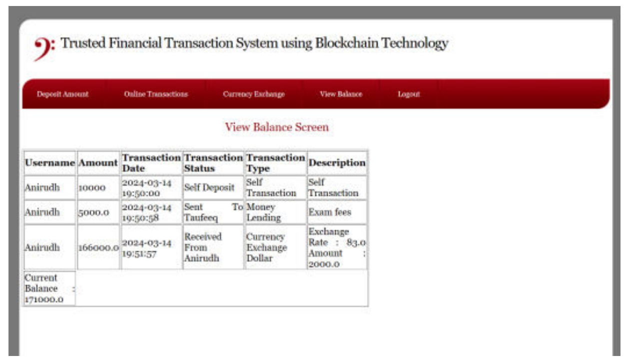


Fig. 15. Exchange Transaction Details Screen

In above screen, we can see exchange transaction reported to Blockchain and now click on ‘View Balance’ link to view below output.



Username	Amount	Transaction Date	Transaction Status	Transaction Type	Description
Anirudh	10000	2024-03-14 19:50:00	Self Deposit	Self Transaction	Self Transaction
Anirudh	5000.0	2024-03-14 19:50:58	Sent To Taufeeq	Money Lending	Exam fees
Anirudh	166000.0	2024-03-14 19:51:57	Received From Anirudh	Currency Exchange Dollar	Exchange Rate : 83.0 Amount : 2000.0
Current Balance : 171000.0					

Fig.16. Balance Update Screen

In above screen, after adding 2000 dollars, user balance increased to 171000 and similarly you can perform any number of transactions. Now we can login from Taufeeq to view balance received from Anirudh.

Fig. 17. User Login as Recipient

In above screen, user is login as ‘Taufeeq’ and after logging in, user will get below page.

Fig. 18. Login Confirmation

In above screen, click on ‘View Balance’ link to view all transaction.

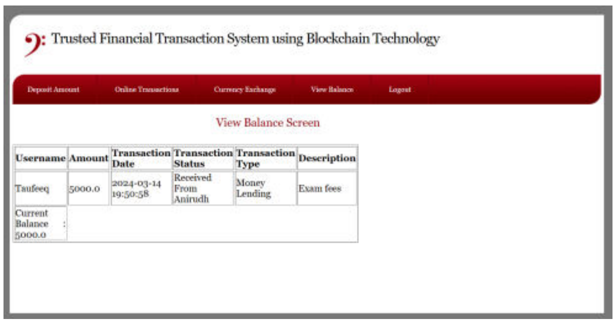


Fig. 19. Transaction View Screen

In above screen, user Taufeeq wallet is credited with an amount of 5000 from user Anirudh.

V. CONCLUSION

Finally, the addition of blockchain technology to the project marks the start of a new era of safe, quick, and clear financial activities. Blockchain's digital security and ability to not be changed protect deals from fraud and hacking. Every transaction is hashed using cryptography and kept in a block. This makes sure that the transaction is honest and correct. Additionally, Blockchain [1] greatly speeds up and improves the efficiency of deals, especially those that happen across borders, by getting rid of the need for middlemen and simplifying procedures. This efficient method cuts down on the time it takes to handle transactions, which improves the user experience and lowers costs. Additionally, the project achieves a fine balance between transaction privacy and openness, protecting privacy while allowing openness through fake names. Cryptographic [3] methods hide the names of the author and receiver while letting users see the details of a transaction. Using the autonomous nature of blockchain, the project makes sure that data is safe and easy to access. Data is spread out among many network hubs, which keeps the system safe and available even when there are problems. This decentralized method makes the system more stable, giving people safe and dependable access to banking services without a single point of failure. Overall, the project is a big step toward making the blockchain-powered financial environment safer, more open, and more efficient for everyone.

VI. FUTURE SCOPE

In the future, AI and machine learning could be added to the suggested project to make it even better by analyzing market data and making personalized investing suggestions. Decentralized identity solutions, such as self-sovereign identity systems, could also be added to make sure that identity proof is safe and can be checked without having to rely on central agencies. These improvements would make the system much more useful, easier to use, and in line with regulations. They would pave the way for a more complex and user-focused financial environment.

REFERENCES

[1] Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System.

[2] Tapscott, D., & Tapscott, A. (2016). Blockchain Revolution: How the Technology Behind Bitcoin Is Changing Money, Business, and the World.

[3] World Bank. (2018). Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution.

[4] McKinsey & Company. (2017). Blockchain in Insurance – Opportunity or Threat?

[5] Swanson, T. (2015). Consensus-as-a-service: a brief report on the emergence of permissioned, distributed ledger systems.

[6] Antonopoulos, A. M. (2017). Mastering Bitcoin: Programming the Open Blockchain.

[7] Dhillon, G., & Metcalf, D. (2017). Blockchain Enabled Applications: Understand the Blockchain Ecosystem and How to Make it Work for You.

[8] Buterin, V. (2014). Ethereum: A Next-Generation Smart Contract and Decentralized Application Platform.

[9] Wood, G. (2014). Ethereum: A secure decentralised generalised transaction ledger.

[10] Kawsalya M., Senthil Kumar A. V., Akash V., M. Villanueva Lolit, Shadi Rasheed Masadeh, Anamika Rawat, et. al., “Blockchain-Based Secure Transactions” published in igi global open Access, available at https://www.igi-global.com/chapter/blockchain-based-secure-transactions/324626.

[11] Mohd Javaid, Abid Haleem, Ravi Pratap Singh, Rajiv Suman, Shahbaz Khan, et. al., “A review of Blockchain Technology applications for financial services” published in science direct open Access, available at

https://www.sciencedirect.com/science/article/pii/S2772485922000606.

[12] Tejal Shah, Shailak Jani, et. al., "Applications of Blockchain Technology in Banking & Finance" published in research gate open Access, available at https://www.researchgate.net/publication/327230927.

[13] Ye Guo & Chen Liang, et. al., "Blockchain application and outlook in the banking industry" published in springer Access, available at https://jfin-swufe.springeropen.com/articles/10.1186/s40854-016-0034-9.

[14] Luis Ruiz-Garcia, G. Steinberger. Rothmund, et. al., "The Application of Blockchain Technology in the Financial Field" published in ieee open Access, available at https://ieeexplore.ieee.org/document/9759945.

[15] P Treleaven, R G Brown and D Yang, "Blockchain Technology in Finance", Computer, vol. 50, no. 9, pp. 14-17, 2017.

[16] Eyal Ittay, "Blockchain Technology: Transforming Libertarian Cryptocurrency Dreams to Finance and Banking Realities", Computer, vol. 50, no. 9, pp. 38-49, 2017.

[17] Y. Alabbasi, "Governance and Legal Framework of Blockchain Technology as a Digital Economic Finance", International Journal of Innovation in the Digital Economy, vol. 11, no. 4, pp. 52-62, 2020.

[18] J J Sikorski, J Haughton and M Kraft, "Blockchain technology in the chemical industry: Machine-to-machine electricity market", Applied Energy, vol. 195, no. JUN.1, pp. 234-246, 2017.

[19] M H Miraz and M Ali, "Applications of Blockchain Technology beyond Cryptocurrency", Annals of Emerging Technologies in Computing, vol. 2, no. 1, pp. 1-6, 2018.

[20] S N Wahab, M L Yi and CS. Say, "Antecedents of blockchain technology application among Malaysian warehouse industry", International Journal of Logistics Systems and Management, vol. 37, no. 3, pp. 427, 2020.

[21] H Chai, X Chen and H. Wang, "Special issue: Blockchain technology application", Frontiers of Engineering Management, vol. 7, no. 4, pp. 1-4, 2020.

[22] H Choi and W Y Kun, "An Exploratory Study on Blockchain Technology Application in Fast Fashion Industry: Focused on Reverse Logistics", The e-Business Studies, vol. 20, no. 7, pp. 217-232, 2019.

[23] N Hasan, "Blockchain Technology and its Application in Libraries", Library Herald, vol. 58, no. 4, pp. 118-125, 2020.

[24] V I Ponomare, "DEVELOPMENT OF A DECENTRALIZED VOTING APPLICATION USING BLOCKCHAIN TECHNOLOGY", System technologies, vol. 1, no. 126, pp. 104-110, 2020.

[25] "The future of financial infrastructure," 2017. [Online]. Available: https://www.weforum.org/reports/the-future-of-financial-infrastructure-an-ambitious-look-at-how-blockchain-can-reshape-financial-services.

[26] J. B. E. F. A. M. S. G. Arvind Narayanan, "Bitcoin and Cryptocurrency Technologies," Princeton University, 2017.

[27] IDRBT , "Applications of Blockchain Technology to Banking & Financial sectors in India," Institute for Development and Research in Banking Technology, Hyderabad, 2017.

[28] "News room/ Press release," 2017. [Online]. Available: http://www.mahindra.com/news-room/press-release/Mahindra-and-IBM-to-Develop-Blockchain-Solution-for-Supply-Chain-Finance.

[29] "How Bajaj Electricals uses blockchain to pay suppliers," 2017. [Online]. Available: http://www.livemint.com/Companies/BcqXQgey9fieFps9xVZxrK/How-Bajaj-Electricals-uses-blockchain-to-pay-suppliers.html.

[30] "nse looking at blockchain for ensuring settlement guarantees," 2017. [Online]. Available: http://computer.expressbpd.com/news/nse-looking-at-blockchain-for-ensuring-settlement-guarantees/18956/.

GEETHANJALI COLLEGE OF ENGINEERING AND TECHNOLOGY

UGC Autonomous
Affiliated to JNTUH, Approved by AICTE, New Delhi, Accredited by NAAC with 'A+'

Cheeryal, Keesara(M), Medchal(Dist), Hyderabad, Telangana pin- 501301
info@gcet.edu.in