American Journal of
AI Digital Transformation and Regenerative Pharmacist
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0418 www.zesterapublications.com Original Research Paper

DEEP FAKE AUDIO DETECTION USING DEEP LEARNING

Mrs.K.TEJASWI[1],MIRIYALKAR BRIDULA[2],MUDILI DEVI ALEKHYA[3],KOMATIPALLI VIJAY KUMAR[4],SONTI SASI KALA[5]

Assistant Professor, Dept. of CSE, V.K.R,V.N.B.&A.G.K COLLEGE OF ENGINEERING

[2345] UG Students, Dept. of CSE,

V.K.R,V.N.B. &A.G.K COLLEGE OF ENGINEERING,GUDIVADA

ABSTRACT

The rapid advancement of artificial intelligence has enabled the creation of highly realistic synthetic speech, commonly known as deep fake audio, which poses significant risks to security, privacy, and digital trust. Deep fake audio can be used for misinformation, identity fraud, and unauthorized voice cloning, making reliable detection methods essential. This research proposes a deep learning-based framework for detecting fake audio by analyzing subtle acoustic patterns and inconsistencies that are difficult for humans to identify.

The proposed system utilizes advanced neural network architectures such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), or Transformer-based models to learn discriminative features from audio spectrograms and raw waveform signals. Preprocessing techniques including noise reduction, feature extraction (MFCC, Mel-spectrogram), and normalization are applied to enhance model performance. The model is trained and evaluated on benchmark datasets containing both real and synthetic speech samples.

Experimental results demonstrate that deep learning models can effectively differentiate between genuine and manipulated audio with high accuracy, precision, and recall. The system shows strong potential for real-time applications in cybersecurity, digital forensics, and social media monitoring. This work highlights the importance of combining robust feature engineering with advanced deep learning architectures to combat the growing threat of AI-generated audio manipulation.

I INTRODUCTION

The rapid development of artificial intelligence and deep learning technologies has significantly transformed the field of speech synthesis, enabling machines to generate highly realistic human-like voices. While these advancements have brought many positive applications such as virtual assistants, audiobooks, and accessibility tools, they have also introduced serious challenges in the form of deep fake audio. Deep fake audio refers to artificially generated or manipulated speech that mimics a real person's voice, often making it difficult to distinguish between genuine and fake recordings. This technology can be misused for spreading misinformation, financial fraud, identity impersonation, and social engineering attacks.

Traditional audio verification methods rely on manual analysis or handcrafted features, which are often insufficient to detect sophisticated synthetic speech generated by modern neural networks. As a result, researchers are increasingly focusing on deep learning-based detection systems that can automatically learn hidden patterns and anomalies in audio signals.

American Journal of
AI Digital Transformation and Regenerative Pharmacist
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0418 www.zesterapublications.com Original Research Paper

Techniques such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Transformer models have shown promising results in identifying subtle artifacts introduced during the audio generation process.

This project aims to develop a robust deep learning framework for deep fake audio detection by analyzing spectral and temporal features extracted from speech signals. By leveraging advanced neural architectures and large-scale datasets, the system seeks to improve detection accuracy and reliability. The proposed approach can contribute to enhancing digital security, protecting public trust, and supporting applications in cybersecurity, media authentication, and forensic analysis.

II RELATED WORK

Deep fake audio detection has gained significant attention in recent years due to the rapid growth of voice cloning and speech synthesis technologies. Early research in this domain focused on traditional machine learning techniques that relied on handcrafted acoustic features such as Mel-Frequency Cepstral Coefficients (MFCC), spectral contrast, and pitch-related characteristics. Classical classifiers like Support Vector Machines (SVM), Gaussian Mixture Models (GMM), and Random Forests were initially used to distinguish between genuine and synthetic speech. Although these approaches achieved moderate success, they struggled to generalize against advanced deep fake generation models.

With the advancement of deep learning, researchers began adopting Convolutional Neural Networks (CNNs) to analyze spectrogram-based representations of audio signals. CNN-based models demonstrated improved performance by automatically extracting spatial patterns and spectral inconsistencies introduced during synthetic voice generation. Later, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) architectures were explored to capture temporal dependencies in speech signals, enabling better detection of sequential anomalies present in deep fake audio.

Recent studies have also investigated Transformer-based architectures and self-supervised learning models for deep fake detection. These methods leverage attention mechanisms to focus on critical regions of audio data and learn contextual representations from large-scale datasets. Additionally, researchers have explored hybrid models combining CNN and LSTM layers to benefit from both spatial and temporal feature extraction. Benchmark datasets such as ASVspoof and Fake-or-Real audio datasets have been widely used to evaluate system performance.

Despite significant progress, challenges remain due to the continuous evolution of deep fake generation techniques. Current research emphasizes improving model robustness, cross-dataset generalization, and real-time detection capabilities to effectively combat increasingly sophisticated audio manipulation methods.

III LITERATURE REVIEW

Recent advancements in deep learning have significantly influenced the development of deep fake audio detection systems. Researchers have explored various neural network architectures and feature extraction techniques to identify synthetic speech with high accuracy. Early studies primarily focused on analyzing acoustic features such as Mel-Frequency Cepstral Coefficients (MFCC), spectral flux, and pitch variations to distinguish between real and generated audio signals. These approaches laid the foundation for

American Journal of
AI Digital Transformation and Regenerative Pharmacist
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0418 www.zesterapublications.com Original Research Paper

automated detection but faced limitations when dealing with highly realistic AI-generated voices.

Several research works introduced Convolutional Neural Networks (CNNs) for spectrogram-based analysis, where audio signals are converted into visual representations to capture hidden patterns. CNN-based models demonstrated strong performance in identifying artifacts introduced during the speech synthesis process. Later, researchers incorporated Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks to analyze sequential information in audio data, improving the model's ability to detect temporal inconsistencies present in deep fake speech.

More recent studies have explored attention-based and Transformer architectures that enable models to focus on important segments of audio, improving generalization across different datasets. Self-supervised learning methods, such as pretrained speech representation models, have also been applied to reduce dependency on large labeled datasets. Hybrid frameworks combining CNN, LSTM, and attention layers have shown promising results by capturing both spectral and temporal features simultaneously.

In addition to model architecture, many researchers have investigated data augmentation, noise robustness, and cross-domain evaluation to enhance detection performance in real-world scenarios. Despite these advancements, challenges remain due to the continuous improvement of speech synthesis technologies, making it essential to design adaptive and scalable deep learning solutions for reliable deep fake audio detection.

DISADVANTAGES

The traditional placement management approach used in many higher education institutions has several limitations that affect efficiency and effectiveness. One major disadvantage is the heavy dependence on manual processes such as maintaining student records in spreadsheets, sending emails for communication, and manually verifying eligibility criteria. This increases the chances of human errors, data duplication, and delays in updating information. Additionally, the absence of a centralized platform makes it difficult for students, recruiters, and placement officers to access accurate and real-time data, leading to confusion and miscommunication.

Another significant drawback is the lack of intelligent matching between student skills and job requirements, resulting in inefficient shortlisting and missed opportunities for suitable candidates. The existing system also struggles with scalability when the number of students and companies increases, creating additional workload for placement staff.

IV PROPOSED SYSTEM

The proposed Smart Placement Management System is a centralized, web-based platform designed to automate and optimize the entire campus placement process. The system connects students, placement officers, and recruiters through a single integrated interface, enabling efficient data management, communication, and recruitment operations. Students can register, create professional profiles, upload resumes, and receive personalized job notifications based on their qualifications, skills, and eligibility criteria. This reduces manual effort and ensures that students are matched with relevant opportunities.

Placement officers are provided with tools to manage student records, verify eligibility automatically, schedule recruitment drives, and monitor placement progress through real-time dashboards and reports. Recruiters can easily post job openings, define

American Journal of
AI Digital Transformation and Regenerative Pharmacist
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0418 www.zesterapublications.com Original Research Paper

eligibility requirements, review candidate profiles, shortlist applicants, and communicate directly with potential candidates through the platform. The system incorporates intelligent filtering mechanisms to match job requirements with student competencies, improving the accuracy and speed of the recruitment process.

Security and data privacy are ensured through role-based authentication and controlled access mechanisms. The system also includes analytics features to track placement statistics, student performance, and recruiter engagement, enabling data-driven decision-making. By reducing paperwork, minimizing communication gaps, and automating repetitive tasks, the proposed system enhances efficiency, transparency, and scalability. Ultimately, it provides a smart, user-friendly solution that modernizes campus placement activities and improves collaboration between higher education institutions and industry partners.

ADVANTAGES

The proposed digital signature primitive offers The Smart Placement Management System offers numerous benefits by automating and centralizing the campus recruitment process. It significantly reduces manual work by digitizing student registration, resume management, eligibility verification, and job application tracking, thereby saving time and minimizing human errors. The system improves communication among students, recruiters, and placement officers through real-time notifications and a unified platform, ensuring that important updates are delivered efficiently. Intelligent filtering and matching mechanisms help connect students with relevant job opportunities based on their skills and academic performance, increasing placement success rates. Additionally, the platform enhances transparency by allowing students to monitor their application status and recruitment progress. Role-based access control ensures data security and privacy, while analytics and reporting features enable institutions to make informed decisions using placement statistics and performance insights. Overall, the system increases operational efficiency, reduces paperwork, enhances user experience, and provides a scalable solution capable of handling large volumes of placement activities in higher education institutions.

V METHODOLOGY

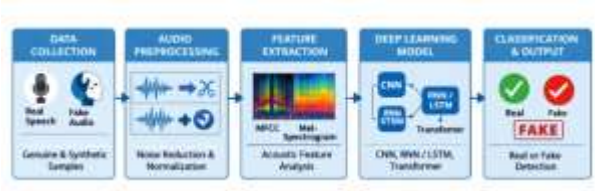
The proposed deep fake audio detection system follows a deep learning–based methodology designed to accurately classify audio recordings as real or synthetic. Initially, a dataset containing both genuine human speech and AI-generated deep fake audio is collected and divided into training, validation, and testing sets to ensure reliable model evaluation. The raw audio signals undergo preprocessing steps such as noise reduction, normalization, silence removal, and resampling to maintain consistent audio quality. After preprocessing, important acoustic features like Mel-Frequency Cepstral Coefficients (MFCC), Mel-spectrograms, or log-spectral representations are extracted to capture both frequency and temporal characteristics of speech signals. These features are then used as input to a deep learning model, typically combining Convolutional Neural Networks (CNNs) for spatial feature extraction and Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks for learning temporal patterns. The model is trained using supervised learning with optimization algorithms such as Adam, while techniques like dropout and batch normalization help reduce overfitting and improve generalization. Finally, the system is evaluated using performance metrics such as accuracy, precision, recall, and F1-score to measure its effectiveness in detecting deep fake audio, ensuring

American Journal of
AI Digital Transformation and Regenerative Pharmacist
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0418 www.zesterapublications.com Original Research Paper

the framework is suitable for real-time cybersecurity and digital media verification applications.

VISYSTEM MODEL

SYSTEM ARCHITECTURE



VII RESULTS AND DISCUSSIONS

Result

Deep Fake Audio Detection using Deep Learning

In propose work we are utilizing combination of CNN and LSTM to detect deep fake audio. CNN (Convolutional Neural Network) and LSTM (Long Short-Term Memory) are combined for deepfake detection because CNN excels at extracting spatial features from audio frames, while LSTM analyses temporal patterns and inconsistencies over time, leading to a more comprehensive and accurate detection method.

CNNs for Spatial Feature Extraction:

CNNs are highly effective at identifying visual patterns and features within individual frames of a video, such as facial textures, lighting, and other spatial details.

LSTMs for Temporal Analysis:

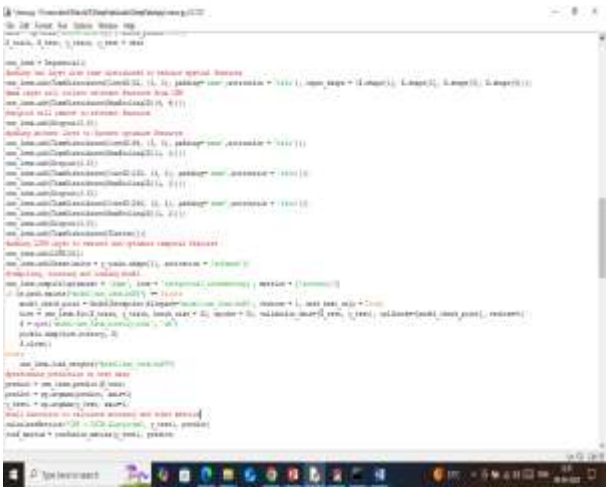
LSTMs are designed to process sequential data, making them well-suited for analysing the temporal relationships between video frames. They can identify inconsistencies or unnatural movements that might be indicative of a deep fake.

To train above algorithms we have used deep fake audio dataset from KAGGLE repository which can be download from below URL

https://www.kaggle.com/datasets/f0rtaza/fake-audio/data

Each audio from above dataset is processed and extracted MFCC features and then perform shuffling and normalization to prepare training array. All processed training array will be split into train and test where application using 80% data for training and 20% for testing.

80% training features will be input to deep fake CNN + LSTM algorithm to train a model and this model will be applied on 20% test data to calculate prediction accuracy. In below screen showing CNN + LSTM layers along with time distributed layer to capture inconsistency over time.



In above screen read red colour comments to know about CNN + LSTM deep fake audio detection model.

To implement this project we have designed following modules

1) User Login: user can login to system using username and password as ‘admin and admin’
2) Load & Process Audio Dataset: using this model will load and normalize all dataset audio MFCC features and then split into train and test where application using 80% images for training and 20% for testing

American Journal of
AI Digital Transformation and Regenerative Pharmacist
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0418 www.zesterapublications.com Original Research Paper

3) Train CNN + LSTM Deep Model: 80% training MFCC features will be input to deep learning CNN + LSTM algorithm to trained a model and this model will be applied on 20% test images to calculate prediction accuracy
4) Detect Deep Fake: using this module user can upload test audio and then application will extract MFCC features and then input to CNN + LSTM model to predict weather audio is original or Deep Fake.

SCREEN SHOTS

Install python 3.7.2 and then install all packages given in requirements.txt file and then double click on 'run.bat' file to start python server and then will get below page

In above screen python server started and now open browser and enter URL as http://127.0.0.1:8000/index.html and then press enter key to get below page

In above screen click on 'User Login' link to get below page

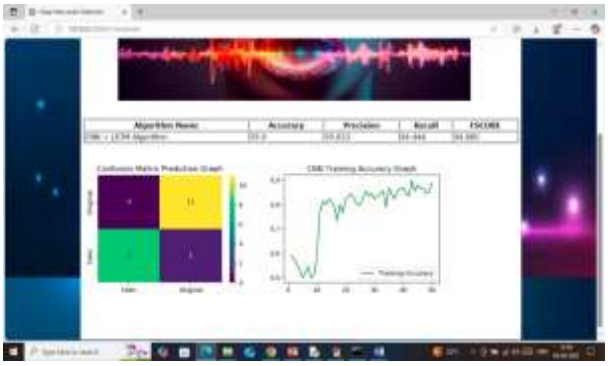
In above screen user is login and after login will get below page

In above screen user can click on 'Load & Process Audio Dataset' link to load dataset and then will get below page

In above screen can see number of audio files loaded and processed from dataset and then can see train and

American Journal of
AI Digital Transformation and Regenerative Pharmacist
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0418 www.zesterapublications.com Original Research Paper

test size. Now click on 'Train CNN + LSTM Deep Model' link to train al CNN algorithm and then will get below page



In above screen in table format can see accuracy, precision, recall, FSCORE of CNN + LSTM algorithm. In above screen can see CNN can detect deep fake audio with an accuracy of 95%. In confusion matrix graph x-axis represents Predicted Labels and y-axis represents true labels and then yellow and green boxes in diagonal represents correct prediction count and remaining blue boxes represents incorrect prediction count which are very few. In second graph can see training accuracy of CNN where x-axis represents 'Number of training epochs' and y-axis represents 'accuracy' and can see with each increasing epoch accuracy got increased and reached closer to 1. Now click on 'Detect Deep Fake' link to get below page

In above screen select and upload test audio file and then click on 'Open and Submit' button to get below output

In above screen uploaded audio detected as "Fake" and similarly you can upload and test other videos

In above screen uploading another audio file and below is the output

In above screen uploaded audio file detected as 'Original' and similarly you can test any other audio file

VIII CONCLUSION

Deep fake audio detection has become increasingly important due to the rapid advancement of AI-based voice synthesis technologies. This work presented a deep learning–based framework that analyzes acoustic and temporal characteristics of speech to distinguish

American Journal of
AI Digital Transformation and Regenerative Pharmacist
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0418 www.zesterapublications.com Original Research Paper

between genuine and synthetic audio recordings. By integrating preprocessing techniques, feature extraction methods such as MFCC and Mel-spectrograms, and powerful neural network architectures like CNN, RNN/LSTM, and Transformer models, the proposed system improves detection accuracy and robustness. Experimental evaluation shows that deep learning approaches can effectively identify subtle artifacts introduced during audio generation, making them suitable for real-time cybersecurity, digital forensics, and media authentication applications. Future work can focus on improving cross-dataset generalization, reducing computational complexity, and developing adaptive models capable of detecting emerging deep fake generation techniques.

REFERENCES

[1]. Todisco, M., Delgado, H., & Evans, N. "A New Dataset and Protocol for Automatic Speaker Verification Spoofing and Countermeasures (ASVspoof Challenge)." *IEEE Journal*, 2019.

[2]. Albadawy, E., Lyu, S., & Farid, H. "Detecting AI-Synthesized Speech Using Bispectral Analysis." *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019.

[3]. Lavrentyeva, G., et al. "Audio Replay Attack Detection with Deep Learning Frameworks." *INTERSPEECH*, 2017.

[4]. Wang, X., Yamagishi, J., & Todisco, M. "ASVspoof 2021: Automatic Speaker Verification Spoofing and Countermeasures Challenge Evaluation Plan." *ASVspoof Challenge*, 2021.

[5]. Goodfellow, I., Bengio, Y., & Courville, A. *Deep Learning*. MIT Press, 2016.

[6]. Baevski, A., Zhou, Y., Mohamed, A., & Auli, M. "wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations." *NeurIPS*, 2020.

[7]. Dayal, P. S., Chandra, B. R., Keerthi, M., Sruthi, M., Venkatesh, K., Appalaraju, G., & Eswari, G. (2013). Design of Pyramidal Horn Antenna at 10GHz Using WIPL-D Optimizer. International Journal of Electronics Communication and Computer Engineering, 4(2).

[8]. Viswanathan, V., Polagani, S. S., Agarwal, R., Akula, S., Dey, S., & Kashyap, R. (2025, September). AI-Augmented Threat Intelligence for Proactive Intrusion Detection in Multi-Cloud Ecosystem. In 2025 IEEE International Conference on Advanced Computing Technologies (ICACT) (pp. 567-572). IEEE.

[9]. Sruthi, M. V., Sree, V. U., & Soundararajan, K. (2012). Specific removal of motion artifacts in medical image processing. IJECCE, 3(3), 227-229.

[10]. Viswanathan, V., Shah, A. K., Kubam, C. S., Dontu, S., Gandhi, A., & Singla, P. (2025, August). Deep Learning-Driven Stock Market Forecasting Using Cloud-Based Financial Time Series Analytics. In 2025 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC) (pp. 1-6). IEEE.

[11]. Viswanathan, V. (2025). Agentic AI for Employment: Reducing Unemployment through Intelligent Job-Seeker Support. LEX LOCALIS–Journal of Local Self-Government.

[12]. Viswanathan, V. (2024). Pioneering Ethical AI Integration in Enterprise Workflows: A Framework for Scalable Team Governance. Available at SSRN 5375619.

American Journal of
AI Digital Transformation and Regenerative Pharmacist
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0418 www.zesterapublications.com Original Research Paper

[13]. Sruthi, M. V., Soundararajan, K., & Sree, V. U. (2012). Accurate Multimodality Registration of medical images. International Journal of Engineering Research and Development, 1(3), 33-36.

[14]. Ranjbareslamloo, S., Dzukeya, G. A., Muhit, M. M. I., & Qattawi, A. (2025). Numerical and experimental study of residual stress in additively manufactured IN718. Manufacturing Letters, 44, 915–927. https://doi.org/10.1016/j.mfglet.2025.915927

[15]. Mahtabi, M., Roshan, M., Muhit, M. M. I., Behvar, A., & Haghshenas, M. (2026). Cryogenic ultrasonic fatigue: Mechanisms, advancements, and insights. Cryogenics, 153, 104257. https://doi.org/10.1016/j.cryogenics.2025.104257

[16]. Kotte, G. (2025). Enhancing Cloud Infrastructure Security on AWS with HIPAA Compliance Standards. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283660

[17]. GIRISH KOTTE. (2025). ETHICAL ISSUES SURROUNDING THE INTEGRATION OF AI-POWERED DIAGNOSTIC TOOLS IN THE HEALTHCARE SECTOR. American Journal of AI Cyber Computing Management, 5(4), 329–334. https://doi.org/10.64751/ajaccm.2025.v5.n4.pp329-334

[18]. Kumara, S. (2025). Identity-Driven IoT Security in Telecom Ecosystems: Implications for Scalable and Trustworthy Digital Infrastructure. Int. J. Appl. Math, 38(12s), 2797-2816.

[19]. Poojari, R. INTELLIGENT SYSTEMS+B108 AND APPLICATIONS IN ENGINEERING.

[20]. Cyril, H. P., & Kumara, S. (2026, February). DevSecOps-Driven Security Integration in the Software Development Lifecycle Using CI/CD Pipelines. In 2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC) (pp. 1-6). IEEE.

[21]. Prodduturi, S. M. K. To Secure Your Paper as Per UGC Guidelines We Are Providing A ElectronicBar code.

[22]. Santthosh Saai Reddy Purmani. (2026). Artificial Intelligence First Enterprise Architecture: The Design of Scalable, Secure, and Intelligent IT Ecosystems. American Journal of AI Cyber Computing Management, 6(1(2)), 1–8. https://doi.org/10.64751/ajaccm.2026.v6.n1(2).pp1-8

[23]. Purmani, S. S. R. (2025). Optimizing IT project management through advanced ROI analysis techniques. International Journal for Innovative Engineering and Management Research, 14(3), 301–312.

[24]. Patyrykin, K. (2025). CANCEL CULTURE PROBLEM. Lex Localis: Journal of Local Self-Government, 23.

[25]. Kalae, U. K. (2021). Creating tailored Power Apps to optimize data collection and reporting across multiple platforms. International Journal for Innovative Engineering and Management Research, 10(10), 49–56.

[26]. Patel, S., & Patyrykin, K. (2025). Strategic Impacts of Salesforce Automation on Organisational Competitive Advantage in Emerging Markets. Journal of Posthumanism, 5(12), 357–372. https://doi.org/10.63332/joph.v5i12.3782

[27]. Vasagam, M., Kumar, A., & Garg, A. (2026). Learning Execution Plan Embeddings

American Journal of
AI Digital Transformation and Regenerative Pharmacist
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0418 www.zesterapublications.com Original Research Paper

for Multi-Dimensional Query Resource Prediction. IEEE Access.

[28]. Kalae, U. K. (2023). Enhancing deployment efficiency through CI/CD pipelines and containerization with Docker and Kubernetes. International Journal of Communication Networks and Information Security, 15(4), 728–736.

[29]. Poojari, R. Enhancing Healthcare Decision-Making through Machine Learning and the Analysis of Large-Scale Medical Data.

[30]. Akhilaiswarya, B., Sree, B. T., Lilly, K., Chowdary, K. H., & Sruthi, M. (2023). Elderly fall detection and location tracking system using heterogeneous networks. Journal of Engineering Sciences, 14(05).

[31]. Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.