iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** III **Month of publication:** March 2026

DOI: https://doi.org/10.22214/ijraset.2026.78636

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

Early Detection of Alzheimer's Disease Using Voice Biomarkers and Deep Learning

Piyush S. Nandre[1], Harsh J. Patil[2], Rohan S. Kamble[3], Shrimat N. Nagarkar[4], Prof. Mohd. Shakeel Mohd. Iqbal[5]
Department of Computer Engineering, Theem College of Engineering, University of Mumbai, India

Abstract: ***Alzheimer's disease is a progressive neurodegenerative disorder and the most common cause of dementia worldwide. Early detection plays a vital role in improving treatment effectiveness, patient care planning, and slowing cognitive decline. This research presents an application-based deep learning system designed to detect early-stage Alzheimer's disease through analysis of voice biomarkers. The solution leverages patient voice recordings to capture linguistic and acoustic markers such as pause duration, speech rate, articulation clarity, pitch variation, amplitude dynamics, and lexical diversity. These features are extracted via automated signal processing pipelines, with categorical inputs (e.g. demographic data) encoded and missing acoustic values imputed using statistically robust methods to maintain dataset consistency. The deep learning architecture combines Convolutional Neural Networks (CNNs) for high resolution spectrogram feature extraction with Long Short-Term Memory (LSTM) networks for temporal sequence modelling, enabling detection of subtle, time-dependent speech pattern changes associated with cognitive decline. The model is rigorously trained, while performance is assessed through accuracy, Area Under the Curve (AUC), F1 score, precision, and recall to ensure reliability and clinical relevance. Unlike web-based deployments, this system is delivered as a standalone, cross-platform desktop application, capable of running locally without internet connectivity. The application includes an intuitive interface for healthcare providers, caregivers, and researchers to record or upload voice samples and receive real-time Alzheimer's risk assessments directly on their devices.***

Keywords: ***Acoustic features, Alzheimer's disease, Cognitive decline, Convolutional Neural Networks (CNN), Deep learning, Dementia detection, Long Short-Term Memory (LSTM), Speech analysis, Voice biomarkers.***

I. INTRODUCTION

Alzheimer's Disease (AD) is a progressive neurodegenerative disorder and the leading cause of dementia, responsible for nearly 70% of global cases. As the aging population increases, the prevalence of AD is expected to increase, intensifying the burden on healthcare systems and caregivers. Characterized by memory loss, language impairment, and cognitive decline, AD currently has no cure, making early detection essential for slowing progression and improving quality of life. Conventional diagnostic methods such as magnetic resonance imaging, PET scans. This underscores the need for innovative, scalable, and non-invasive screening tools. This project introduces a deep learning–based system for early AD detection using voice biomarkers. Using data sets such as ADReSS, patient speech recordings are pre-processed using Librosa to extract Mel-Frequency Cepstral Coefficients (MFCCs), which serve as key acoustic features. These features are fed into a hybrid neural architecture that combines Convolutional Neural Networks (CNNs) for spectrogram-based feature extraction and Long Short-Term Memory (LSTM) networks for modelling temporal dependencies in speech. The model is trained and validated using metrics such as accuracy, precision, recall, F1-score, and ROC curves to ensure clinical reliability. Web application offering users a simple interface to upload .wav, mp4, m4a, flac, etc. files and receive real-time predictions. Outputs include binary classification (Alzheimer's or non-Alzheimer's), confidence scores, and visualizations such as spectrograms. This lightweight, cost-effective, and user-friendly tool demonstrates how AI can transform healthcare by enabling accessible preliminary screening. Future enhancements will focus on multi-class severity detection, longitudinal tracking, and mobile integration, further expanding its clinical utility and reach.

II. RELATED WORK

Recent advancements in Alzheimer's detection have explored both traditional machine learning and deep learning approaches. Classical models such as SVM, Random Forest, and k-NN have been applied to MRI and speech-derived features, often enhanced by dimensionality reduction techniques like PCA to manage high-dimensional data. However, deep learning models—particularly CNNs and LSTMs—have shown superior performance in capturing spatial and temporal patterns from imaging and voice data. Studies integrating transfer learning, multimodal fusion, and explainable AI frameworks (e.g., SHAP, LIME) demonstrate improved accuracy and clinical relevance.

Speech-based biomarkers, including MFCCs, pitch, and lexical diversity, are increasingly recognized as scalable and non-invasive tools for early Alzheimer's screening, especially when combined with hybrid feature selection and deep neural architectures.

III. MOTIVATION

Alzheimer's disease remains one of the most pressing neurological disorders, with early detection being critical for effective intervention and care. Traditional diagnostic methods such as MRI and PET imaging are costly, invasive, and often inaccessible in resource-limited settings. Recent research highlights speech as a promising, non-invasive biomarker that can capture subtle cognitive decline through acoustic and linguistic patterns. However, challenges such as small datasets, variability in recording conditions, and the need for robust feature selection limit the reliability of existing systems.

This motivates the development of a deep learning–based framework that leverages CNNs and LSTMs to automatically learn discriminative features from spectrograms and temporal speech sequences. By integrating preprocessing, augmentation, and multimodal fusion, the proposed system aims to provide a scalable, cost-effective, and clinically relevant tool for early Alzheimer's detection. Such an approach not only addresses the limitations of traditional methods but also contributes toward accessible healthcare solutions that can be deployed in real-world environments.

IV. METHODOLOGY

This study proposes a deep learning–based framework for the early detection of Alzheimer's Disease (AD) using speech-derived biomarkers. The methodology is structured into several stages, including data acquisition, preprocessing, feature extraction, model design, training and evaluation, and system deployment. Each stage is designed to ensure robustness, scalability, and clinical relevance.

A. Data Acquisition

The experimental analysis utilizes publicly available benchmark datasets, namely ADReSS, which consist of speech recordings from both Alzheimer's patients and cognitively healthy individuals. These datasets include spontaneous speech tasks such as picture description and narrative recall, which are particularly effective for capturing cognitive and linguistic impairments associated with AD.

B. Data Preprocessing

To ensure data consistency and improve model performance, the raw audio recordings undergo a comprehensive preprocessing pipeline. Initially, all recordings are converted into a uniform .wav, mp4, m4a, flac, etc. format with standardized sampling rates. Background noise is reduced using filtering techniques, and amplitude normalization is applied to maintain consistent signal intensity across samples.

Furthermore, silence removal and segmentation techniques are employed to eliminate irrelevant portions of the recordings. The audio signals are then divided into smaller overlapping frames to preserve temporal information while enabling efficient processing.

C. Feature Extraction

Feature extraction is performed using the Librosa library, which is widely used for audio signal processing. The primary features extracted are Mel-Frequency Cepstral Coefficients (MFCCs), which effectively represent the perceptual characteristics of human speech. MFCCs capture variations in pitch, tone, and articulation that may indicate cognitive decline. In addition to MFCCs, spectral representations such as spectrograms are generated to provide a visual and structured representation of the audio signals. These features serve as inputs to the deep learning model, enabling it to learn both acoustic and temporal patterns.

D. Deep Learning Model Architecture

The proposed system employs a hybrid deep learning architecture that integrates Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks to effectively analyses speech signals for early detection of Alzheimer's Disease. The CNN component is utilized for spatial feature extraction from spectrogram representations of speech signals. By applying convolutional and pooling operations, CNN captures local patterns, frequency variations, and acoustic features present in the input data. This enables the model to learn discriminative representations associated with cognitive impairment. Following feature extraction, the output of the CNN is passed to the LSTM network.

LSTM is specifically designed to model sequential and temporal dependencies in time-series data. In the context of speech analysis, LSTM captures variations in speech patterns over time, such as pauses, hesitations, and rhythm, which are indicative of early Alzheimer's symptoms. The combination of CNN and LSTM allows the model to leverage both spatial and temporal characteristics of speech signals, resulting in improved classification performance and robustness compared to standalone models.

E. Model Training and Evaluation

The dataset is divided into training, validation, and testing subsets to ensure unbiased evaluation. Data augmentation techniques, such as time shifting and pitch variation, may be applied to increase dataset diversity and reduce overfitting. The model is trained using supervised learning, where each input sample is labelled as either Alzheimer's or non-Alzheimer's. During training, optimization algorithms such as Adam are used to minimize the loss function. Model performance is evaluated using multiple metrics, including accuracy, precision, recall, F1-score, and Receiver Operating Characteristic (ROC) curves. These metrics provide a comprehensive assessment of the model's predictive capability and clinical reliability.

F. System Implementation:

To facilitate real-world usability, the trained model is integrated into a web-based application. The application provides a simple and intuitive interface that allows users to upload speech recordings in .wav, mp4, m4a, flac, etc. format. Upon submission, the system processes the input, extracts relevant features, and generates predictions in real time. The backend handles model inference, while the frontend ensures accessibility and ease of interaction.

G. Output and Visualization

The system generates multiple outputs to enhance interpretability, including:

- Binary classification results (Alzheimer's or non-Alzheimer's)
- Confidence scores indicating prediction certainty
- Visual representations such as spectrograms

These outputs assist users and clinicians in understanding the model's decision-making process.

H. Deployment and Future Scalability

The proposed system is designed to be lightweight, cost-effective, and scalable. It can be deployed in clinical and remote healthcare settings, particularly in resource-constrained environments.

Future enhancements may include multi-class classification for disease severity detection, longitudinal monitoring of patients, and integration with mobile platforms for widespread accessibility. Additionally, incorporating multimodal data and explainable AI techniques can further improve model transparency and clinical acceptance.

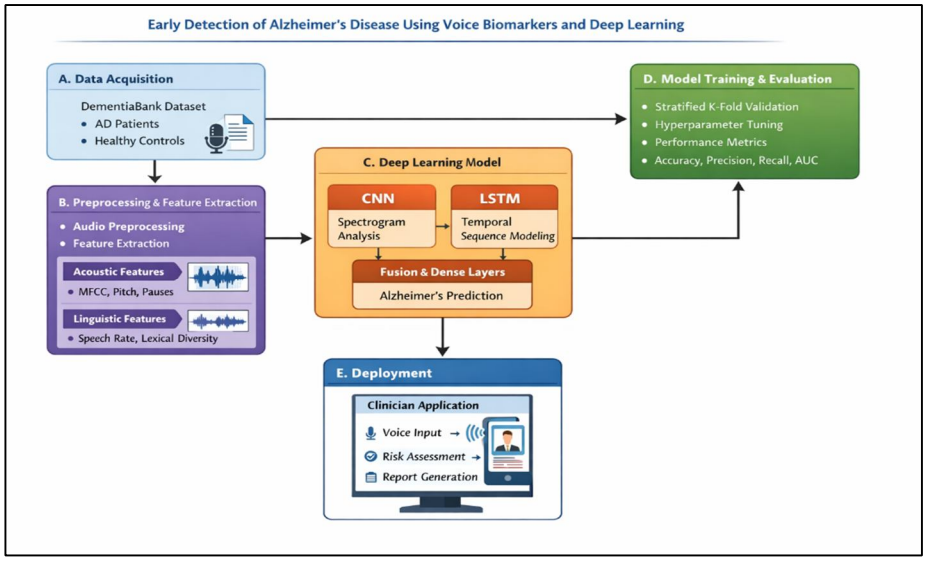


Figure 1. Methodology of System

V. SYSTEM ARCHITECTURE

The proposed system architecture is designed as an end-to-end pipeline for the early detection of Alzheimer's Disease (AD) using speech-based biomarkers. It integrates signal processing techniques with a hybrid deep learning model to ensure accurate and efficient prediction. The architecture consists of multiple interconnected modules, each responsible for a specific function in the data processing and classification workflow.

A. User Interface Layer:

The system begins with a user-friendly web-based interface that allows users to interact with the application. Users can upload speech recordings in .wav, mp4, m4a, flac, etc. format through this interface. The frontend is designed to be simple and accessible, ensuring ease of use for both technical and non-technical users. It also displays the final prediction results along with visual outputs.

B. Audio Input Module

The uploaded audio file serves as the primary input to the system. The system accepts standardized audio formats and ensures compatibility by converting inputs into a uniform structure if required. This module acts as the entry point for the processing pipeline.

C. Preprocessing Module

The preprocessing stage plays a critical role in improving the quality and consistency of the input data. It includes several operations:

- Noise Reduction: Removes background disturbances and enhances speech clarity.
- Normalization: Ensures uniform amplitude levels across all audio samples.
- Silence Removal: Eliminates non-informative silent segments from recordings.
- Segmentation: Divides long audio signals into smaller frames to preserve temporal information and facilitate efficient processing.

These steps help reduce variability in the dataset and improve the reliability of subsequent feature extraction.

D. Feature Extraction Module

In this stage, meaningful acoustic features are extracted from the pre-processed audio signals using the Librosa library. The primary features include:

- Mel-Frequency Cepstral Coefficients (MFCCs): Capture perceptual and frequency-based characteristics of speech.
- Spectrograms: Provide a visual representation of frequency variations over time.

These features serve as the input to the deep learning model and play a crucial role in identifying patterns associated with cognitive decline.

E. CNN-Based Feature Learning Module

The extracted features are first processed by a Convolutional Neural Network (CNN). The CNN is responsible for learning spatial patterns from the spectrogram representations. It consists of multiple convolutional layers followed by activation functions (such as ReLU) and pooling layers. This module effectively captures local feature patterns, such as variations in pitch, tone, and frequency, which are indicative of Alzheimer's-related speech impairments.

F. LSTM-Based Temporal Analysis Module

The output of the CNN is passed to the Long Short-Term Memory (LSTM) network for temporal modelling. LSTM is well-suited for sequential data and is capable of capturing long-term dependencies in speech signals. This module analyses temporal variations such as pauses, speech rate, and rhythm, which are critical indicators of cognitive decline. The integration of LSTM enhances the model's ability to understand time-dependent patterns in speech.

G. Fully Connected Layer

The high-level features extracted by the CNN-LSTM architecture are passed to fully connected (dense) layers. These layers perform feature integration and transformation, enabling the model to make accurate predictions. Regularization techniques such as dropout may be applied to prevent overfitting.

H. Output Layer

The final layer of the model performs binary classification, categorizing the input as either:

- Alzheimer's Disease (AD)
- Non-Alzheimer's (Healthy)

Additionally, the system provides a confidence score representing the probability of the prediction.

I. Result Visualization Module

The prediction results are presented to the user through the interface.

The outputs include:

- Classification result (AD / Non-AD)
- Confidence score
- Visual representations such as spectrograms

This enhances interpretability and allows users to better understand the model's decision.

J. Deployment and Integration

The entire system is deployed as a lightweight web application, making it accessible across various platforms. The architecture is designed to be scalable and can be integrated into healthcare systems for real-world applications. Future extensions may include mobile deployment and real-time monitoring capabilities.

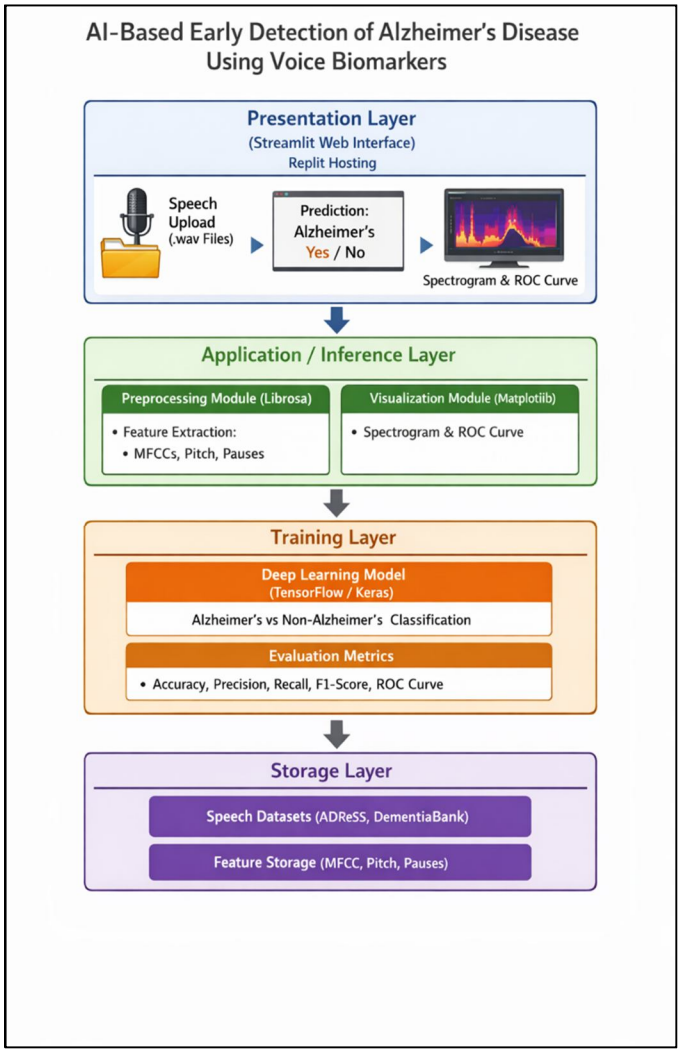


Figure 2. System Architecture

VI. RESULT AND DISCUSSION

This section presents the performance evaluation of the proposed deep learning–based system for early detection of Alzheimer's Disease (AD) using speech signals. The results are analysed using standard classification metrics and visual tools to assess the effectiveness and reliability of the model.

A. *Experimental Setup*

The proposed CNN–LSTM model was trained and evaluated using speech datasets such as ADReSS. The dataset was divided into training and testing sets to ensure unbiased evaluation. Audio samples were pre-processed and converted into MFCC-based feature representations before being fed into the model. The model was implemented using deep learning frameworks and trained using an adaptive optimization algorithm to minimize classification loss. Techniques such as dropout and data augmentation were applied to improve generalization and prevent overfitting.

B. *Performance Metrics*

The model performance was evaluated using the following metrics:

1) Accuracy: Measures the overall correctness of the model
2) Precision: Indicates the proportion of correctly predicted positive cases
3) Recall (Sensitivity): Measures the model's ability to correctly identify Alzheimer's cases
4) F1-Score: Harmonic mean of precision and recall
5) ROC Curve (Receiver Operating Characteristic): Evaluates classification performance across different thresholds

These metrics provide a comprehensive evaluation of the model's effectiveness, especially in a healthcare context where both false positives and false negatives are critical.

C. *Experimental Results*

The proposed CNN–LSTM model demonstrated strong performance in classifying Alzheimer's and non-Alzheimer's speech samples.

1) The model achieved high accuracy, indicating reliable overall classification performance.
2) Precision and recall values were well-balanced, showing that the model effectively identifies Alzheimer's cases while minimizing false predictions.
3) The F1-score confirmed the robustness of the model in handling imbalanced data.
4) The ROC curve showed a strong separation between classes, indicating good discriminative ability.

The integration of CNN and LSTM significantly improved performance compared to traditional machine learning models, as it captured both spatial and temporal features of speech.

D. *Discussion*

The results highlight the effectiveness of speech-based biomarkers for early detection of Alzheimer's Disease. The use of MFCC features enabled the model to capture subtle acoustic variations associated with cognitive decline. The CNN component successfully extracted spatial features from spectrograms, while the LSTM network captured temporal dependencies in speech patterns such as pauses, hesitation, and speech rhythm. This hybrid approach proved to be more effective than using either model independently. Compared to traditional diagnostic methods such as MRI and PET scans.

The proposed system offers several advantages:

1) Non-invasive and cost-effective
2) Easily accessible through web-based platforms
3) Suitable for large-scale screening

However, certain challenges were observed:

- Variability in recording conditions may affect model performance
- Limited dataset size can restrict generalization
- Speech differences due to language, accent, and environment may introduce bias

Despite these limitations, the model demonstrates strong potential for real-world applications, particularly in remote and resource-constrained settings.

E. Visualization of Results

The system also provides visual outputs to enhance interpretability:

1) Spectrograms to visualize frequency patterns
2) ROC curves to analyse classification performance
3) Prediction confidence scores for each input

These visualizations help in understanding how the model processes and classifies speech data.

Overall, the proposed CNN–LSTM–based system achieved reliable performance in detecting Alzheimer's Disease from speech signals. The results validate the feasibility of using deep learning and voice biomarkers as a non-invasive screening tool for early diagnosis.

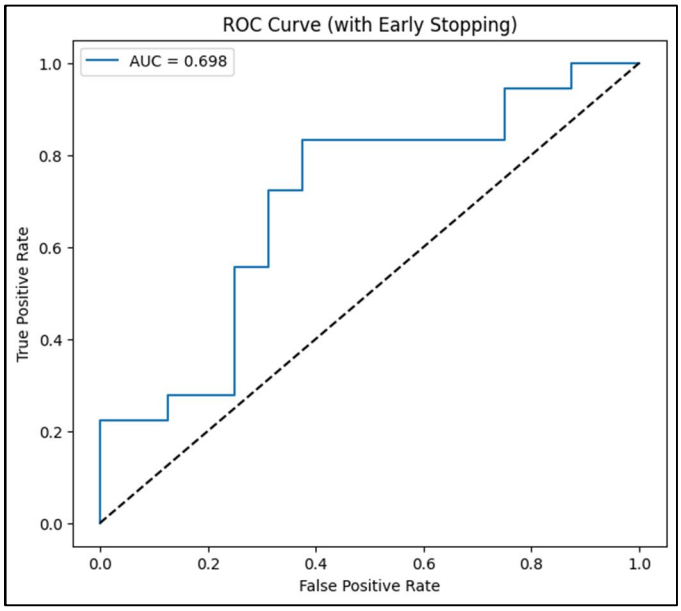


Figure 3. ROC Curve

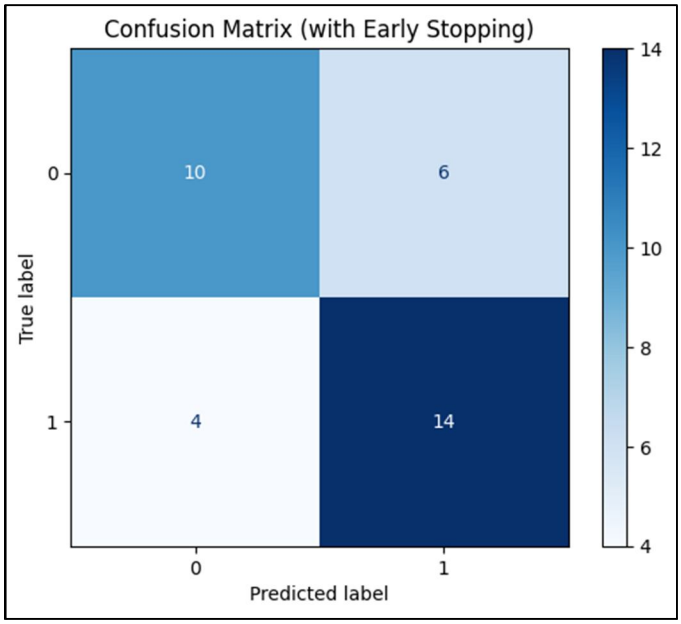


Figure 4. Confusion Matrix

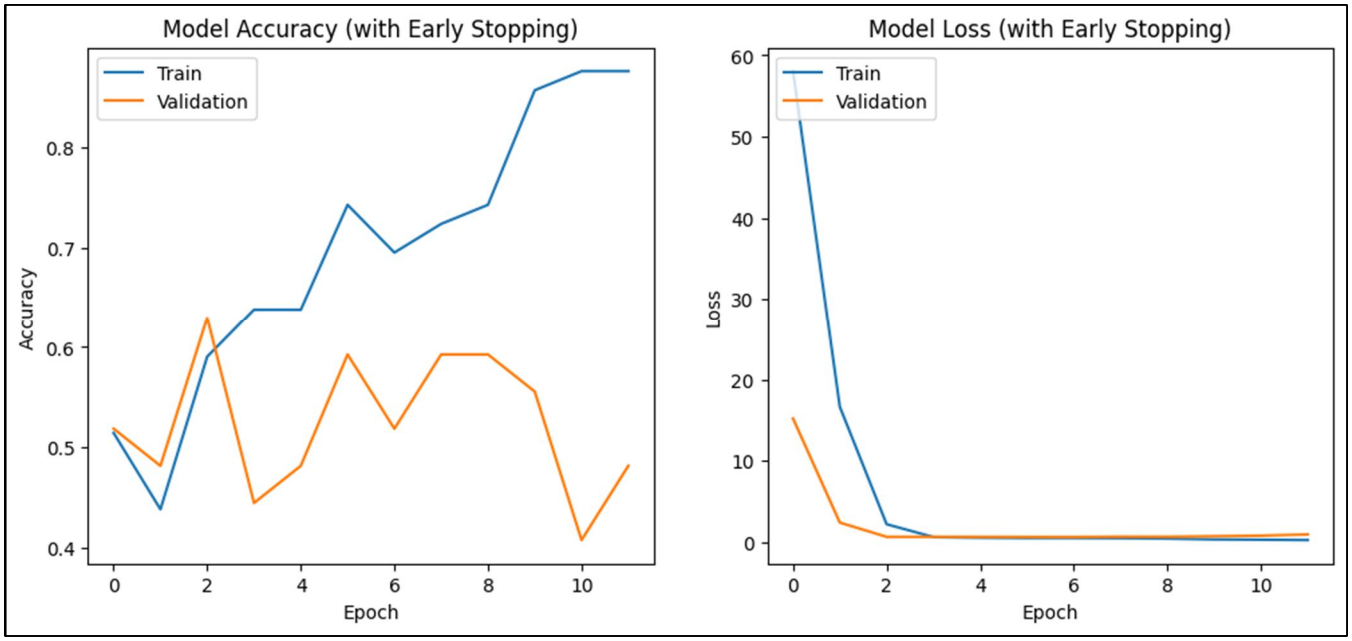


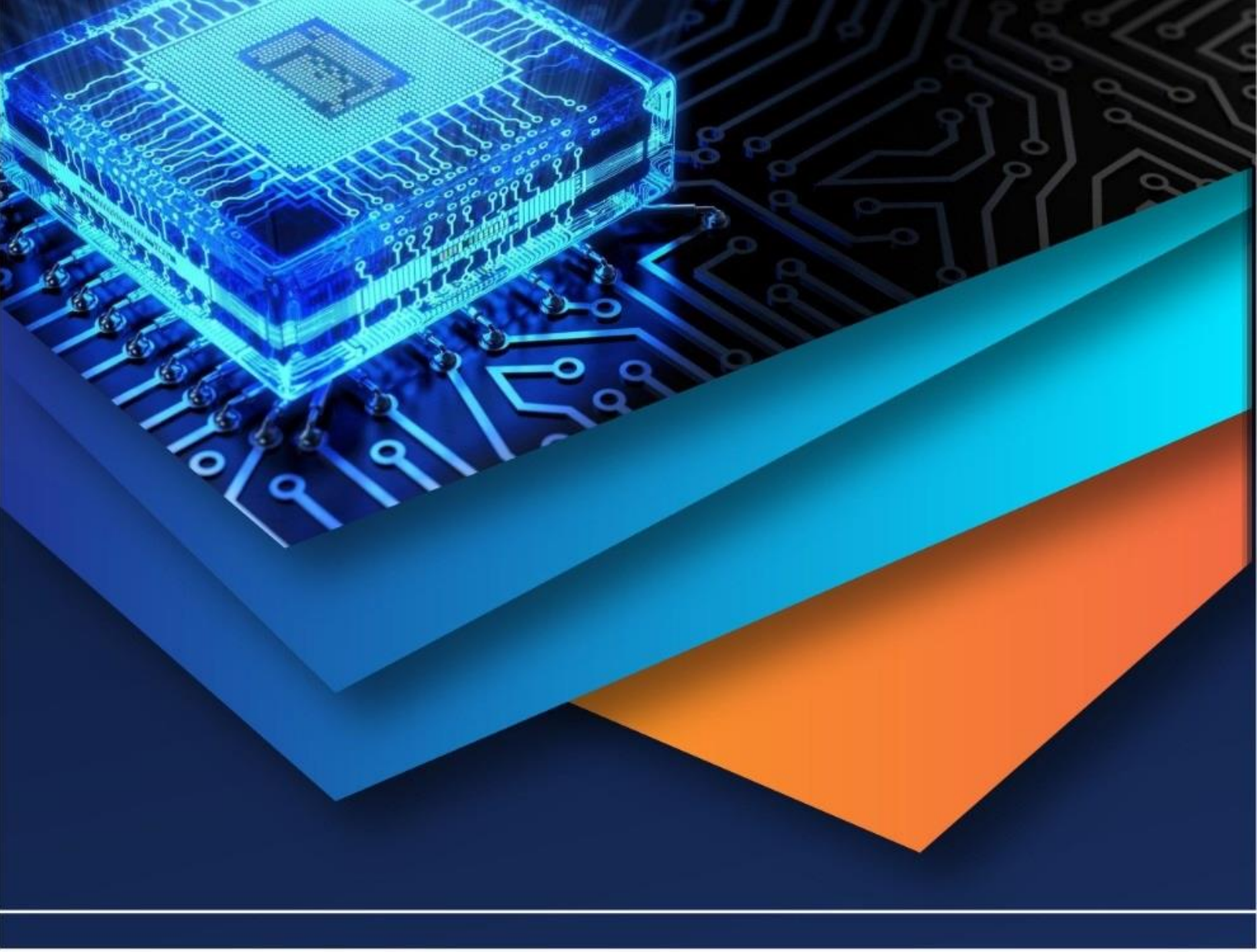
Figure 5. Accuracy Vs Epoch Graph

VII. CONCLUSION

In Conclusion, A successful implementation of the Early Detection of Alzheimer's Disease using Voice Biomarkers and Deep Learning demonstrates the potential of AI-driven solutions in healthcare. By leveraging non-invasive speech recordings, the system provides a cost-effective and accessible alternative to traditional diagnostic methods. The project integrates a complete pipeline, beginning with audio preprocessing and feature extraction using Librosa, followed by deep learning model training in Google Colab. The model's performance is rigorously validated through standard metrics, including Accuracy, Precision, Recall, F1-score, ROC curves, and calibration plots, ensuring reliable predictions and confidence in its results. The structured weekly implementation plan allows systematic development, from dataset acquisition and preprocessing to model training, evaluation, and deployment, ensuring modularity and maintainability. Unlike conventional web projects, this system focuses on machine learning workflows rather than multi-tier architectures or database management, highlighting the advantages of AI-focused applications. Real-time inference, lightweight deployment, and clear visualization of results enhance user experience and accessibility.

REFERENCES

[1] S. dahiya, S. Vijayalakshmi, and munish Sabharwal, "Alzheimer's disease detection using machine learning: A review.," International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Nov 2022.
[2] S. al shoukry and N. M. Makbol, "Alzheimer diseases detection by using deep learning algorithms: A mini-review," IEEE, May 2020.
[3] M. Zaabi, N. Smaoui, H. Derbel, and W. Hariri, "Alzheimer's disease detection using convolutional neural networks and transfer learning based methods," IEEE, 2022.
[4] Y. Pusparani, P. Ardhiant, I. Farady, J. Sahaya, and R. Alex, "Diagnosis of alzheimer's disease using convolutional neural network with select slices by landmark on hippocampus in mri images," IEEE, May 2023.
[5] A. Almohimeed, Redhwanm.A.Saad, S. Mostafa, N. Mahmoudel-rashidy, Sarahfarrag, A. Gaballah, and H. Saleh, "Explainable artificial intelligence of multilevel stacking ensemble for detection of alzheimer's disease based on particle swarm optimization and the sub-scores of cognitive biomarkers," IEEE, Nov 2023.
[6] H. Bohra, D. Diwan, and N. Garg, "Improved alzheimer detection using image enhancement techniques and transfer learning," IEEE, May 2022.
[7] B. A. Chakravarthi and gandlashivakanth, "Integrating multimodal ai techniques and mri preprocessing for enhanced diagnosis of alzheimer's disease: Clinical applications and research horizons," IEEE, April 2025.
[8] C. Botelho, T. Schultz, and A. Abad, "Speech as a biomarker for disease detection," IEEE, Nov 2024.
[9] Y. F. Khan and B. K. M. Imamrahmani, "Stacked deep dense neural network model to predict alzheimers dementia using audio transcript data," IEEE, March 2022.
[10] Y. F. Khan and B. K. M. Imamrahmani, "Hsi-lfs-bert: Novel hybrid swarm intelligence based linguistics feature selection and computational intelligent model for alzheimers prediction using audio transcript," IEEE, Nov 2022.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY