IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.15537

Alzheimer's Disease Detection Using AI

Divyanshi Kashyap[1], Divya pandey[2], Mansi Pandey[3], Divyanshi Yadav[4], Dr. Nikhat Akhtar[5]

Student, Goel Institute of Technology and Management, Lucknow, India[1-4]

Professor, Goel Institute of Technology and Management, Lucknow, India[5]

Abstract: Alzheimer's disease is a progressive neurological condition that impacts memory, thinking ability, and speech functions. Early detection of Alzheimer's is very crucial because timely intervention can help slow the progression of symptoms and enhance the overall quality of life for patients. Conventional diagnostic methods mainly depend on neuroimaging, cognitive assessments, and clinical evaluations, which are often costly, time-consuming, and not easily accessible to all individuals.

Recent studies indicate that speech patterns and language usage can provide early indications of cognitive decline. Individuals affected by Alzheimer's disease frequently show variations in speech fluency, vocabulary diversity, pause pat-terns, and sentence formation. These variations can be exam-ined using computational methods along with machine learning techniques.

In this work, we present a machine learning-based approach for identifying Alzheimer's disease through speech recordings and linguistic analysis. Acoustic features such as Mel Frequency Cepstral Coefficients (MFCC), pitch, energy, speech rate, and pause duration are derived from audio signals. Linguistic features are obtained from transcribed speech using TF-IDF representation.

Multiple models, including text-based, audio-based, and combined feature models, are tested using the XGBoost classifier. The experimental findings indicate that integrating both acoustic and linguistic features leads to a noticeable improvement in prediction performance. The proposed hybrid model attains an accuracy of 76.44

1 INTRODUCTION

Alzheimer's disease (AD) is a chronic neurological disorder that gradually reduces an individual's ability to think, re-member, communicate, and make decisions. With the global increase in life expectancy, the number of people living with Alzheimer's is steadily growing, creating significant emotional and financial challenges for families as well as health-care systems. Early identification of the disease is highly important, as intervention at an initial stage can slow down cognitive deterioration and allow patients to remain independent for a longer duration.

Conventional diagnostic approaches such as MRI, PET scans, and comprehensive clinical assessments are reliable but are often costly, time-intensive, and not widely available in many areas. These challenges have motivated researchers to investigate alternative and non-invasive markers of cognitive decline. Among various options, speech has proven to be a valuable and natural indicator. Since speaking involves memory, language processing, attention, and motor coordination, even minor cognitive impairments can affect a person's way of speaking.

People affected by Alzheimer's commonly exhibit clear variations in both speech production and content. These may include limited vocabulary, slower speaking rate, increased pauses, and simplified sentence construction. With the rapid progress in artificial intelligence, it is now possible to automatically detect and analyze such subtle variations.

In this research, we propose a machine learning frame-work that utilizes both acoustic features of speech and linguistic patterns derived from transcripts to identify signs of Alzheimer's disease. By integrating these two types of information, the proposed system aims to enhance reliability and assist in early-stage screening within real-world healthcare settings.

2 PROBLEM STATEMENT

Alzheimer's disease is considered one of the most difficult cognitive disorders to identify at an early stage. While med-ical imaging techniques and neurological assessments pro-vide accurate results, they depend on specialized equipment, skilled professionals, and considerable financial investment. Such requirements restrict their accessibility, particularly in rural and resource-limited areas. In addition, the initial symptoms of Alzheimer's are usually subtle and often confused with normal aging, leading to late diagnosis and reduced effectiveness of treatment.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.15537

Speech, on the other hand, serves as a natural and easily accessible source of information that reflects underlying cognitive abilities. Difficulties in word retrieval, sentence forma
tion, and speech fluency are commonly observed during the early stages of Alzheimer's disease. However, many existing studies tend to focus on either acoustic features or linguis-tic features separately. Depending on only one type of fea-ture limits the ability to fully capture the range of cognitive changes present in speech.

A further challenge is the limited availability of publicly accessible datasets, which restricts the training of deep learning models that typically require large volumes of data. This situation highlights the need for efficient approaches that can perform effectively even with smaller datasets while still maintaining good accuracy.

Therefore, the main problem addressed in this work is to design a reliable, scalable, and non-invasive system for detecting Alzheimer's disease using speech data. The objective is to combine acoustic and linguistic features within a unified model and apply machine learning techniques to identify patterns that can differentiate healthy individuals from those exhibiting early signs of cognitive decline.

3 RELATED WORK

Earlier studies have widely investigated the application of speech and language analysis for the detection of Alzheimer's disease. Initial research mainly concentrated on linguistic features obtained from speech transcripts. Fraser et al. showed that factors such as lexical diversity, syntac-tic structure, and semantic consistency can effectively differentiate between healthy individuals and those affected by Alzheimer's disease.

Later research extended this work by incorporating acous-tic analysis. Features including pitch variation, speaking rate, articulation rate, and pause duration have been found to in-dicate cognitive impairment. These characteristics represent the temporal and prosodic properties of speech, which are commonly influenced in neurodegenerative disorders.

With the growth of machine learning techniques, several algorithms such as Support Vector Machines (SVM), Random Forest, and Logistic Regression have been applied for classification purposes. Although these models produce encouraging results, they have limitations in capturing complex relationships among multiple features.

Deep learning methods, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have also been utilized for speech-based Alzheimer's detec-tion. CNNs are well-suited for processing spectrogram-based representations of audio signals, whereas RNNs and Long Short-Term Memory (LSTM) networks are effective for han-dling sequential data. However, these approaches generally require large-scale datasets and significant computational re-sources.

More recent studies emphasize the advantages of multi-modal approaches that integrate both acoustic and linguistic features. By combining information from these two domains, such methods achieve better classification performance.

In this study, we employ a hybrid feature strategy along with the XGBoost algorithm, which provides a good trade-off between accuracy and computational efficiency. This makes the proposed approach practical for real-world applications.

4 DATASET DESCRIPTION

The dataset used in this study consists of speech recordings collected from both healthy individuals and patients diagnosed with Alzheimer's disease. The recordings are obtained from a standardized dataset designed for cognitive assessment tasks, ensuring consistency and reliability.

Each audio sample contains spontaneous speech generated during picture description or narrative tasks. These tasks are commonly used in clinical settings to evaluate cognitive and linguistic abilities. The average duration of each recording is approximately 15 seconds, providing sufficient data for fea-ture extraction.

The dataset initially comprised 174 speech recordings, including 116 from healthy individuals and 58 from Alzheimer's patients. The audio files are sampled at a fre-quency of 16 kHz and stored in WAV format. Transcriptions of the speech recordings are also provided, enabling linguis-tic analysis. After preprocessing, segmentation, and feature extraction, the dataset was expanded to a total of 864 sam-ples, which were used for training and evaluation of the machine learning models.

Prior to feature extraction, several preprocessing steps are applied to enhance data quality. These include noise reduction, silence removal, and normalization. Noise reduction techniques are used to eliminate background interference, while silence removal ensures that only meaningful speech segments are analyzed. Normalization standardizes the amplitude of audio signals, improving consistency across sam-ples.

The dataset is relatively balanced and provides a suit-able foundation for training and evaluating machine learn-ing models. However, the limited size of the dataset poses challenges for deep learning approaches, further justifying the use of efficient algorithms such as XGBoost.

Table 1: Dataset Statistics

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.15537

Parameter	Value
Original Samples (Audio Recordings)	174
Healthy Subjects	116
Alzheimer Patients	58
Processed Samples (After Preprocessing)	864
Average Duration	15 seconds
Sampling Rate	16 kHz
Language	Korean

Before feature extraction, several preprocessing steps were performed including noise removal, silence trimming, and normalization.

5 FEATURE EXTRACTION

Feature extraction is an essential step in converting raw speech signals into meaningful numerical representations. In this work, both acoustic and linguistic features are obtained to capture multiple aspects of speech.
Acoustic features are directly derived from the audio sig-nal and represent the physical characteristics of speech. One of the key acoustic features used in this study is the Mel Fre-quency Cepstral Coefficient (MFCC), which is widely ap-plied in speech processing as it closely models human au-ditory perception. Other acoustic parameters include pitch, energy, speech rate, pause duration, and zero-crossing rate.
Linguistic features are obtained from speech transcripts using the TF-IDF (Term Frequency–Inverse Document Fre-quency) technique. TF-IDF evaluates the importance of words within a document in comparison to a larger corpus, helping to identify distinctive terms that may signal cogni-tive decline.
Combining acoustic and linguistic features results in a more comprehensive representation of speech patterns. While acoustic features capture temporal and prosodic prop-erties, linguistic features focus on vocabulary usage and sen-tence structure.
The extracted features are integrated into a hybrid feature vector, which serves as input to the machine learning model. This multimodal strategy improves the system's capability to identify subtle variations between normal speech and speech affected by Alzheimer's disease.

6 SYSTEM ARCHITECTURE

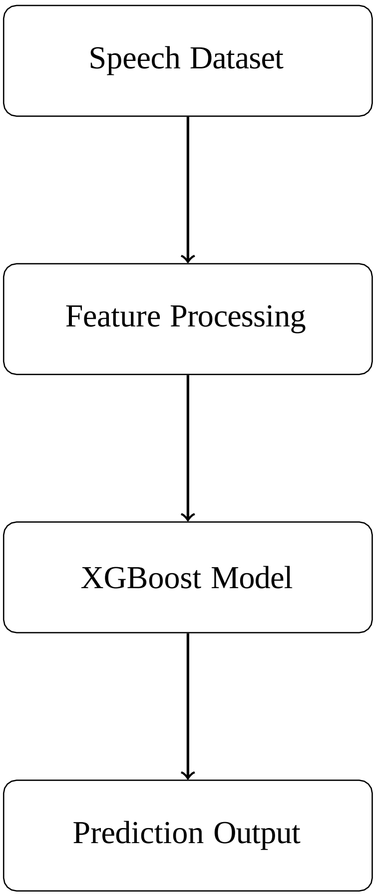
The proposed Alzheimer's disease detection system is struc-tured as a multi-stage pipeline that converts raw speech recordings into meaningful representations for classification. The overall architecture combines signal processing, feature extraction, feature integration, and machine learning-based classification within a unified framework.
The first stage involves the collection of speech data. Recordings are obtained from participants performing cog-nitive tasks such as picture description and narrative speech. These activities are commonly used in clinical evaluations as they capture spontaneous speech patterns, which provide strong indicators of cognitive condition.
In the second stage, preprocessing methods are applied to improve the quality of the speech signals. Noise reduction techniques are used to remove background interference. Si-lence trimming is carried out to eliminate non-essential seg-ments, ensuring that only relevant speech portions are ana-lyzed. In addition, normalization is performed to maintain consistent signal amplitude across different recordings.
After preprocessing, feature extraction is carried out through two parallel streams. The first stream focuses on ex-tracting acoustic features directly from the audio signal, in-cluding Mel Frequency Cepstral Coefficients (MFCC), pitch, energy, zero-crossing rate, and speech rate. These features represent the temporal and spectral characteristics of speech. The second stream deals with linguistic feature extraction. Speech recordings are converted into text using automatic speech recognition (ASR) systems. From these transcripts, features such as TF-IDF vectors, vocabulary diversity, and sentence complexity are derived, offering insights into lan-
guage usage and cognitive processes.
During the feature fusion stage, both acoustic and linguis-tic features are merged into a single hybrid feature vector. This multimodal representation allows the model to utilize complementary information from speech signals as well as language patterns.
The final stage performs classification using the XGBoost algorithm. The model is trained to differentiate between healthy individuals and those with Alzheimer's disease based on the extracted features. The architecture is designed to be efficient in terms of computation and scalable, making it practical for real-world applications.
Overall, the system architecture provides a comprehensive analysis of speech data, supporting accurate and early identi-

IJARCCE
ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.15537



fication of Alzheimer's disease.

Figure 1: Architecture of the proposed Alzheimer's detection sys-tem

6.1 Acoustic Features

Acoustic features describe the physical and spectral charac-teristics of speech signals. These features are extracted directly from audio recordings and offer important information related to voice quality, pitch variation, energy distribution, and articulation behavior. Since Alzheimer's disease impacts brain regions involved in speech production, variations in these acoustic properties can act as early signs of cognitive decline.

One of the most commonly used acoustic features is the Mel Frequency Cepstral Coefficients (MFCC). MFCCs are designed to resemble the human auditory perception system by mapping frequencies onto the mel scale, which better re-flects how humans perceive sound. The process of comput-ing MFCCs includes steps such as pre-emphasis, framing, windowing, Fast Fourier Transform (FFT), mel filter bank analysis, and discrete cosine transform. These coefficients effectively represent the spectral envelope of speech and are highly effective in distinguishing between normal and im-paired speech patterns.

Pitch is another significant acoustic feature that represents the fundamental frequency of speech. Changes in pitch can reflect emotional condition, stress levels, and neurological health. Individuals with Alzheimer's disease often show reduced pitch variation, leading to more monotonous speech. This reduced variability can be measured using pitch extraction techniques.

Energy-based features indicate the intensity of the speech signal. These features help in identifying variations in vo-cal strength and emphasis. Patients with Alzheimer's may exhibit irregular energy patterns due to challenges in control-ling speech output. Commonly used measures in this context include short-term energy and root mean square (RMS) energy.

Speech rate and articulation rate are also important acous-tic parameters. Speech rate refers to the number of words spoken within a specific time frame, whereas articulation rate considers only the spoken portions, excluding pauses. Peo-ple experiencing cognitive decline often demonstrate slower speech rates due to difficulty in retrieving words and forming sentences.

Another useful feature is the Zero Crossing Rate (ZCR), which measures how frequently the signal crosses the zero amplitude line. ZCR provides information about the fre-quency characteristics of the signal and helps differentiate

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.15537

between voiced and unvoiced segments. Variations in ZCR patterns may indicate abnormalities in speech production. In addition, spectral features such as spectral centroid, spectral bandwidth, and spectral roll-off are used to examine how energy is distributed across different frequencies. These features contribute to a deeper understanding of the acoustic structure of speech signals.
Overall, acoustic features are essential for capturing both physiological and neurological aspects of speech production. When integrated with other types of features, they signif-icantly improve the effectiveness of Alzheimer's detection systems.

- Mel Frequency Cepstral Coefficients (MFCC)
- Pitch
- Energy
- Speech Rate
- Pause Duration
- Zero Crossing Rate

MFCC features are especially important as they closely ap-proximate how the human auditory system interprets sound frequencies. They capture essential spectral properties of speech and reflect subtle variations in tone and articulation. In Alzheimer's Disease, these changes help in identifying im-paired speech patterns, making MFCC a reliable feature for accurate classification.

6.2 Speech Fluency Features

Speech fluency features capture the temporal and behavioral aspects of speech production and provide insights into cogni-tive processes involved in language generation. These fea-tures are particularly useful for identifying early signs of Alzheimer's disease.
Pause frequency reflects the number of silent pauses in speech, which tend to increase in Alzheimer's patients due to difficulties in word retrieval and sentence planning. Sim-ilarly, pause duration measures the length of these silent in-tervals, where longer pauses indicate hesitation and higher cognitive effort.
Speech rate is another important indicator, with reduced rates commonly observed in individuals with cognitive im-pairment. Filled pauses such as "um" and "uh" also increase as speakers struggle to organize thoughts.
Sentence length and structure provide additional insights, as Alzheimer's patients often produce shorter and simpler sentences. Repetition rate, indicating repeated words or phrases, is also associated with memory limitations.
Disfluencies such as false starts, corrections, and interrup-tions reflect challenges in planning and delivering coherent speech. Prosodic features, including rhythm and intonation, along with temporal measures like speech segment duration and silence ratio, further help in analyzing speech patterns.
Overall, speech fluency features offer valuable cues for de-tecting cognitive decline and, when combined with acoustic and linguistic features, enhance the accuracy of Alzheimer's detection systems.

Table 2: Speech Fluency Indicators

Feature	Detailed Description
Pause Frequency	Number of silent pauses during speech. Higher values indicate dif-ficulty in word retrieval and sen-tence formation.
Speech Rate	Words spoken per second. Lower speech rate reflects cognitive slow-ing and reduced fluency.
Average Pause Length	Duration of pauses between spoken segments. Longer pauses indicate hesitation and increased cognitive load.
Sentence Length	Average number of words per sen-tence. Alzheimer's patients often use shorter and simpler sentences.
Repetition Rate	Frequency of repeated words or phrases, indicating memory issues and reduced vocabulary control.
Filled Pauses	Use of filler words like "um" and "uh", showing hesitation in speech planning.
	Includes interruptions, corrections, and broken speech patterns reflect-ing instability.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.15537

Disfluency Rate	

6.3 Linguistic Features

Linguistic features focus on the content and structure of spo-ken language and are derived from speech transcripts. They provide insights into vocabulary usage, grammar, syntax, and semantic consistency. Alzheimer's disease affects language processing, leading to noticeable changes in these patterns.
TF-IDF is a commonly used technique for extracting lin-guistic features, as it measures the importance of words within a document relative to a dataset. It helps identify dis-tinctive terms and differentiate between speech samples.
Vocabulary richness reflects the diversity of words used by a speaker. Alzheimer's patients often use a limited vocabulary and repeat common terms, which can be measured using metrics like Type-Token Ratio (TTR).
Syntactic complexity describes sentence structure, with af-fected individuals typically producing shorter and simpler sentences. Features such as sentence length and clause com-plexity help evaluate this aspect.
Semantic coherence represents the logical flow of ideas. Patients may produce disorganized speech due to difficulty maintaining context, which can be analyzed using techniques like word embeddings and similarity measures.
Part-of-speech (POS) tagging further helps identify pat-terns in language usage, as Alzheimer's patients often rely more on pronouns and fewer specific nouns. Repetition and redundancy are also key indicators linked to memory limita-tions.
Advanced approaches, including transformer-based mod-els, can enhance linguistic analysis by capturing contextual relationships between words, improving detection per-formance.

7 MACHINE LEARNING MODEL

The classification process in this study is carried out using the Extreme Gradient Boosting (XGBoost) algorithm, a powerful ensemble learning method based on gradient boosting concepts. XGBoost has become widely popular in machine learning due to its strong performance, scalability, and effec-tiveness in handling structured datasets.
XGBoost works by building an ensemble of decision trees in a sequential way. Each new tree is trained to reduce the errors made by the previous ones, leading to continu-ous improvement in model accuracy. The algorithm relies on gradient-based optimization to minimize a defined loss func-tion, making it highly suitable for classification problems.
A major advantage of XGBoost is its capability to manage heterogeneous feature sets, which makes it appropriate for the hybrid representation used in this work. Since the com-bination of acoustic and linguistic features creates a complex feature space, XGBoost is able to effectively learn interac-tions among these features.
To avoid overfitting, regularization techniques are applied within the model. Parameters such as maximum tree depth and learning rate help control model complexity, while sub-sampling methods enhance generalization ability. In addition, XGBoost supports parallel computation, which helps in reducing training time significantly.
The model is trained on labeled data, where each speech sample is categorized as either healthy or affected by Alzheimer's disease. During the training phase, the algo-rithm iteratively updates its parameters to minimize classification errors. Hyperparameter tuning is also performed to achieve optimal performance.
Overall, XGBoost offers a good balance between accu-racy and computational efficiency, making it a suitable choice for practical healthcare applications where computational re-sources may be constrained.

8 EXPERIMENTAL SETUP

The experimental setup of the proposed Alzheimer's detec-tion system was designed to ensure reliable and reproducible results, with the primary goal of evaluating the effectiveness of combining acoustic and linguistic features.
The dataset consists of speech recordings from healthy in-dividuals and Alzheimer's patients, each labeled accordingly. It was split into training and testing sets using an 80:20 ra-tio to allow effective model training and unbiased evaluation. Stratified sampling was applied to maintain a balanced distri-bution of classes in both sets, which is important for medical datasets.
Preprocessing was performed on raw audio signals, includ-ing noise reduction using spectral gating, silence trimming

Table 3: XGBoost Hyperparameters

Parameter	Value and Description

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.15537

Learning Rate	0.1 (Controls the step size dur-ing model training and helps prevent overfitting by updating weights gradually)
Max Depth	6 (Defines the maximum depth of each decision tree, con-trolling model complexity and learning capacity)
Number of Trees	200 (Total number of boost-ing rounds used to build the ensemble model for better performance)
Subsample Ratio	0.8 (Fraction of training sam-ples used for each tree to re-duce overfitting and improve generalization)
Regularization	1.0 (Controls model complexity by penalizing large weights and preventing overfitting)

to remove non-informative segments, and normalization to standardize amplitude levels.

Feature extraction was carried out in two stages. Acous-tic features such as MFCC, pitch, energy, zero-crossing rate, speech rate, and pause duration were extracted using the Li-brosa library, capturing the physical and temporal properties of speech. Linguistic features were obtained from transcripts using TF-IDF vectorization, converting text into numerical representations based on word importance.
These features were combined into a hybrid feature vec-tor, enabling the model to capture both speech characteristics and language patterns. Classification was performed using the XGBoost algorithm, selected for its efficiency and ability to handle structured data. Hyperparameters such as learning rate, tree depth, and number of estimators were tuned to op-timize performance.
The implementation was carried out in Python using li-braries such as NumPy, Pandas, Scikit-learn, and Librosa. The system was executed on a standard computing setup without GPU support, demonstrating its computational effi-ciency.

9 EVALUATION METRICS

Evaluation metrics are essential for assessing the perfor-mance and reliability of a machine learning model, espe-cially in healthcare applications where incorrect predictions can have serious consequences. In this study, multiple met-rics are used for comprehensive evaluation.
Accuracy measures the proportion of correctly predicted instances among all observations. While it provides an over-all performance estimate, it may be misleading for imbal-anced datasets.
Precision is the ratio of true positives to all predicted posi-tive cases, indicating how many predicted Alzheimer's cases are correct. High precision helps reduce false positives and unnecessary medical actions.
Recall (sensitivity) is the ratio of true positives to ac-tual positive cases, reflecting the model's ability to identify Alzheimer's patients. High recall is crucial to avoid missed diagnoses.
The F1-score, the harmonic mean of precision and recall, provides a balanced measure by considering both false positives and false negatives.
Additionally, a confusion matrix is used to analyze pre-diction outcomes in detail, offering insights into correct and incorrect classifications.

10 RESULTS

The experimental findings highlight the effectiveness of the proposed hybrid feature-based model for Alzheimer's disease detection. The model achieved an overall accuracy of 76.44
The precision value of 0.74 indicates that most of the cases predicted as Alzheimer's are correctly classified. This sug-

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.15537

gests that the model maintains a relatively low false positive rate, which is important in medical applications.
The recall score of 0.71 shows that the model is able to correctly identify a considerable proportion of actual Alzheimer's cases. Although there is scope for further im-provement, the result demonstrates the model's capability in detecting cognitive impairment.
The F1-score of 0.72 represents a balanced trade-off be-tween precision and recall, indicating that the model per-forms consistently without favoring one metric excessively over the other.
A comparative evaluation with traditional models such as Logistic Regression, Random Forest, and Support Vector Machine reveals that the proposed hybrid approach performs significantly better. This improvement can be attributed to the effective combination of both acoustic and linguistic fea-tures.

11 CONFUSION MATRIX ANALYSIS

The confusion matrix offers a detailed representation of the classification outcomes by comparing actual labels with pre-dicted labels. It is composed of four key elements: true posi-tives, true negatives, false positives, and false negatives.
In this study, the confusion matrix shows that the model correctly identified 42 healthy individuals and 18 patients with Alzheimer's disease. However, 6 healthy individu-als were incorrectly predicted as Alzheimer cases, while 8 Alzheimer patients were misclassified as healthy.
False positives correspond to instances where healthy in-dividuals are classified as having Alzheimer's disease. Al-though this may cause unnecessary concern, it is generally

Table 4: Model Performance

Metric	Score and Description
Accuracy	76.44% (Represents the overall correctness of the model by mea-suring the proportion of correctly classified samples)
Precision	0.74 (Indicates how many of the predicted Alzheimer cases are actually correct, reflecting false positive control)
Recall	0.71 (Measures the model's ability to correctly identify Alzheimer's cases from all actual positive instances)
F1 Score	0.72 (Harmonic mean of preci-sion and recall, providing a bal-anced evaluation of model perfor-mance)

considered less critical than false negatives.
False negatives refer to cases where patients with Alzheimer's disease are predicted as healthy. This is a more serious concern, as it can lead to delayed diagnosis and treat-ment.
The comparatively low number of incorrect predictions suggests that the model performs effectively overall. Nev-ertheless, further improvements can be achieved by refining feature selection methods and tuning model parameters.

12 FEATURE IMPORTANCE ANALYSIS

Understanding which features have the greatest influence on model predictions is important for interpreting machine learning systems, especially in healthcare applications.
The XGBoost model provides feature importance scores that reflect how relevant each feature is during the construc-tion of decision trees. In this study, MFCC features emerged as some of the most significant acoustic features, as they ef-fectively capture the spectral properties of speech signals.

ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.15537

Speech fluency features, including pause frequency and speech rate, were also found to have a strong impact on model performance. Individuals with Alzheimer's disease often exhibit longer pauses and slower articulation, patterns that the model is able to detect effectively.

Among the linguistic features, vocabulary diversity and sentence length demonstrated high predictive importance. Patients with Alzheimer's tend to produce shorter sentences and show increased repetition compared to healthy individu-als.

Feature importance analysis confirms that the model is capturing meaningful patterns in the data rather than relying on random relationships.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.15537

13 COMPARATIVE ANALYSIS

Table 5: Comparison with Other Models

Model	Accuracy and Description
Logistic Regression	48% (Baseline linear model with limited capability to cap-ture complex speech patterns)
Random Forest	58% (Ensemble model that im-proves performance but lacks efficiency in handling sequen-tial speech features)
Support Vector Ma-chine	63% (Effective for high-dimensional data but strug-gles with large-scale feature interactions)
Proposed Hybrid Model	76.44% (Combines acoustic and linguistic features, achieving the highest accuracy and better generalization)

14 FEATURE IMPORTANCE DISCUSSION

Understanding which features have the greatest impact on model predictions is essential in healthcare-related machine learning applications. In this study, feature importance anal-ysis was carried out using the XGBoost model to determine the most influential speech and linguistic attributes.
Among the acoustic features, Mel Frequency Cepstral Co-efficients (MFCC) contributed the most to classification performance. These features effectively capture the spectral properties of speech and provide a reliable representation of human vocal patterns.
Speech rate and pause duration were also identified as key indicators of cognitive decline. Individuals with Alzheimer's disease often demonstrate slower speech and longer pauses during communication due to difficulties in language processing.
Linguistic features such as vocabulary diversity and sen-tence complexity also had a significant impact. Patients with Alzheimer's typically produce simpler sentence structures and show a higher tendency to repeat words compared to healthy individuals.
The integration of these acoustic and linguistic features en-ables the machine learning model to capture subtle variations in speech patterns, resulting in improved classification accu-racy.

15 PERFORMANCE VISUALIZATION

Performance visualization plays a key role in model evalua-tion as it offers a clear and intuitive representation of results. In this study, a bar graph is used to compare the accuracy of different machine learning models.

The visualization clearly indicates that the proposed hy-brid model achieves the highest accuracy when compared to Logistic Regression, Random Forest, and Support Vector Machine. This demonstrates the effectiveness of integrating both acoustic and linguistic features.
Visualization techniques also assist in identifying trends and patterns in model performance, making it easier to interpret results and present findings to a wider audience.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.15537

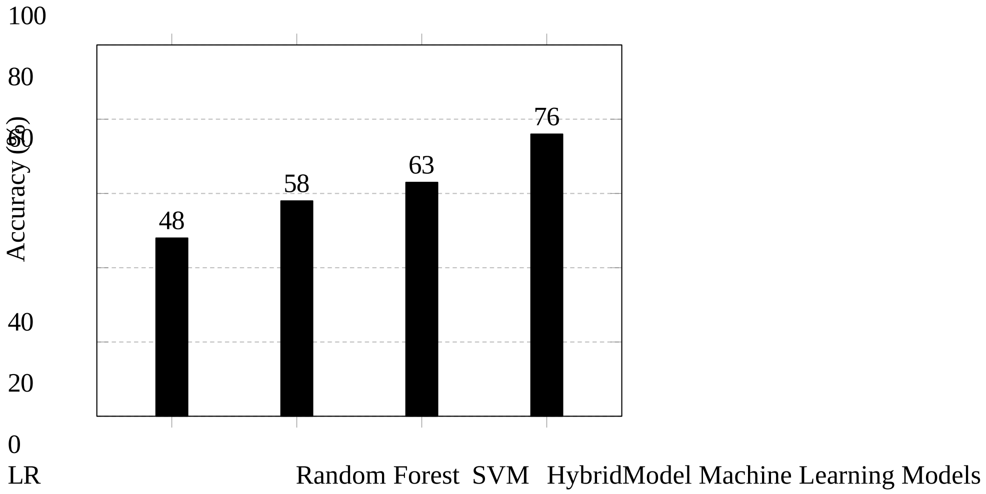


Figure 2: Accuracy comparison of different models

The graphical representation in Fig. 2 clearly illustrates the improvement achieved by the proposed hybrid model. Traditional models such as Logistic Regression and Random For-est show lower accuracy because they rely on limited feature representations.
The hybrid model combines both acoustic and linguis-tic features, allowing the classifier to capture more complex patterns related to cognitive decline. As a result, the hy-brid model achieves the highest accuracy of 76%, demonstrating the effectiveness of multimodal feature analysis for Alzheimer's detection.

16 DISCUSSION

The findings of this study demonstrate the strong potential of speech-based analysis as a non-invasive approach for early detection of Alzheimer's disease. The experimental results show that the proposed hybrid model, combining acoustic and linguistic features, effectively captures subtle patterns associated with cognitive decline. These observations are consistent with prior research, where speech and language impairments are considered early indicators of Alzheimer's progression.
A major strength of this approach lies in integrating mul-tiple feature types. Acoustic features such as MFCC, pitch, and speech rate reflect the physical and temporal properties of speech, while linguistic features derived from transcripts capture higher-level aspects like vocabulary usage, sentence structure, and semantic coherence. The combination of these features enables a more comprehensive representation of an individual's cognitive state.

The improved performance of the hybrid model compared to single-modality models further highlights the importance of multimodal integration. Acoustic-only models may miss linguistic abnormalities, whereas text-based models may ig-nore prosodic cues. By combining both, the proposed system enhances classification accuracy and robustness.
The use of the XGBoost algorithm also contributes to the effectiveness of the system. Its ability to handle structured data and capture complex feature interactions makes it suitable for hybrid feature spaces. The results indicate that well-optimized traditional machine learning methods can still achieve strong performance when supported by effective fea-ture engineering.
From a practical perspective, the proposed system offers advantages over conventional diagnostic methods. Unlike MRI and PET scans, which are expensive and require special-ized infrastructure, speech-based analysis can be performed using simple audio recordings. This makes it more accessi-ble and suitable for large-scale screening, while also being non-invasive and user-friendly.
The results further confirm that features such as pause du-ration, speech rate, and repetition patterns are strong indicators of cognitive decline. These findings align with clinical observations, where patients often exhibit slower speech, in-creased hesitation, and reduced linguistic complexity.
However, the results should be interpreted considering dataset limitations. The relatively small size and limited diversity may affect generalization to real-world scenarios. De-spite this, the study provides a strong foundation for future research in speech-based Alzheimer's detection.
In summary, integrating acoustic and linguistic features with an efficient machine learning model significantly im-

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.15537

proves Alzheimer’s detection. The proposed approach en-hances performance while offering a practical and scalable solution for early diagnosis.

17 LIMITATIONS AND FUTURE WORK

Despite the promising results, several limitations should be considered to provide a balanced evaluation of the proposed approach. Identifying these limitations is essential for guid-ing future improvements.
A major limitation is the relatively small dataset size, which may not capture the full diversity of real-world speech patterns. This can lead to overfitting, where the model per-forms well on training data but fails to generalize to unseen samples. In healthcare applications, strong generalization is critical for reliable performance across different patient groups.
Another limitation is the lack of demographic diversity. Speech patterns vary with age, gender, language, and cultural background, and a less diverse dataset may reduce model consistency across populations. Future work should focus on larger and more diverse datasets to improve robustness.
Variations in recording conditions, such as background noise and device quality, also affect performance. Although preprocessing techniques like noise reduction and normaliza-tion were applied, they may not fully remove such inconsis-tencies. More advanced signal processing methods could fur-ther enhance reliability.
While the XGBoost model provides strong results, it may not fully capture temporal dependencies in speech data. Deep learning approaches such as LSTM and CNN models could better model sequential and hierarchical patterns, potentially improving performance.
Additionally, incorporating other modalities such as facial expressions or physiological signals could provide complementary information and improve detection accuracy. Real-time implementation using mobile or wearable devices is another important direction for practical deployment.
Ethical considerations, including data privacy and secu-rity, must also be addressed to ensure responsible use in real-world applications.
In conclusion, although the proposed system shows en-couraging results, improvements in dataset size, diversity, feature representation, and model design are necessary to de-velop a more robust and scalable Alzheimer’s detection system.

18 REAL-WORLD APPLICATIONS

The proposed speech-based Alzheimer’s detection system shows strong potential for real-world use, particularly in healthcare. Its key advantage lies in providing a non-invasive, cost-effective, and accessible method for early detection of cognitive disorders.
In telemedicine, the system can be integrated into remote platforms for early screening. Patients can submit speech samples via smartphones or computers, which are then an-alyzed for signs of cognitive decline. This is especially beneficial for individuals with limited access to healthcare or mo-bility constraints.
Mobile health applications offer another important use case. By embedding the model into apps, users can regularly monitor cognitive health through simple speech-based tests. These systems can provide instant feedback and alert users or healthcare providers when abnormalities are detected, en-abling early intervention.
In rural and underserved areas, where advanced diagnostic tools like MRI and PET scans are scarce, speech-based analysis serves as a practical alternative. Its low cost and minimal infrastructure requirements make it suitable for large-scale deployment.
Elderly care facilities can also benefit from continuous monitoring of residents’ speech patterns, helping caregivers detect early cognitive decline and provide timely, personal-ized care.
The system can further support clinical decision-making by acting as a preliminary screening tool to identify high-risk individuals, thereby reducing the burden on health-care systems. Integration with wearable devices and smart home technologies can enable continuous, passive monitor-ing through everyday interactions.
Overall, the proposed approach offers diverse real-world applications and has the potential to improve early detection and management of Alzheimer’s disease through accessible and scalable solutions.

19 FUTURE SCOPE

The field of Alzheimer’s disease detection using artificial in-telligence is rapidly evolving and offers significant opportu-nities for future research. While the proposed system demon-strates promising results using machine learning and hybrid feature extraction, further improvements are needed in accu-racy, scalability, real-time deployment, and security.
A key direction for future work is the adoption of deep learning techniques for speech and text analysis. Unlike tra-

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.15537

ditional models such as XGBoost, deep learning approaches including CNNs, RNNs, and LSTM networks can automati-cally learn complex patterns and temporal dependencies from raw data, improving the detection of subtle cognitive im-pairments. Transformer-based models like BERT can further enhance linguistic analysis by capturing contextual meaning and relationships between words more effectively than TF-IDF methods.

Another promising area is multimodal learning, where ad-ditional data sources such as facial expressions, eye move-ments, and neuroimaging can be combined with speech and text. This integration can provide a more comprehensive un-derstanding of cognitive health and improve diagnostic accu-racy.

Federated learning is an important approach for ensuring data privacy in healthcare. It enables multiple institutions to collaboratively train models without sharing raw data, thereby maintaining data security while improving model generalization across diverse populations.

Security and privacy remain critical challenges. Advanced techniques such as encryption, secure communication, and access control must be implemented to protect sensitive pa-tient data. Additionally, models should be designed to resist adversarial attacks that may compromise prediction reliabil-ity.

Explainable AI is also essential in healthcare applications. Future systems should provide interpretable results to help clinicians understand model decisions and build trust. Tech-niques such as SHAP and LIME can improve transparency.

Scalability and real-time implementation are important for practical use. Optimized models deployed on cloud, mobile, or wearable platforms can enable continuous monitoring and early detection. Personalized models can further improve performance by adapting to individual speech patterns using techniques like transfer learning.

Future research can also explore longitudinal analysis, where speech data collected over time is used to track cogni-tive changes and detect early decline more effectively. Inte-gration with telemedicine platforms can enhance accessibility, especially for patients in remote areas.

Ethical considerations, including fairness, accountability, and data protection, must be addressed to ensure responsible deployment. Additionally, creating larger and more diverse datasets will be essential for building robust and generaliz-able models.

In conclusion, combining deep learning, federated learn-ing, and secure, explainable systems can significantly ad-vance Alzheimer's detection. These developments have the potential to provide accurate, scalable, and accessible solu-tions for early diagnosis and improved patient care.

20 CONCLUSION

In this study, a machine learning-based framework for de-tecting Alzheimer's disease using speech and textual features was developed and evaluated. The main objective was to de-sign an efficient, non-invasive, and scalable system capable of identifying early signs of cognitive decline.

The results indicate that combining acoustic and linguis-tic features leads to a significant improvement in classifica-tion performance. The hybrid model achieved an accuracy of 76.44

The XGBoost algorithm proved to be well-suited for han-dling the hybrid feature space and modeling complex rela-tionships within the data. Feature importance analysis fur-ther revealed that the model relies on meaningful indicators, including MFCC features, speech rate, pause duration, and vocabulary diversity.

A key contribution of this work is the presentation of a practical and accessible solution for Alzheimer's detection. Unlike traditional diagnostic approaches, the proposed sys-tem does not depend on expensive equipment or invasive pro-cedures, making it suitable for large-scale screening and de-ployment in resource-limited environments.

However, certain limitations must be considered, including the relatively small dataset and limited diversity. Addressing these issues will be essential for improving the generalization capability and robustness of the model.

Future research should focus on expanding the dataset, exploring advanced deep learning approaches, and incorpo-rating additional modalities to further enhance performance. Real-time implementation and validation in real-world sce-narios will also be important for practical adoption.

In conclusion, the proposed speech-based Alzheimer's de-tection system represents a promising step toward early di-agnosis and improved patient care. With continued research and development, this approach has the potential to become a valuable tool in addressing the challenges associated with Alzheimer's disease.

REFERENCES

[1] Fraser, K., Meltzer, J., & Rudzicz, F. (2016). Linguis-tic features identify Alzheimer's disease in narrative speech.World Health Organization. Dementia Fact Sheet.
[2] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system.
[3] Satt, A., Hoory, R., Ko¨nig, A., Aalten, P., & Robert, P. (2017). Speech-based automatic detection of early de-mentia.
[4] Toth, L., et al. (2018). Automatic detection of mild cog-nitive impairment using speech recognition.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⩼ Peer-reviewed & Refereed journal ⩼ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.15537

[5] Konig, A., et al. (2015). Automatic speech analysis for Alzheimer's disease assessment.
[6] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
[7] Jurafsky, D., & Martin, J. H. (2021). Speech and Lan-guage Processing.
[8] Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.
[9] Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python.
[10] McFee, B., et al. (2015). Librosa: Audio signal process-ing in Python.
[11] Eyben, F., et al. (2010). Opensmile: Audio feature ex-traction toolkit.
[12] Devlin, J., et al. (2019). BERT: Deep bidirectional transformers.
[13] Hinton, G., et al. (2012). Deep neural networks for speech recognition.
[14] Graves, A., et al. (2013). Speech recognition with RNNs.
[15] Alzheimer's Association. (2023). Alzheimer's Disease Facts and Figures.
[16] WHO. (2021). Global status report on dementia.
[17] Young, T., et al. (2018). Deep learning for NLP.
[18] Zhang, Z., & Schuller, B. (2019). Speech emotion recognition.
[19] Ravi, D., et al. (2017). Deep learning for activity recog-nition.
[20] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature.
[21] He, K., et al. (2016). Deep residual learning for image recognition.
[22] Vaswani, A., et al. (2017). Attention is all you need.
[23] Szegedy, C., et al. (2015). Going deeper with convolu-tions.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.15537

[24] Kingma, D., & Ba, J. (2014). Adam: A method for stochastic optimization.
[25] Breiman, L. (2001). Random forests.
[26] Cortes, C., & Vapnik, V. (1995). Support-vector net-works.
[27] Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification.
[28] Duda, R., Hart, P., & Stork, D. (2001). Pattern Classifi-cation.
[29] Rabiner, L. (1989). A tutorial on hidden Markov mod-els.
[30] Schuller, B., et al. (2011). Recognizing realistic emo-tions.
[31] Cummins, N., et al. (2015). Review of depression de-tection using speech.
[32] Low, L. A., et al. (2020). Automated speech analysis in dementia.
[33] Pakhomov, S., et al. (2010). Computerized analysis of speech for dementia.
[34] Fraser, K., et al. (2019). Detecting Alzheimer’s using speech and NLP.
[35] Mirheidari, B., et al. (2018). Dialogue-based Alzheimer detection.
[36] Balagopalan, A., et al. (2020). Speech-based detection of Alzheimer’s.
[37] Haider, F., et al. (2020). Acoustic features for cognitive decline detection.
[38] Weiner, M., et al. (2017). Alzheimer’s Disease Neu-roimaging Initiative.
[39] Jack, C., et al. (2018). NIA-AA research framework for Alzheimer’s.