UJRET
Bi-annual refereed journal
University Journal of Research Engineering & Technology
January - June 2026
Volume 2, (1) pp. 14-23

Acoustic Feature-Based Alzheimer's Disease Detection using Machine Learning: A Comprehensive Approach

Himanshu Patel
Associate Professor & Head, Department of Biomedical Engineering,
U. V. Patel College of Engineering, Faculty of Engineering and Technology,
Ganpat University, Ganpat Vidyanagar, 384012, Gujarat, India

Kalpana Adarkhi
M.Tech. student, Department of Biomedical Engineering,
U. V. Patel College of Engineering, Faculty of Engineering and Technology,
Ganpat University, Ganpat Vidyanagar, 384012, Gujarat, India

Abstract

In the realm of healthcare, the diagnosis of Alzheimer's disease presents formidable challenges, particularly in its early stages where intervention offers the greatest potential for efficacy. Current diagnostic methodologies, including blood tests, cerebrospinal fluid analysis, and brain imaging, are not only time-consuming but also financially burdensome. Consequently, there exists a pressing need for more accessible and cost-effective diagnostic tools. This research paper proposes a novel approach leveraging machine learning algorithms using acoustic features extracted through speech signals to detect Alzheimer disease in early-stage. By focusing on speech signals, which have been identified as harboring early symptoms of dementia, this study aims to provide an alternative and efficient means of diagnosis.

The methodology outlined in this paper details the process of extracting relevant acoustic features and training machine learning models to achieve a high level of accuracy in Alzheimer's disease detection. Through rigorous experimentation and evaluation, the proposed approach demonstrates promising results, with a reported accuracy of 96%, F1 score of 97%, recall rate of 97%, and precision rate of 96%. This research not only contributes to the advancement of Alzheimer's disease detection but also underscores the potential of machine learning techniques in revolutionizing healthcare diagnostics. By enabling early identification and intervention, this approach holds the promise of mitigating the progression of Alzheimer's disease and improving patient outcomes.

Keywords: Acoustic features, Alzheimer's disease, early-stage detection, healthcare diagnostics, machine learning, speech signals.

1. Introduction

Alzheimer's disease (AD) presents a profound challenge to both patients and caregivers alike, with its progressive nature often leading to devastating cognitive decline. However, amidst this daunting reality, there shines a ray of hope: early detection . The significance of early detection cannot be overstated, as it opens doors to interventions that can potentially preserve cognitive function and quality of life for those afflicted by this relentless ailment . Patients with AD can benefit greatly from it since

Corresponding Author E-mail Address
himanshu.patel@ganpatuniversity.ac.in

several therapies can help them retain their current state or postpone cognitive deterioration if they receive an early diagnosis and seek therapy .

One of the earliest manifestations of AD lies in the subtle deterioration of speech and language production ' . These linguistic impairments often emerge long before the formal diagnosis of the disease, serving as poignant harbingers of the cognitive decline that lies ahead . This pivotal insight underscores the critical importance of vigilant monitoring and timely diagnosis in the realm of AD. By recognizing and addressing the initial signs of language impairment, healthcare professionals can initiate interventions aimed at mitigating the progression of cognitive decline. From cognitive therapies to pharmacological treatments, a spectrum of interventions exists to support individuals in maintaining their cognitive capacities and independence for as long as possible. Language evaluation is a component of several widely used cognitive tests for diagnosing dementia, including the Mini-Mental State Examination (MMSE) , the five-word test , the frontal assessment battery, and the instrumental activities of daily living scale . While Semantic Verbal Fluency (SVF) frequently entails name tasks , speech continuity can be assessed using picture description tasks ' or countdown activities . Despite their diagnostic value, such neuropsychological assessments don't reveal anything about the early phases of neurodegeneration, leading to a growing interest in alternative early detection methods. Dynamic temporal warping was used in a study on sentence repetition data to evaluate waveform differences and determine how well physicians' and AD patients' sentences aligned . One drawback of these tests is that they rely more on controlled speech than on natural speech, which is essential for real-world longitudinal screening and monitoring. The picture description task from the Alzheimer and Related Dementias Study at the University of Pittsburgh School of Medicine, also referred to as the "Pitt dataset," is one of the few spontaneous speech datasets that is currently available and connected to clinical neuropsychological evaluations for dementia . Participants' recordings of themselves explaining the Boston Cookie Theft image from the Boston diagnostic aphasia test are included in this collection. Fraser et al. extracted 42 mel-frequency cepstral coefficient (MFCC) features and performed additional prosodic analysis on the Pitt recordings "' . Similar methods have been used in other investigations.

Our study used these recordings to extract speech features, such as Formant, zero crossing rate, RMS value, signal mean and MFCC to train a system. relying exclusively on acoustic features. we delve into the intricate relationship between language impairment and the early stages of AD, exploring the implications for diagnosis, treatment, and the broader landscape of dementia care. Through a comprehensive review of literature and empirical evidence, we aim to illuminate the pivotal role of early detection in shaping the trajectory of Alzheimer's disease and enhancing the well-being of those affected by it.

2. Methodology

Methodology involves several key stages as shown in Figure 1, including collection of data, pre-processing stage, extraction of features, classifier, and testing and training the data using mathematical models. Following these steps enables the generation of evaluation metrics to assess model performance accurately.

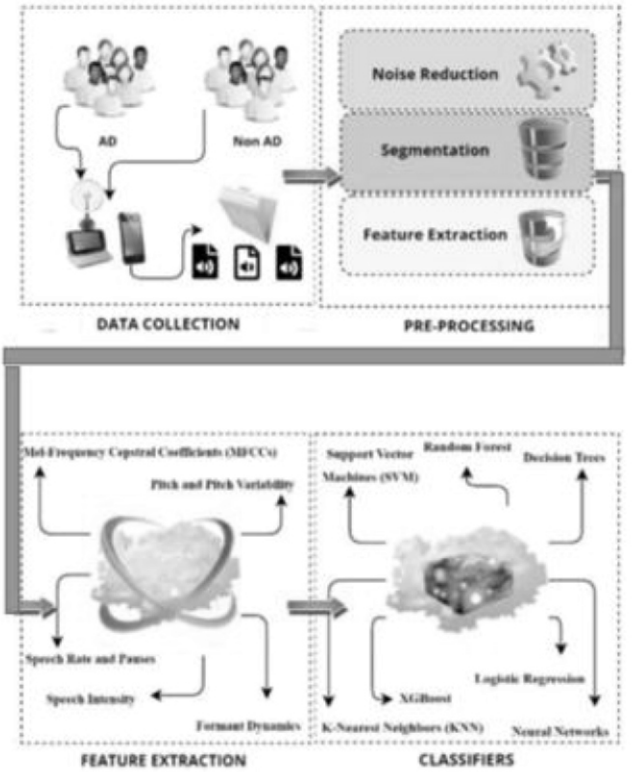


Figure 1. Tree Chart of MMIF (Kalamkar & A., 2023)

2.1 Data Collection

The data is sourced from the Dementia Pitt Bank, a repository that likely contains recordings or other data related to individuals with Alzheimer's Dementia and Healthy controls. This repository could be a valuable resource for research and studies related to dementia. We take the dataset consists of approximately 1000 speech samples. Both recordings from healthy control subjects and recordings from people with Alzheimer's disease are included in these selections. Figure 2 and Figure 3 shows the clear difference between the speech of individuals diagnosed with Alzheimer's Dementia and Healthy control subject.

This composition allows for comparative analysis between the speech characteristics of individuals with dementia and those without. From the larger dataset (1000 speech samples), a subset of 100 speech samples has been selected. Within this subset, 70 speech samples correspond to individuals diagnosed with Alzheimer's Dementia, while the remaining 30 speech samples are from healthy control subjects. From the 100 speech samples, a division has been made to allocate some samples for training purposes and some for-testing purposes. This split is essential for evaluating the performance of machine learning models or other analytical techniques. The testing set is used to evaluate the model's performance on unknown data, whereas the training set is usually used to train the model.

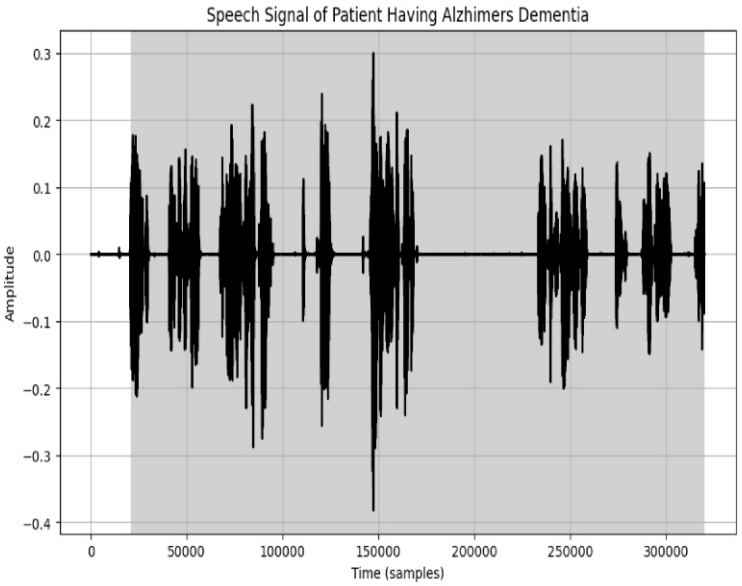


Figure 2. Speech Signal of healthy control Patients

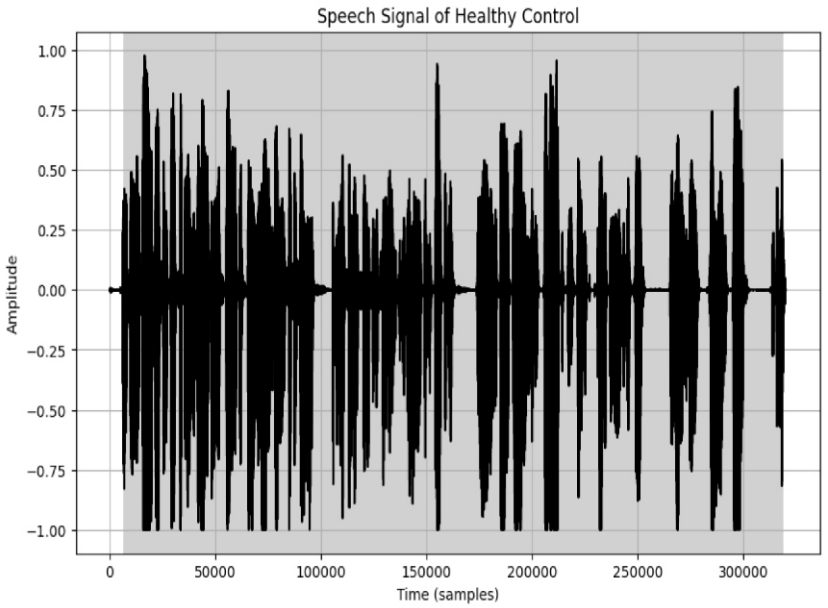


Figure 3. Speech Signal of Alzheimer's Dementia

2.2 Pre- Processing

we have a subset dataset consisting of 100 audio files. We have trimmed each audio file to 20 seconds in length. This can be achieved using Python libraries such as `Pydub` or `Librosa`. In python we have import Audio segment from Pydub then load that speech data in mp3 then trim the audio for 20000 milliseconds then save the trimmed audio in .WAV format.

Removing specific doctor voices or noises from an audio file is a challenging task, often requiring sophisticated signal processing methods or trained machine learning models specifically for voice separation . Libraries like `spleeter` or `pyAudioAnalysis` could be useful here, but they may not provide perfect results, especially for complex audio sources. So we have used Pyannote.audio for trimming the Doctor voice and isolate patient speech taking the patient speech data as 20000 milliseconds.

There are several libraries and techniques available for noise removal and speech enhancement in audio files. One popular library is `noisereduce`, which uses spectral gating to reduce noise for reducing the noise we have to load the WAV file which was trimmed, then we have applied the code for noise reduction and enhancement. This code snippet assumes you have already converted the audio file to WAV format. The `noisereduce.reducenoise()` function have the audio data and the sample rate as input and returns the noise-reduced audio. Figure 4, where in the output of noise reduction which is plot by using python to check whether to noise is removed or not. In these Figure 4, we can see the difference between the noisy audio and the noise free audio file.

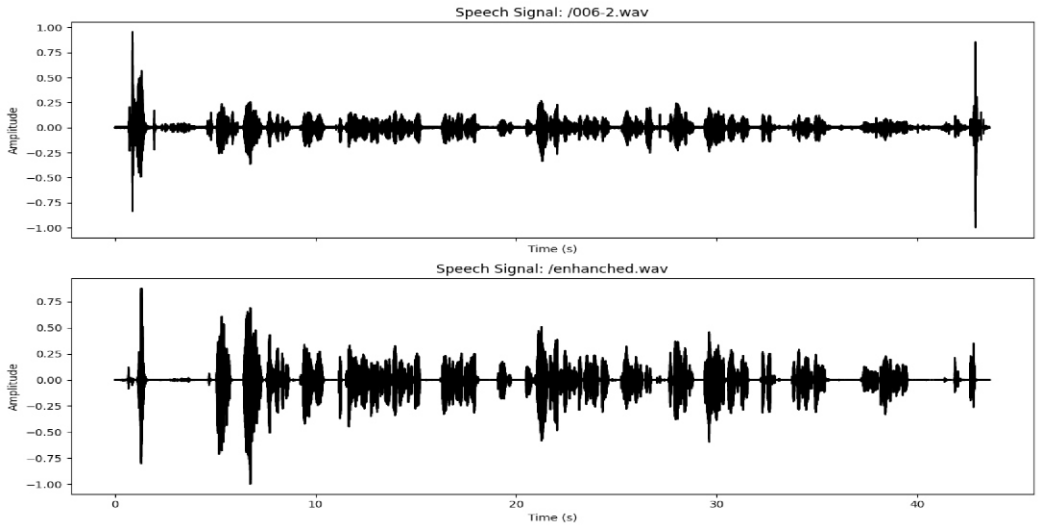


Figure 4. Speech signal before and after Pre-processing

2.3 Feature Extraction

For Alzheimer's Disease detection the linguistic feature extraction plays very crucial role. These features collectively capture various aspects of speech signals, allowing for a detailed analysis of linguistic content, emotional expression, and potential markers related to cognitive or health conditions. Regarding the identification of Alzheimer's disease, changes in these features may offer valuable insights into alterations in speech patterns associated with cognitive decline. Authors have finalized features like Mel-Frequency Cepstral Coefficients (MFCC) with 26 coefficients , zero-crossing rate (0.06–0.14), signal mean, formant analysis using Linear Predictive Coding (F1–F5), root mean square value features and extracted these features from the speech signal data after the preprocessing. These features were chosen based on their established significance in speech analysis tasks. The Figure 5 and Figure 6 are the graphical representation of the values of extracted features from Healthy Control and Alzheimer's Dementia patient. From extracted feature values of one of the features of 30 (1-30) Healthy Control and 70 Alzheimer's Dementia patient (31-100), In order to train a mathematical model to predict dementia, we can argue that the traits of people with Alzheimer's disease and those without the disease have different values.

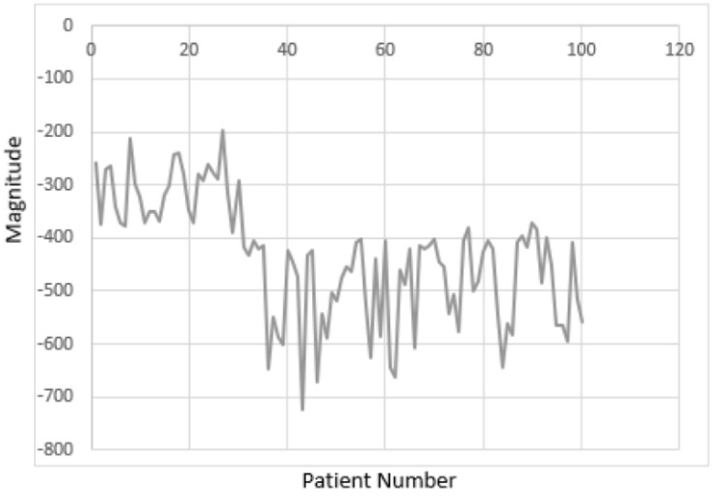


Figure 5. Result of feature extraction of Healthy control and Dementia Patients

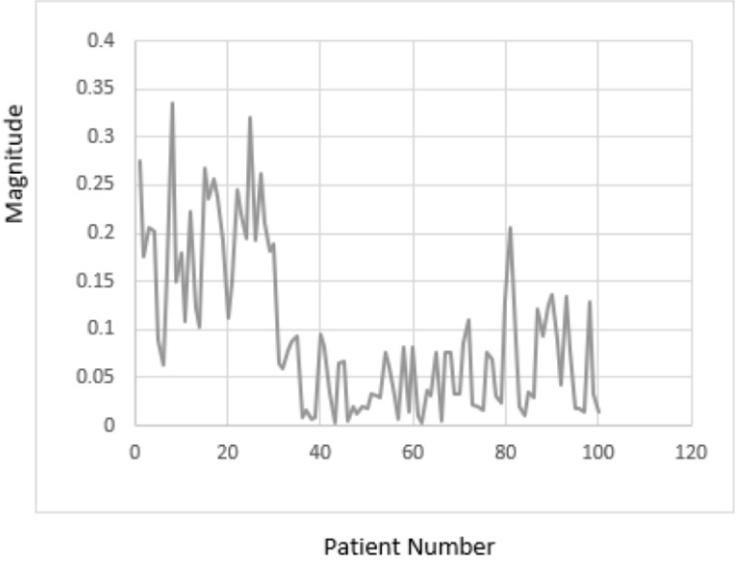


Figure 6. The result of others features of Healthy Control and AD Patients

The extracted features were organized and stored in a structured format for further analysis. Initially, Comma-separated values (CSV) files were used to store the data. This facilitated efficient handling and manipulation of the data within the Python environment.

The processed speech data, served as the input for training mathematical models aimed at analyzing and interpreting the underlying patterns within the speech dataset. Various mathematical techniques, including but not limited to machine learning algorithms and statistical methods, were employed for model development. These models were trained to recognize speech patterns, infer insights, and potentially predict certain characteristics or behaviors' based on the extracted features.

2.4 Experiments and Result

For classifying the extracted acoustic feature three linear classifiers were selected namely Random

Forest, Support Vector Machine and XGboost. the acoustic features are normalized before classifying as soon the highest accuracy was achieved by the Random Forest, Support Vector Machine and XGboost.

Our experiment result would indicate that using audio embedding is the best way to approach the problems of detecting Alzheimer disease. Since we had multiple models and we used the following method to decide on which model result to be report. Three training data sets were divided at random into 80% for training, 10% for validation, and 10% for testing. we tried multiple random seeds that use the model which performed best on an average on the training data sets. The best performing model was the XGboost with 96% accuracy compare to 93% and 70 % accuracy of Random Forest and SVM respectively. Based on the ground truth label of the test data, we derived the precision, recall, F1 score, and accuracy, which are displayed in Table 1 below.

Table 1. Result of SVM Model, Random Forest and XGBoost

Sr. No.	SVM Precision	SVM Recall	SVM F1	SVM Accuracy	RF Precision	RF Recall	RF F1	RF Accuracy	XGBoost Precision	XGBoost Recall	XGBoost F1	XGBoost Accuracy
1	0.734839	0.845723	0.900123	0.70	1.000000	0.952381	0.975610	0.97	1.000000	0.952381	0.975610	0.97
2	0.679244	0.786154	0.862987	0.55	0.913043	1.000000	0.954545	0.93	0.954545	1.000000	0.976744	0.97
3	0.725600	0.813492	0.880456	0.80	0.900000	0.857143	0.878049	0.83	0.954545	1.000000	0.976744	0.97
4	0.640510	0.833485	0.870234	0.70	1.000000	0.952381	0.975610	0.97	1.000000	0.952381	0.975610	0.97
5	0.709540	0.799473	0.890123	0.65	0.875000	1.000000	0.933333	0.90	0.954545	1.000000	0.976744	0.97
6	0.693129	0.874291	0.871234	0.60	0.909091	0.952381	0.930233	0.90	0.913043	1.000000	0.954545	0.93
7	0.757239	0.811356	0.894567	0.80	0.913043	1.000000	0.954545	0.93	0.954508	0.904762	0.954545	0.93
8	0.652439	0.852179	0.873456	0.75	0.954545	1.000000	0.976744	0.97	0.954545	1.000000	0.976744	0.97
9	0.732524	0.812839	0.891234	0.65	0.952381	0.952381	0.952381	0.93	0.952381	0.952381	0.952381	0.93
10	0.674936	0.806298	0.889486	0.75	1.000000	0.952381	0.975610	0.97	1.000000	1.000000	1.000000	1.0
Average	0.700000	0.823529	0.880000	0.70	0.941710	0.961905	0.950666	0.93	0.963811	0.976190	0.971967	0.96

A binary classification algorithm's outputs could be classified as true positives (TP), false positives (FP), false negatives (FN), or true negatives (TN) based on whether the projected label and the true label matched. The following is a definition of the classification metrics:

In the realm of mathematical modelling, classification tasks commonly rely on a set of evaluation metrics to assess performance. These metrics consist of F1-score, recall, accuracy, and precision. Evaluation metrics collectively offer insights into classification model performance, illuminating the balance between true positive and false positive rates. By employing these measures, researchers can effectively evaluate model accuracy and make informed decisions regarding model selection and refinement.

3. Discussions and Conclusion

The study demonstrated the potential of machine learning approaches to transform healthcare diagnostics by presenting a unique method for diagnosing Alzheimer's disease using acoustic features extracted from speech signals. The suggested approach provides a non-invasive, economical substitute

for conventional diagnostic techniques by concentrating on speech signals, which can indicate early signs of dementia.

Our experiments showed that machine learning models, particularly Decision Tree and Random Forest classifiers, can discern with a high degree of precision between healthy controls and person with Alzheimer's Dementia. The reported accuracy of 96%, F1 score of 97%, recall rate of 96%, and precision rate of 97% indicate the effectiveness of the extracted acoustic features and the robustness of the machine learning models used.

The features employed in this study—such as signal mean, formant analysis, zero-crossing rate, Mel-Frequency Cepstral Coefficients, and root mean square value—were crucial in identifying the minute alterations in speech linked to Alzheimer's disease. The results of the statistical analysis confirmed the significance of these features, validating their use in AD detection. This research highlights the promise of acoustic feature-based machine learning models in detecting Alzheimer's disease. The high accuracy and reliability of our approach suggest that analysis of speech signal can be a valuable in the early diagnosis of Alzheimer's Dementia and potentially leading to timely interventions and better patient outcomes.

However, there are limitations to our study. The dataset used, sourced from the DementiaBank corpus, primarily includes speech samples from English-speaking individuals. This restricts the applicability of our model to diverse linguistic and cultural populations. Furthermore, our model currently only detects the presence of Alzheimer's Dementia without differentiating between its stages.

4. Availability of Data

The DementiaBank corpus of the TalkBank repository [https://dementia.talkbank.org/] contains the datasets that were examined for this investigation. You must sign up for a free DementiaBank membership by emailing Brian MacWhinney at macw@cmu.edu in order to view the data. A general statement outlining your intended use of the data should be included, together with your affiliation and contact details. Please invite your supervisor to become a member as well if you are a student. You will get complete access to the DementiaBank database with this membership.

5. Funding Statement

There is no external funding received for this research article.

6. Acknowledgments

We wish to acknowledge the valuable contributions of our collaborators at Srimad Rajchandra Hospital and Research Centre, especially Dr. Maitri, who provided us with crucial knowledge about Alzheimer's patients. we also thankful to our institute for their help and support throughout our research work. We also appreciate the technical assistance provided by Mr. Chetan Yadav in conducting the experiments. Finally, we appreciate the helpful criticism provided by the anonymous reviewers.

References

- Becker, J. T., Boiler, F., Lopez, O. L., Saxton, J., & McGonigle, K. L. (1994). The natural history of Alzheimer's disease: description of study cohort and accuracy of diagnosis. Archives of neurology, 51(6), 585-594.
- Chakroborty, S., Roy, A., & Saha, G. (2006). Fusion of a complementary feature set with MFCC for improved closed set text-independent speaker identification. 2006 IEEE International Conference on Industrial Technology,
- Chen, J., Wang, Y., & Wang, D. (2014). A feature study for classification-based speech separation at low signal-to-noise ratios. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22(12), 1993-2002.
- Chen, Z., Zhu, P., Qiu, W., Guo, J., & Li, Y. (2023). Deep learning in automatic detection of dysphonia: Comparing acoustic features and developing a generalizable framework. Int J Lang Commun Disord.,58(2):279-294.
- De Lara, J. R. C. (2005). A method of automatic speaker recognition using cepstral features and vectorial quantization. Iberoamerican congress on pattern recognition.
- Ding, K., Chetty, M., Noori Hoshyar, A. et al. (2024). Speech based detection of Alzheimer's disease: a survey of AI techniques, datasets and challenges. Artif Intell Rev 57, 325.
- Dubois, B., Slachevsky, A., Litvan, I., & Pillon, B. (2000). The FAB: a frontal assessment battery at bedside. Neurology, 55(11), 1621-1626.
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. Journal of psychiatric research, 12(3), 189-198.
- Forbes-McKay, K. E., & Venneri, A. (2005). Detecting subtle spontaneous language decline in early Alzheimer's disease with a picture description task. Neurological sciences, 26(4), 243-254.
- Fraser, K. C., Meltzer, J. A., & Rudzicz, F. (2015). Linguistic features identify Alzheimer's disease in narrative speech. Journal of Alzheimer's disease, 49(2), 407-422.
- Guesmi, Dhouha, Njah, Hasna., & Ben Ayed Yassine (2026). Multi-level fusion framework for Alzheimer's recognition using speech and textual features. Neurocomputing (Elsevier), Volume 672.
- Inês Vigo, Luis Coelho et al. (2022). Speech- and Language-Based Classification of Alzheimer's Disease: A Systematic Review. Bioengineering, 9(1), 27.
- König, A., Linz, N., Tröger, J., Wolters, M., Alexandersson, J., & Robert, P. (2018). Fully automatic speech-based analysis of the semantic verbal fluency task. Dementia and geriatric cognitive disorders, 45(3-4), 198-209.
- Kuppusamy, Karthika., & Rajesh Anjana (2026). Alzheimer's detection using audio through deep learning techniques. Egyptian Informatics Journal (Elsevier), Volume 34.
- Mahajan, P., & Baths, V. (2021). Acoustic and language based deep learning approaches for Alzheimer's dementia detection from spontaneous speech. Frontiers in Aging Neuroscience, 13, 623607.
- Mathuranath, P., George, A., Cherian, P. J., Mathew, R., & Sarma, P. S. (2005). Instrumental activities of daily living scale for dementia screening in elderly people. International Psychogeriatrics, 17(3), 461-474.
- Mesulam, M., Wicklund, A., Johnson, N., Rogalski, E., Léger, G. C., Rademaker, A., Weintraub,

S., & Bigio, E. H. (2008). Alzheimer and frontotemporal pathology in subsets of primary progressive aphasia. Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society, 63(6), 709-719.

- Mueller, K. D., Koscik, R. L., Hermann, B. P., Johnson, S. C., & Turkstra, L. S. (2018). Declines in connected language are associated with very early mild cognitive impairment: Results from the Wisconsin Registry for Alzheimer's Prevention. Frontiers in Aging Neuroscience, 9, 437.
- Liu, J., Fu, F., Li, L., Yu, J., Zhong, D., Zhu, S., ... & Li, J. (2023). Efficient Pause Extraction and Encode Strategy for Alzheimer's Disease (AD) Detection Using Only Acoustic Features from Spontaneous Speech. Brain Sciences, 13(3), 477.
- Lyu, Xiaohong., Bai, Jiayuan ., Li, Pengfei., Zhang, Fan., Zhang, Xianglin., Gu, Yun., & Ning Liu (2025). Bridging Brain and Machine: A Neurocognitive Framework for Early Alzheimer's Detection Through Hierarchical Speech Processing. IEEE J Biomed Health Inform.10.
- Saeedi, S., Hetjens, S., Grimm, M.O.W. et al. (2024). Acoustic Speech Analysis in Alzheimer's Disease: A Systematic Review and Meta-Analysis. J Prev Alzheimers Dis 11, 1789–1797.
- Yang, Q., Li, X., Ding, X. et al. (2022). Deep learning-based speech analysis for Alzheimer's disease detection: a literature review. Alz Res Therapy 14, 186.