Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 2199–2207
https://doi.org/10.54941/ahfe1008446

Speech Analysis Integration for Assistance in Identification of Potential Alzheimer's Disease Early Symptoms

Lili-Marlene Camacho-Bustamante, Alejandro Diaz-Ruiz, and Sergio Navarro-Tuch

Tecnologico de Monterrey, Mexico City, 14380, Mexico

ABSTRACT

The number of Alzheimer's disease cases is increasing worldwide. Diagnosing this disease at its early stages remains a major challenge, particularly in low-resource countries. Early detection is essential to help patients and caregivers prepare for future care needs. Therefore, the development of affordable screening tools is of significant interest. In this study, a preliminary framework based on machine learning and speech analysis is explored as a non-invasive approach for identifying features associated with Alzheimer's disease. The Pitt Corpus from the DementiaBank database was used to extract linguistic and acoustic features. The analysis was conducted using the associated text transcripts. The proposed pipeline involved segmenting the transcripts and analyzing the resulting text data to evaluate temporal speech patterns and discourse structure. Key features, including speech pauses, repetitions, and grammatical complexity, were examined to assess differences between healthy controls and individuals with dementia. Subsequently, a neural network model was trained using the filtered feature set to identify early signs of cognitive decline. Hyperparameters were optimized to improve model performance and efficiency. These results demonstrate that subtle speech alterations associated with Alzheimer's disease can be identified despite the significant class imbalance present in the dataset.

Keywords: Alzheimer's disease, Assistive technology, Speech analysis, Pre-diagnosis, Biomarkers

INTRODUCTION

Dementia has become an increasingly prevalent global health concern due to population aging. It is estimated that, by 2050, approximately 153 million people worldwide will be living with dementia (Nichols et al., 2022). Alzheimer's disease (AD) is the most common form of dementia, accounting for approximately 60–80% of all cases ("2026 Alzheimer's Disease Facts and Figures," 2026). It is considered one of the most common conditions affecting individuals over the age of 65, although it can also occur at younger ages in cases of early-onset AD (Mendez, 2019; World Health Organization, 2025). As the disease progresses, it affects not only memory but also language and the ability to perform activities of daily living (Clemmensen et al., 2020).

Received September 25, 2026; Revised November 3, 2026; Accepted November 15, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

In advanced stages, individuals may experience significant impairments in their ability to communicate and interact with their surrounding environment (Woodward, 2013).

Early diagnosis has emerged as one of the primary priorities addressed in most public policy strategies related to dementia, highlighting its critical importance, as it is essential for slowing disease progression and improving patients' quality of life (Dubois et al., 2016; World Health Organization, 2017). However, achieving an early diagnosis remains challenging due to the variability of symptoms and their overlap with other forms of dementia (Turner et al., 2020). Despite advances in neuroimaging techniques and biomarker-based approaches, these methods are often costly and are not accessible in all clinical settings (Suresh et al., 2026).

Recent research has explored the potential of detecting AD through language and voice analysis. Studies suggest that alterations in verbal fluency, word production, and discourse coherence may emerge years before other clinically evident symptoms (Kokje et al., 2022). These findings indicate that language analysis could provide an early window for detecting cognitive decline and may offer a promising, accessible, and non-invasive approach to supporting the early detection of AD (Blazquez-Folch et al., 2026).

Artificial intelligence (AI) is increasingly becoming a key tool in healthcare, contributing to improved diagnostic accuracy and patient outcomes (Gu et al., 2025). Its ability to process and analyze large volumes of data enables faster and more precise diagnostic assessments (Taiyeb Khosroshahi et al., 2025). In the context of AD, AI plays a critical role in early detection through machine learning models capable of identifying subtle patterns that may be associated with the disease (Mohsen, 2025). This study aims to explore an initial approach to identifying and examining the variables that may support the identification of individuals with Alzheimer's disease using voice-based analysis.

BACKGROUND

Key studies highlight the potential of Machine Learning (ML) and Natural Language Processing (NLP) in analysing speech patterns (Shakeri & Farmanbar, 2025). Some approaches show that ML models can classify Alzheimer's patients with high accuracy by analyzing features like fluency and lexical complexity (Campbell et al., 2021; Caro et al., 2025; Siddiqui et al., 2026), while others focus on identifying linguistic markers such as syntactic errors and discourse incoherence (Nyongesa et al., 2025).

Standardized datasets, like the ADReSS Challenge corpus, have been instrumental in benchmarking and improving speech analysis models by providing structured samples from Alzheimer's patients and healthy controls. These resources have enabled the refinement of ML algorithms to detect subtle speech variations across linguistic and cultural contexts (Luz et al., 2020).

Multimodal approaches that integrate speech data with neuroimaging and genetic information provide a more comprehensive diagnostic framework. For example, one study demonstrated that combining MRI, clinical, and genetic data with deep learning methods significantly improves diagnostic accuracy achieving 80% precision when combining voice analysis with neuroimaging (Golovanevsky et al., 2022). Furthermore, recent approaches have highlighted the potential of large language models, such as GPT-like architectures, which achieved 78% precision in identifying nuanced linguistic patterns (Mo et al., 2024).

Another approach is the use of deep sequential neural networks to detect mild cognitive impairment (MCI) from spontaneous speech, obtaining a mean classification accuracy of 83 percent (Themistocleous et al., 2018).

These advancements highlight the transformative potential of ML and speech analysis in Alzheimer's early detection, providing a non-invasive, scalable, and efficient alternative to traditional diagnostic methods.

DEVELOPMENT

The Pitt Corpus is a dataset that includes both audio recordings and transcripts from 397 participants (Becker, 1994). The participants were initially classified into three groups: 104 control participants, 208 participants with dementia, and 85 participants with an unknown diagnosis. The dataset provides recordings and corresponding transcripts of participants while completing four different tasks, including the Cookie Theft picture description task, verbal fluency, recall, and sentence construction tasks. For the present study, the Cookie Theft task was selected because it elicits spontaneous speech and has been shown to be particularly useful for identifying changes associated with cognitive decline.

Inclusion Criteria

- Participants labeled with a diagnosis of probable Alzheimer's disease (AD) or possible AD.
- Participants belonging to the control group.

Exclusion Criteria

- Participants diagnosed with other types of dementia, Parkinson's disease, vascular dementia, or dementia with Lewy bodies.
- Participants diagnosed with mild cognitive impairment (MCI).

Feature Extraction

For this initial approach, the transcripts associated with each audio recording were used. The transcripts are provided in CHAT format and are accompanied by documentation that explains the meaning and annotation

conventions associated with each transcription code. A set of linguistic and speech-related features was extracted from the transcripts, including the number of unfilled pauses (short, medium, and long), filled pauses, interruptions, self-corrections, filler words, total number of words, number of unique words, total number of sentences, pronoun rate, syntactic connector rate, and average word length.

Feature Selection

A total of 18 features were initially extracted. The mean values for each feature were subsequently calculated for each group, as presented in Table 1. Features related to retracing, reformulation, and false starts were then excluded because none of these events were detected in the available recordings.

Subsequently, to reduce the risk of overfitting and identify the most relevant features, the statistical significance of each explanatory variable with respect to the target variable was evaluated by calculating its p-value. Only features demonstrating statistical significance were retained for further analysis. The selected features are presented in Figure 1.

Table 1: Features.

Feature	AD Group	Control Group
Self-interruptions	0.047059	0.078189
Unfilled_pauses_short	1.129412	0.728395
Unfilled_pauses_medium	0.431373	0.279835
Unfilled_pauses_long	0.105882	0.069959
Repetition	1.796078	0.724280
Retracing	0.0	0.0
Reformulation	0.0	0.0
False_start	0.0	0.0
Filled_pauses	3.329412	3.374486
Trailing_off	2.192157	1.304527
Total_words	110.250980	119.497942
Unique_words	69.694118	76.748971
Repeated_words	17.984314	17.864198,
Total_sentences	24.992157	25.325103
Average_sentence_length	4.442078	4.776584
Pronoun_rate	0.063216	0.053599
Connector_rate	0.057298	0.048364
Average_word_length	5.114902	5.065967

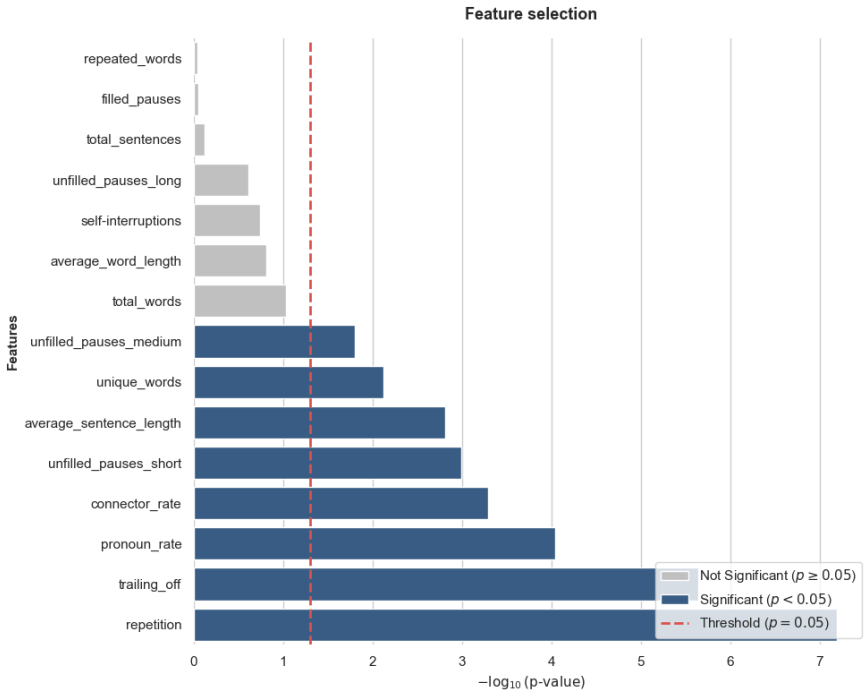


Figure 1: Selection of input features using statistical significance testing.

Neural Network

A sequential architecture neural network was created in order to classify the subjects as either having possible Alzheimer's disease (AD) or serving as controls, based on 8 features and involving approximately 498 cases (255 cases of AD and 243 control cases). Before any hyperparameter tuning took place, the dataset was divided in a stratified way into a development and an independent test subset, thus making sure that the test subset remained completely separate from the optimization process. Rather than using just one fixed set of settings, we tested a number of different combinations involving batch size, the number of neurons per layer, the optimizer, and dropout.

RESULTS

The hyperparameter search identified batch size = 16, a three-layer architecture ([16, 8, 4] neurons), the Adam optimizer, and a dropout rate of 0.3 as the best-performing configuration. This configuration achieved a mean cross-validated precision of 0.773 and a mean area under the curve (AUC) of 0.806 across the 5 folds of the development subset (n = 398). It was then retrained on the full development subset and evaluated once on the independent test subset (n = 100), which had remained untouched throughout the search process. On this test subset, the model achieved a precision of 0.660 and an AUC of 0.677.

The decrease observed between cross-validated performance during model selection (AUC = 0.806) and performance on the test subset (AUC = 0.677) is consistent with the expected optimism of hyperparameter search, in which the best-performing configuration among many candidates tends to reflect some degree of favorable variation in its validation folds. This gap may also be partly attributable to the limited size of the test subset (n = 100), which increases the variance of point estimates such as AUC. The test-set result is considered the more reliable indicator of the model's expected performance on unseen data, as it was obtained without any influence from the hyperparameter selection process.

Table 2: Best-performing hyperparameter configurations.

Rank	Batch Size	Architecture	Optimizer	Dropout	Mean Precision
1	16	[16, 8, 4]	Adam	0.3	0.773
2	16	[8]	RMSprop	0.3	0.758
3	16	[16, 8]	RMSprop	0.3	0.757
4	32	[8]	RMSprop	0.2	0.752
5	16	[16, 8]	RMSprop	0.2	0.746

Table 2 presents the five best-performing hyperparameter configurations, ranked by mean precision across the 5-fold cross-validation performed on the development subset. Notably, the five top-ranked configurations exhibited relatively similar performance, suggesting that model performance was reasonably stable across this region of the hyperparameter space rather than being driven by a single optimal combination. RMSprop appeared in four of the five top configurations, indicating that it performed consistently well even though Adam achieved the highest overall precision. Similarly, a dropout rate of 0.2 or 0.3 was present in all five top configurations, which may reflect the benefit of moderate regularization given the relatively small feature set used in this study.

CONCLUSION

This study demonstrates the significant potential of integrating speech analysis with machine learning for the early diagnosis of Alzheimer's disease. Linguistic features are systematically extracted from spontaneous speech, enabling the model to identify subtle indicators of cognitive decline and offering a non-invasive, accessible diagnostic approach. Hyperparameter optimization enhances the neural network's capacity to recognize complex speech patterns and achieve high diagnostic accuracy.

This study does not serve as a definitive screening method but instead aims to inform the development of such tools by identifying potentially valuable features. Substantial further work is required, including the integration of additional variables and comprehensive analysis of the dataset, particularly regarding the mild cognitive stage, which could facilitate earlier detection.

Beyond its immediate implications, this research lays the groundwork for future advancements in Alzheimer's detection and monitoring.

First, developing more robust models based on advanced architectures such as recurrent neural networks or transformers will allow more precise capture of the complex linguistic alterations that characterize cognitive decline.

ACKNOWLEDGMENT

The authors would like to acknowledge the CIMB Laboratory at Tecnologico de Monterrey, Mexico City Campus, for its support of this research. The study has received financial assistance from Tecnologico de Monterrey through scholarship A01551004 and from the Secretaría de Ciencia, Humanidades, Tecnología e Innovación (Secihti) under CVU No. 1151200.

REFERENCES

2026 Alzheimer's disease facts and figures. (2026). *Alzheimer's & Dementia*, *22*(4). https://doi.org/10.1002/alz.71345

Becker, J. T. (1994). The Natural History of Alzheimer's Disease. *Archives of Neurology*, *51*(6), 585. https://doi.org/10.1001/archneur.1994.00540180063015

Blazquez-Folch, J., Calm, B., Hinojosa-Calleja, A., García-Gutiérrez, F., Alegret, M., Muñoz, N., Cano, A., Fernández, M. V., Miguel, A., Solivar, A., De Rojas, I., Valenzuela-Seba, A., García-González, P., Puerta, R., Olivé, C., Capdevila-Bayo, M., Muñoz-Morales, Á., Bayón-Buján, P., Montrreal, L., … Valero, S. (2026). Listening for Alzheimer's clues: Machine learning analysis of multidomain speech features for cognitive impairment screening. *Frontiers in Aging Neuroscience*, *18*. https://doi.org/10.3389/fnagi.2026.1816747

Campbell, E. L., Mesía, R. Y., Docío-Fernández, L., & García-Mateo, C. (2021). Paralinguistic and linguistic fluency features for Alzheimer's disease detection. *Computer Speech & Language*, *68*, 101198. https://doi.org/10.1016/j.csl.2021.101198

Caro, I., Pérez, G. N., Bisé, J. V., Ponferrada, J., Ferrante, F. J., Welford, A. S., Gauder, L., Ferrer, L., Ibanez, A., Slachevsky, A., & Garcia, A. M. (2025). Testing speech biomarkers against cognitive and neural signatures of Alzheimer's dementia. *Alzheimer's & Dementia*, *21*(S2). https://doi.org/10.1002/alz70856_107220

Clemmensen, F. K., Hoffmann, K., Siersma, V., Sobol, N., Beyer, N., Andersen, B. B., Vogel, A., Lolk, A., Gottrup, H., Høgh, P., Waldemar, G., Hasselbalch, S. G., & Frederiksen, K. S. (2020). The role of physical and cognitive function in performance of activities of daily living in patients with mild-to-moderate Alzheimer's disease - a cross-sectional study. *BMC Geriatrics*, *20*(1), 513. https://doi.org/10.1186/s12877-020-01926-9

Dubois, B., Padovani, A., Scheltens, P., Rossi, A., & Dell'Agnello, G. (2016). Timely Diagnosis for Alzheimer's Disease: A Literature Review on Benefits and Challenges. *Journal of Alzheimer's Disease : JAD*, *49*(3), 617–631. https://doi.org/10.3233/JAD-150692

Golovanevsky, M., Eickhoff, C., & Singh, R. (2022). Multimodal attention-based deep learning for Alzheimer's disease diagnosis. *Journal of the American Medical Informatics Association*, *29*(12), 2014–2022. https://doi.org/10.1093/jamia/ocac168

Gu, Z., Ge, B., Wang, Y., Gong, Y., & Qi, M. (2025). Artificial intelligence technologies for enhancing neurofunctionalities: A comprehensive review with applications in Alzheimer's disease research. *Frontiers in Aging Neuroscience*, *17*. https://doi.org/10.3389/fnagi.2025.1609063

Kokje, E., Celik, S., Wahl, H.-W., & von Stutterheim, C. (2022). Can discourse processing performance serve as an early marker of Alzheimer's disease and mild cognitive impairment? A systematic review of text comprehension. *European Journal of Ageing*, *19*(1), 3–18. https://doi.org/10.1007/s10433-021-00619-5

Luz, S., Haider, F., Fuente, S. de la, Fromm, D., & MacWhinney, B. (2020). Alzheimer's Dementia Recognition Through Spontaneous Speech: The ADReSS Challenge. *Interspeech 2020*, 2172–2176. https://doi.org/10.21437/Interspeech.2020-2571

Mendez, M. F. (2019). Early-onset Alzheimer Disease and Its Variants. *Continuum*, *25*(1), 34–51. https://doi.org/10.1212/CON.0000000000000687

Mo, T., Lam, J. C. K., Li, V. O. K., & Cheung, L. Y. L. (2024). *Leveraging Large Language Models for Identifying Interpretable Linguistic Markers and Enhancing Alzheimer's Disease Diagnostics*. https://doi.org/10.1101/2024.08.22.24312463

Mohsen, S. (2025). Alzheimer's disease detection using deep learning and machine learning: A review. *Artificial Intelligence Review*, *58*(9), 262. https://doi.org/10.1007/s10462-025-11258-y

Nichols, E., Steinmetz, J. D., Vollset, S. E., Fukutaki, K., Chalek, J., Abd-Allah, F., Abdoli, A., Abualhasan, A., Abu-Gharbieh, E., Akram, T. T., Al Hamad, H., Alahdab, F., Alanezi, F. M., Alipour, V., Almustanyir, S., Amu, H., Ansari, I., Arabloo, J., Ashraf, T., … Vos, T. (2022). Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: An analysis for the Global Burden of Disease Study 2019. *The Lancet Public Health*, *7*(2), e105–e125. https://doi.org/10.1016/S2468-2667(21)00249-8

Nyongesa, C. A., Hogarth, M., & Pa, J. (2025). Artificial intelligence-driven natural language processing for identifying linguistic patterns in Alzheimer's disease and mild cognitive impairment: A study of lexical, syntactic, and cohesive features of speech through picture description tasks. *Journal of Alzheimer's Disease*, *106*(1), 120–138. https://doi.org/10.1177/13872877251339756

Shakeri, A., & Farmanbar, M. (2025). Natural language processing in Alzheimer's disease research: Systematic review of methods, data, and efficacy. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *17*(1). https://doi.org/10.1002/dad2.70082

Siddiqui, A., Kathiresan, T., Opler, M., Cohen, A., Alber, J., Snyder, P. J., & Vogel, A. P. (2026). The Utility of Speech and Language Analytics for Screening Alzheimer's Disease. *Neuro-Degenerative Diseases*, 1–13. https://doi.org/10.1159/000552094

Suresh, S., Maffei, L., Bauermeister, S., & Raymont, V. (2026). Blood biomarkers for diagnosis and differential diagnosis of Alzheimer's disease in real-world clinical populations: A systematic review. *Journal of Alzheimer's Disease : JAD*, *109*(4), 1574–1589. https://doi.org/10.1177/13872877251408510

Taiyeb Khosroshahi, M., Morsali, S., Gharakhanlou, S., Motamedi, A., Hassanbaghlou, S., Vahedi, H., Pedrammehr, S., Kabir, H. M. D., & Jafarizadeh, A. (2025). Explainable Artificial Intelligence in Neuroimaging of Alzheimer's Disease. *Diagnostics (Basel, Switzerland)*, *15*(5). https://doi.org/10.3390/diagnostics15050612

Themistocleous, C., Eckerström, M., & Kokkinakis, D. (2018). Identification of Mild Cognitive Impairment From Speech in Swedish Using Deep Sequential Neural Networks. *Frontiers in Neurology*, *9*. https://doi.org/10.3389/fneur.2018.00975

Turner, R. S., Stubbs, T., Davies, D. A., & Albensi, B. C. (2020). Potential New Approaches for Diagnosis of Alzheimer's Disease and Related Dementias. *Frontiers in Neurology*, *11*, 496. https://doi.org/10.3389/fneur.2020.00496

Woodward, M. (2013). Aspects of communication in Alzheimer's disease: Clinical features and treatment options. *International Psychogeriatrics*, *25*(6), 877–885. https://doi.org/10.1017/S1041610213000318

World Health Organization. (2017). *Global Action Plan on the Public Health Response to Dementia 2017–2025*. World Health Organization. https://www.who.int/publications/i/item/9789241513487

World Health Organization. (2025, October 1). *Ageing and health*. World Health Organization News. https://www.who.int/news-room/fact-sheets/detail/ageing-and-health