# A Fuzzy Pause-Aware Framework for Alzheimer's Disease Detection from Spontaneous Speech

1st Rishabh
*Department of Computer Science*
*University of Delhi*
Delhi, India
0009-0001-4913-227X

2nd Kuldeep Singh
*Department of Computer Science*
*University of Delhi*
Delhi, India
0000-0001-5289-362X

3rd Dhirendra Kumar
*Department of Applied Mathematics*
*Delhi Technological University*
Delhi, India
0000-0002-8902-5022

4th Yogendra Meena
*School of Computer and Systems Sciences*
*Jawaharlal Nehru University*
New Delhi, India
0009-0009-9578-6790

***Abstract*—Early detection of Alzheimer's disease via speech-based biomarkers offers a scalable, non-invasive alternative to traditional diagnostics. Prior studies have examined semantic or prosodic features separately, but the unified integration of pause cues to the semantic representation remain under-explored. The paper introduces a fuzzy pause-aware framework for the detection of Alzheimer's disease (AD) using spontaneous speech. The suggested technique improves language representations by directly modeling pause dynamics taken from word-level transcriptions based on Whisper. Using fuzzy C-means clustering, pause durations are sorted into groups that make sense semantically. The resultant pause information is added using pause tokens and fuzzy membership-based pause embeddings. These are then combined with frozen BERT-based lexical embeddings to capture both semantic and prosodic signals associated to cognitive disfluency. A Transformer-based model with attention pooling and pause statistics fusion is then used to arrange the merged representations. Experiments conducted on the ADReSSo dataset show that the suggested method yields an accuracy of 77.46%, which is better than several state-of-the-art baselines. The findings validate that adaptive, data-driven modeling of pause patterns substantially improves the discriminative efficacy of speech-based Alzheimer's disease detection systems. This work underscores the importance of integrating prosodic information with linguistic representations to maintain a robust cognitive.**

***Index Terms*—Alzheimer's disease detection, pause-aware modeling, fuzzy C-means clustering, Whisper, BERT embeddings, Transformer classifier, attention pooling, ADReSSo.**

## I. Introduction

Alzheimer's Disease (AD) is a neurodegenerative condition where patients experience aphasia caused by brain lesion. Traditional methods for detecting Alzheimer's disease are time-consuming and costly, imposing a significant burden on families and society [1], [2]. Alzheimer's typically begins with subtle symptoms, and it is difficult to recognize at early stage. People may start forgetting recent events, conversations, or familiar places. They might struggle with everyday tasks. Language difficulties also emerge, patients often take longer pauses during conversations, lose their train of thought, or have trouble finding the right word. Cognitive impairments are generally recognized through various methods, including clinical evaluations and cognitive assessments [3] (such as MMSE and MoCA), and imaging modalities [4] (such as MRI, and PET). These methods tend to be costly and may not be readily available. Since, speech-based detection of Alzheimer's Disease offers advantages such as low cost, non-invasiveness, and convenient operation [5], [6], it holds great potential for broad application.

## II. Related Work

The detection of Alzheimer's Disease (AD) through speech recordings has gained considerable attention in recent research, has led to the exploration of natural language processing techniques for reliable AD detection [7]. Prior works [8]–[11] have focused on acoustic and linguistic features, but temporal markers such as pauses remain underutilized, despite their established role in reflecting hesitation, memory retrieval, and slowed processing in impaired individuals.

Yuan et al. [12] fine-tune the pre-trained Transformer language models (BERT and ERNIE) on spontaneous speech transcripts to detect AD, focusing on linguistic disfluencies. Zhu et al. [13] proposes WavBERT, a hybrid model that combines Wav2vec, and BERT, that leverages both semantic and non-semantic information from speech. Also extended WavBERT model to estimate inter-word pause duration, and to include the non-semantic information in dementia detection. Haulcy et al. [14] examined the effectiveness of both audio-based and text-based features extracted from speech for AD classification and MMSE score prediction using ADReSS dataset. Demonstrated the strong performance of pre-trained BERT embedding as feature for both classification and regression. Liu et al. [8] proposed a non-text based, acoustic-only method for detecting AD using pause-patterns in spontaneous speech. Extracted Voice Activity Detection (VAD) based pause features which introduces a binary/non-pause feature sequence from raw audio. Roshabzamir et al. [15] used pre-trained transformer-based language models (like BERT, XLNet, and

XLM) eliminating the need for hand-crafted features. Among these, BERTLarge sentence-level embeddings with logistic regression. Pu et al. [16] introduces a novel method to enhanve AD detection from spontaneous speech by integrating pause information into a language model (BERT), capturing both semantic and temporal features. Wang et al. [17] introduces a language-independent, interpretable and statistical rule-based method for detecting AD across multiple languages using scale-based z-score features.

## III. Proposed Work

Speech has gained attention as a promising biomarker for early detection of AD. Pause patterns have been consistently identified as critical indicators, and their correlation with cognitive load, hesitation, and language planning deficits seen in AD patients has been well established. Recent studies have utilized fixed, hard-coded pause thresholds. To address these limitations, this work introduces a cluster-based approach using fuzzy c-means clustering. It categories pauses into three categories (short, medium, and long). This clustering-based method aims to reflect the speech features, associated with cognitive decline more accurately. These pause categories are then transformed into pause tokens and incorporated into transcription to enable pause-aware embedding extraction using a pre-trained BERT model. The resulting embedding features are subsequently used for classification with a Transformer-based model to distinguish AD from cognitively normal (CN) speech. The overall architecture of the proposed pause-aware Alzheimer's detection system is illustrated in Figure 1. It integrates pause clustering and BERT-style embeddings into a Transformer classifier.

### *A. Dataset*

This study utilizes Alzheimer's dementia recognition through spontaneous speech only (ADReSSo) [18] dataset, introduced at Interspeech 2021, consisting of 237 speech recordings acquired from a picture description task for AD and CN. The training set contains 166 files (87 AD, 79 CN), and the test set includes 71 samples (36 AD, 35 CN).

### *B. Speech Transcription and Pause Extraction*

Each speech recordings are first transcribed into word-level sequence using Whisper, an automatic speech recognition (ASR) model. It returns the time-aligned transcription, where each word is associated with its start and end timestamps. This enables accurate modeling of temporal speech dynamics. Pause are extracted using the word-level timestamps, as the temporal gap between the end time of a word and the start of the subsequent word. A minimum duration threshold of 50ms were removed to avoid noise. Each detected pause is represented with its duration and the relative position within the word sequence. These pause information helps capturing hesitation and processing time associated with cognitive impairment, form the basis for pause modeling and embedding.

### *C. Pause Modeling via Hard Thresholding and Fuzzy Clustering*

In ADReSSo speech dataset , the presence of pauses can reflect symptoms of cognitive impairments, as they are closely related to cognitive load, hesitation, and language planning difficulties [19], [20]. To incorporate pause features into our framework, pause durations are extracted from speech samples using word-level timestamps obtained from the Whisper model. Given a sequence of words with start and end times, pauses are computed as:

$$P = \{\delta_i = s_{i+1} - e_i \mid \delta_i \geq \theta\},$$

where $s_i$ and $e_i$ denote the start and end times of the $i^{\text{th}}$ word, respectively, and $\theta = 50$ms is used as the minimum pause duration threshold.

*1) Hard Threshold-Based Pause Categorization:* Pause durations are categorized using fixed, manually defined thresholds. Each pause $\delta_i$ is assigned to one of the category (e.g., short, medium, or long) based on predefined duration ranges [19], [20]. While this approach is simple and computationally efficient, it assumes sharp boundaries between pause types. It can lead to misinformation about the temporal fluency of the conversation.

*2) Fuzzy C-Means Clustering Based Pause Modeling:* To overcome the limitations of hard thresholding, Fuzzy C-Means (FCM) clustering is employed to enable soft partitioning of pause durations. The FCM partitions the set of pause durations into $C$ clusters by minimizing the objective function:

$$J_m = \sum_{i=1}^{M} \sum_{j=1}^{C} u_{ij}^m \|\delta_i - c_j\|^2,$$

where $c_j$ denotes the centroid of the $j^{\text{th}}$ cluster, $u_{ij} \in [0, 1]$ represents the degree of membership of pause duration $\delta_i$ to cluster $j$, and $m > 1$ controls the fuzziness of the partition. In this study, $m$ is set to 2, and the number of clusters $C$ is varied from 2 to 5, to study its impact on performance.

Instead of assigning each pause to a single cluster, the membership vector $\mathbf{u}(\delta) = [u_1(\delta), \ldots, u_C(\delta)]$ is employed and used to construct a pause embedding via a weighted combination of learnable cluster prototypes. This formulation better reflect natural speech variability while preserving the pause related information. Figure 2 visualizes the pause duration clusters obtained using FCM.

### *D. Pause-Aware Embedding Representation*

To incorporate temporal pause information into lexical representations, a pause-aware embedding mechanism is introduced on top of frozen BERT embeddings. Given a Whisper-generated word sequence augmented with pause tokens, the sequence is first encoded using a pretrained BERT model. Let $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_T]$, where $\mathbf{h}_t \in \mathbb{R}^{768}$ denotes the contextual embedding of the $t^{\text{th}}$ token.

For each pause occurrence, the corresponding pause duration $\delta$ is mapped to a fuzzy membership vector $\boldsymbol{\mu}(\delta) \in \mathbb{R}^C$ obtained from Fuzzy C-Means clustering. A learnable pause

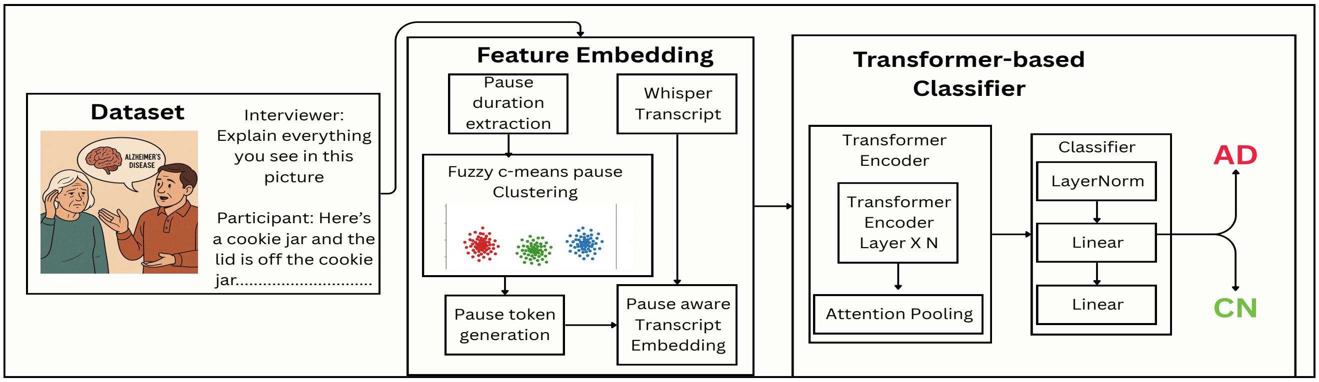
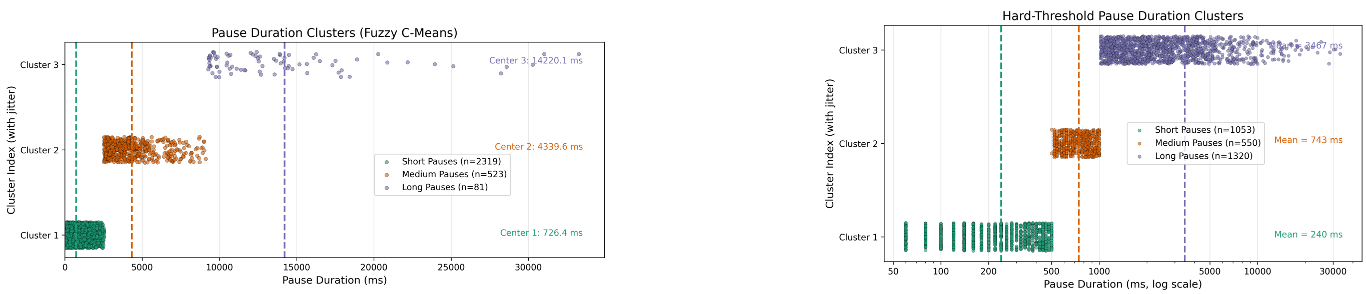



Fig. 1: Overview of the proposed system for Alzheimer's Disease (AD) classification from speech. The pipeline consists of three stages: (1) Dataset comprising picture description tasks, (2) Feature Embedding via Whisper transcription and fuzzy C-means clustering of pauses to generate pause-aware embeddings, and (3) Transformer-based classification using attention pooling and fully connected layers to distinguish AD from cognitively normal (CN) participants.




(a) Fuzzy C-Means based pause clustering.

(b) Hard-threshold based pause clustering (log scale).

Fig. 2: Comparison of pause duration clustering approaches for both Fuzzy C-Means based and Hard-threshold pause clustering.

embedding is then computed as a weighted combination of cluster prototypes:

$$\mathbf{p}(\delta) = \boldsymbol{\mu}(\delta)^\top \mathbf{W}_p,$$

where $\mathbf{W}_p \in \mathbb{R}^{C\times 768}$ is a trainable pause embedding matrix.

The pause-aware token representation is formed by injecting the pause embedding into the corresponding BERT embedding:

$$\tilde{\mathbf{h}}_t = \begin{cases} \mathbf{h}_t + \mathbf{p}(\delta), & \text{if token } t \text{ corresponds to a pause,} \\ \mathbf{h}_t, & \text{otherwise.} \end{cases}$$

This additive fusion strategy preserves the original lexical semantics while enriching the representation with temporal pause information. The resulting pause-aware embedding sequence $\tilde{\mathbf{H}}$ is then passed to the downstream Transformer-based classifier.

### *E. Transformer-Based Classification with Pause Statistics*

The pause-aware embedding sequence is classified using a Transformer-based architecture designed to model long-range contextual dependencies. Given the pause-enriched token representations $\tilde{\mathbf{H}} \in \mathbb{R}^{T\times 768}$, a multi-layer Transformer encoder is applied to further capture higher-order interactions among lexical and pause-related cues.

To obtain a fixed-dimensional utterance-level representation, an attention pooling mechanism is employed.

Specifically, a learnable attention layer assigns importance weights to each token representation, producing a weighted sum that emphasizes diagnostically salient regions of the sequence. This results an embedding$\mathbf{z} \in \mathbb{R}^{768}$. In addition to token-level representations, global pause statistics are incorporated to capture coarse temporal speech features. For each recording, three pause-level features are computed: (i) the total number of pauses, (ii) the mean pause duration, and (iii) the standard deviation of pause durations. These statistics are normalized and concatenated with the pooled Transformer representation:

$$\mathbf{z}' = [\mathbf{z}; \mathbf{s}],$$

where $\mathbf{s} \in \mathbb{R}^3$ denotes the pause statistics vector.

The fused representation $\mathbf{z}'$ is passed through a fully connected classification head to predict the cognitive status of the speaker. This joint modeling of contextual embeddings and global pause statistics enables the classifier to leverage both fine-grained lexical patterns and coarse temporal speech behavior relevant to Alzheimer's disease detection.

## IV. Experimental Results

### A. Evaluation Metrics

In this study, the evaluation metrics used to validate our proposed model are same as suggested in the ADReSSo challenge. AD classification will be evaluated through the following metrics:

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{N}} \tag{1}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{2}$$

$$\text{F1-score} = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} \ + \text{FP} + \text{FN}} \tag{3}$$

where, N is the number of patients, TP is the number of true positives, TN is the number if true negatives, FP is the number of false positive, and FN is the number of false negative.

### B. Experimental Setup

The proposed pause-aware classification framework was tested on the ADReSSo dataset. We used 10-fold stratified cross-validation on the training set to make sure that the class distribution was even across all folds. The AdamW optimizer with a learning rate of $3 \times 10^{-4}$ was used to train the model for 30 epochs for each fold, with weighted cross-entropy loss. Standard scaling based on the training data was used to normalize the pause-level statistical characteristics. Experiments were performed in a Kaggle environment utilizing an NVIDIA Tesla GPU and the PyTorch framework. To speed up repeated runs and make sure they could be repeated, whisper transcriptions were stored.

### C. Ablation Study

Table I summarizes the model variations, indicating the presence or absence of lexical embeddings, pause tokens, fuzzy membership modeling, pause embeddings, and the number of pause clusters applied. Table II displays the outcome for each version given in Table I, illustrating the relevance of FCM pause token integration.

The model M1 utilized BERT-based lexical embeddings, without including pause information and achieves an accuracy of 71.83% with an F1-score of 69.70%. Models M2 and M3 introduces hard-threshold pause tokens with fixed duration bounds. M2 employs thresholds of $[500, 1000]$ ms to distinguish short with long pauses, while M3 uses wider thresholds of $[500, 1500]$ ms to capture longer hesitation patterns. These variants might increase recall in some situations but they don't always improve overall performance. This suggests that fixed pause boundaries don't accurately represent the differences across speakers.

On the other hand, models M4–M7 apply fuzzy C-means clustering to represent pause lengths adaptively. M4 associates with 2 clusters, M5 with 3 and so on. Among these, M5 (three pause clusters) obtains the highest performance, with an accuracy of 77.46% and an F1-score of 78.38%, with a considerable improvement in recall (82.86%). Increasing the number of clusters beyond three leads to no improvement, demonstrating that too much detailing may produce noise rather than useful difference. Overall, the ablation results suggest that adaptive pause modeling using fuzzy clustering is more successful than hard-coded thresholding and that a reasonable number of pause clusters offers the optimal trade-off between expressiveness and robustness.

TABLE I: Ablation configurations showing the presence (✓) or absence (×) of pause-aware components and the number of pause clusters used.

| ID | Text (BERT) | Pause Token | Fuzzy Membership | Pause Embed. | # Clusters |
|---|---|---|---|---|---|
| M1 | ✓ | × | × | × | – |
| M2 | ✓ | ✓ | × | ✓ | – |
| M3 | ✓ | ✓ | × | ✓ | – |
| M4 | ✓ | ✓ | ✓ | ✓ | 2 |
| M5 | ✓ | ✓ | ✓ | ✓ | 3 |
| M6 | ✓ | ✓ | ✓ | ✓ | 4 |
| M7 | ✓ | ✓ | ✓ | ✓ | 5 |

### D. Comparison with State-of-the-Art

Table III compares the proposed methodology with the existing SOTA methods on the ADReSSo dataset. Existing methods, including CNN-based models and VAD-based pause features with eGeMAPS, results in accuracies between 64.79% to 71.74%. The proposed pause-aware framework consistently outperforms these approaches. In particular, the M5 configuration results the highest accuracy (77.46%) and F1-score (78.38%), highlighting the value of incorporating pause dynamics directly into contextualized language embeddings rather than considering pauses as isolated acoustic variables. These results imply that modeling pauses as soft, cluster based semantic signals permits more efficient merging of lexical and prosodic information, leading to enhanced identification of cognitive impairment.

### E. Discussion

The experimental results suggest that pause patterns provide a significant cue for speech-based Alzheimer's identification when modeled adequately. Fixed pause thresholds provide little flexibility, but fuzzy clustering captures complex temporal fluctuations associated with cognitive stress and speech planning deficiencies.

Moreover, integrating pause information directly into token-level representations allows the Transformer classifier to leverage pause–word interactions that are unavailable to typical acoustic feature pipelines. The continuous improvement found with fuzzy clustering-based models shows the necessity of adaptive, data-driven pause modeling for robust and interpretable clinical speech analysis.

**Computational Complexity Analysis:** The hard-threshold method involves simple operations with negligible cost. The proposed approach introduces a small additional computational overhead due to clustering; however, this overhead is minimal and does not significantly affect the overall pipeline.

TABLE II: Ablation results on the ADReSSo21 diagnostic test set.

| ID | Accuracy (%) | Precision (%) | Recall (%) | Specificity (%) | F1-score (%) |
|---|---|---|---|---|---|
| M1 | 71.83 | 74.19 | 65.71 | 77.78 | 69.70 |
| M2 | 73.24 | 72.22 | 74.29 | 72.22 | 73.24 |
| M3 | 71.83 | 70.27 | 74.29 | 69.44 | 72.22 |
| M4 | 64.79 | **75.00** | 42.86 | **86.11** | 54.55 |
| M5 | **77.46** | 74.36 | 82.86 | 72.22 | **78.38** |
| M6 | 69.01 | 63.83 | **85.71** | 52.78 | 73.17 |
| M7 | 71.83 | 70.27 | 74.29 | 69.44 | 72.22 |

TABLE III: Comparison with state-of-the-art methods on ADReSSo dataset

| Method | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| **Baseline [18]** | 64.79% | –% | –% | –% |
| **4-Layer FCNN [17]** | 71.74% | 69.57% | 72.73% | 71.11% |
| **VAD Pause feature and eGeMAPS [8]** | 70.70% | –% | –% | –% |
| **Proposed (M1)** | 71.83% | 74.19% | 65.71% | 69.70% |
| **Proposed (M3)** | 71.83% | 70.27% | 74.29% | 72.22% |
| **Proposed (M5)** | **77.46%** | **74.36%** | **82.86%** | **78.38%** |

### *F. Conclusion and Future Work*

We propose a pause-aware approach for detecting Alzheimer's disease using spontaneous speech in this study. The proposed approach captures temporal disfluencies that indicate cognitive impairment by combining fuzzy C-means clustering of pause durations with Whisper-based transcriptions. The pause information is added by using distinct pause tokens and fuzzy membership-based pause embeddings, which are combined with frozen BERT-based lexical representations. We employ a Transformer-based architecture with attention pooling and pause statistics fusion to classify the resulting pause-aware embeddings. This works better on the ADReSSo dataset than previous baselines.

The experimental results show that adaptive, data-driven pause modeling works better than set hard-threshold techniques. The findings additionally show that a reasonable number of pause clusters gives the optimum balance between robustness and discriminative power. Future research will investigate more structured representations of pause dynamics, such graph-based or hierarchical Transformer models, to improve diagnostic accuracy and model interpretability.

## References


[1] World Health Organization and Alzheimer's Disease International. *Dementia: A Public Health Priority*. World Health Organization, Geneva, 2012. Accessed: 2025-07-10.

[2] Xiaoke Qi, Qing Zhou, Jian Dong, and Wei Bao. Noninvasive automatic detection of alzheimer's disease from spontaneous speech: a review. *Frontiers in Aging Neuroscience*, Volume 15 - 2023, 2023.

[3] Nicole D. Anderson. State of the science on mild cognitive impairment (mci). *CNS Spectrums*, 24(1):78–87, 2019.

[4] Marwan N. Sabbagh, M. Boada, S. Borson, M. Chilukuri, B. Dubois, J. Ingram, A. Iwata, A.P. Porsteinsson, K.L. Possin, G.D. Rabinovici, B. Vellas, S. Chao, A. Vergallo, and H. Hampel. Early detection of mild cognitive impairment (mci) in primary care. *The Journal of Prevention of Alzheimer's Disease*, 7(3):165–170, 2020.

[5] Xuchu Chen, Yu Pu, Jinpeng Li, and Wei-Qiang Zhang. Cross-lingual alzheimer's disease detection based on paralinguistic and pre-trained features. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–2, 2023.

[6] Jinpeng Li and Wei-Qiang Zhang. Whisper-based transfer learning for alzheimer disease classification: Leveraging speech segments with full transcripts as prompts. In *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 11211–11215, 2024.

[7] Ying Qin, Wei Liu, Zhiyuan Peng, Si-Ioi Ng, Jingyu Li, Haibo Hu, and Tan Lee. Exploiting pre-trained asr models for alzheimer's disease recognition through spontaneous speech, 2021.

[8] Jiamin Liu, Fan Fu, Liang Li, Junxiao Yu, Dacheng Zhong, Songsheng Zhu, Yuxuan Zhou, Bin Liu, and Jianqing Li. Efficient pause extraction and encode strategy for alzheimer's disease detection using only acoustic features from spontaneous speech. *Brain Sciences*, 13(3), 2023.

[9] Yilin Pan, Mingyu Lu, Yanpei Shi, and Haiyang Zhang. A path signature approach for speech-based dementia detection. *IEEE Signal Processing Letters*, 31:2880–2884, 2024.

[10] Felix Agbavor and Hualou Liang. Predicting dementia from spontaneous speech using large language models. *PLOS Digital Health*, 1(12):1–14, 12 2022.

[11] Prachee Priyadarshinee, Christopher Johann Clarke, Jan Melechovsky, Cindy Ming Ying Lin, Balamurali B. T., and Jer-Ming Chen. Alzheimer's dementia speech (audio vs. text): Multi-modal machine learning at high vs. low resolution. *Applied Sciences*, 13(7), 2023.

[12] Jiahong Yuan, Yuchen Bian, Xingyu Cai, Jiaji Huang, Zheng Ye, and Kenneth Church. Disfluencies and fine-tuning pre-trained language models for detection of alzheimer's disease. In *Interspeech 2020*, pages 2162–2166, 2020.

[13] Youxiang Zhu, Abdelrahman Obyat, Xiaohui Liang, John A Batsis, and Robert M Roth. Wavbert: Exploiting semantic and non-semantic speech using wav2vec and bert for dementia detection. In *Interspeech*, volume 2021, page 3790, 2021.

[14] R'mani Haulcy and James Glass. Classifying alzheimer's disease using audio and text-based representations of speech. *Frontiers in Psychology*, Volume 11 - 2020, 2021.

[15] Alireza Roshanzamir, Hamid Aghajan, and Mahdieh Soleymani Baghshah. Transformer-based deep neural network language models for Alzheimer's disease risk assessment from targeted speech. *BMC Medical Informatics and Decision Making*, 21(1):92, March 2021.

[16] Yu Pu and Wei-Qiang Zhang. Integrating pause information with word embeddings in language models for alzheimer's disease detection from spontaneous speech. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, page 1–5. IEEE, April 2025.

[17] Yuancheng Wang, Haoran Zheng, Qi Sun, Yong Ma, Shihu Zhu, Le Zhang, and Wei-Qiang Zhang. Cross-lingual alzheimer's disease detection based on scale criteria. In *2024 IEEE 14th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, pages 491–495, 2024.

[18] Saturnino Luz, Fasih Haider, Sofia de la Fuente, Davida Fromm, and Brian MacWhinney. Detecting cognitive decline using speech only: The adresso challenge, 2021.

[19] Patricia Pastoriza-Domínguez, Iván G. Torre, Faustino Diéguez-Vide, Isabel Gómez-Ruiz, Sandra Geladó, Joan Bello-López, Asunción Ávila Rivera, Jordi A. Matías-Guiu, Vanesa Pytel, and Antoni Hernández-Fernández. Speech pause distribution as an early marker for alzheimer's disease. *Speech Communication*, 136:107–117, 2022.

[20] Roelant Ossewaarde, Yolande Pijnenburg, Antoinette Keulen, Roel Jonkers, and Stefan Leijnen. Role of pause duration in primary progressive aphasia. *Aphasiology*, 39(5):601–619, 2025. PMID: 40303008.