

MULTIMODAL PROTOTYPICAL FEW-SHOT LEARNING FOR DEMENTIA DETECTION FROM SPEECH SIGNAL

Chunduri Yoshitha¹ · G. Vishwa Prakashini¹ · Penumarthi Hima Varshini¹ · Debanjali Bhattacharya²

¹Dept. of CSE, Amrita School of Computing | ²Dept. of AI, Amrita School of Artificial Intelligence | Amrita Vishwa Vidyapeetham, Bengaluru, India

aaic2026.1e80373

KEY TAKEAWAY A Conformer + BERT framework with cross-modal attention and prototypical few-shot learning detects dementia from a 2-minute speech sample with **85.2% accuracy and 91.2% AUC** — enabling affordable, non-invasive cognitive screening with no clinical facility required.

INTRODUCTION

- Dementia is a neurodegenerative disorder impairing memory, language, and cognition in millions worldwide; diagnosis is usually expensive and infrastructure-dependent.
- Speech production is directly governed by memory, language generation, and semantics — dementia produces measurable speech disorders: excessive pauses, poor fluency, word-finding trouble, repeated speech.
- Most prior systems rely on a single modality, limiting robustness to overlapping speech, noise, and natural variation.

Research Objective

- Design a multimodal dementia detection system from DementiaBank speech using Cookie Theft, Sentence Repetition, and Verbal Fluency tasks — combining acoustic, linguistic, and behavioral cues with few-shot learning for small-data robustness.

UN SDG Alignment

- SDG 3 (Good Health) · SDG 9 (Innovation) · SDG 10 (Reduced Inequalities) — affordable diagnosis without specialized clinical facilities.

METHODS

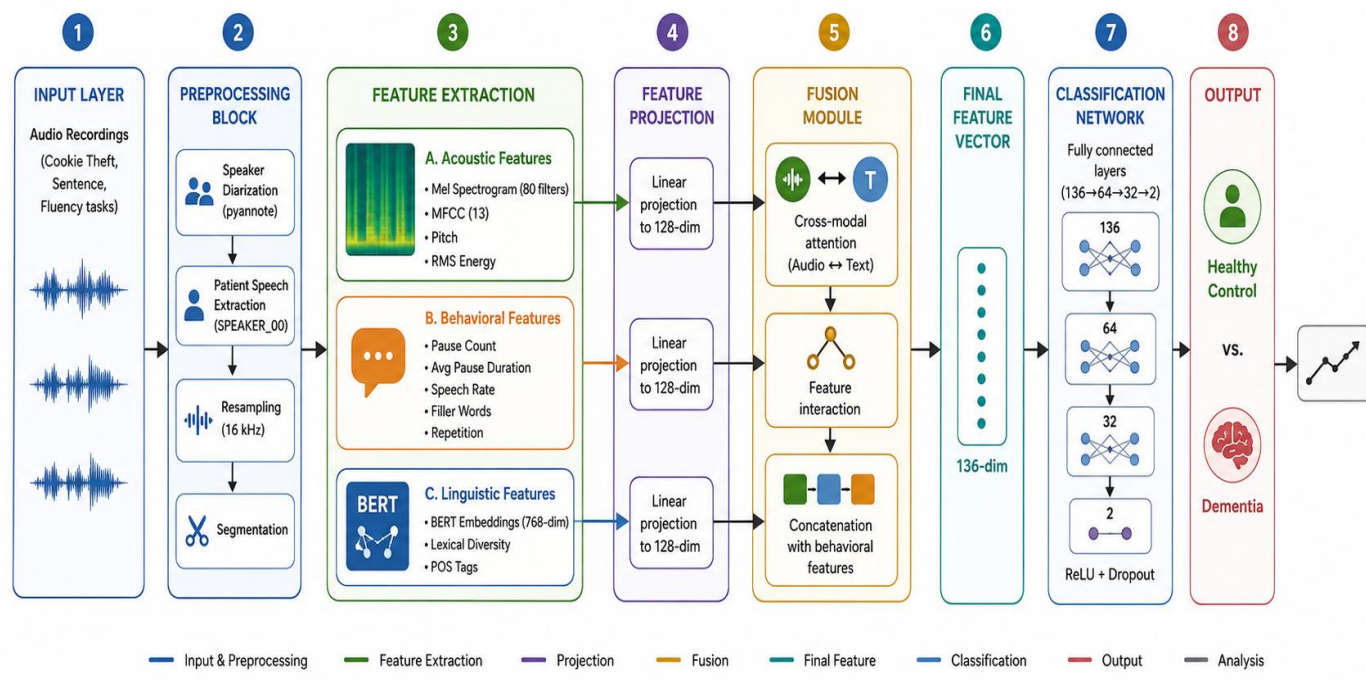


Fig. 1 — Overall architecture of the proposed multimodal dementia detection framework

Dataset: DementiaBank Pitt Corpus

- Cookie Theft (309 Dementia / 505 HC) · Sentence Repetition (236/1) · Verbal Fluency (235/2).

Preprocessing

- Speaker diarization (pyannoteAI) → patient-only extraction (SPEAKER_00) → .mp3 → .wav, 16 kHz → segmentation.

Feature Streams

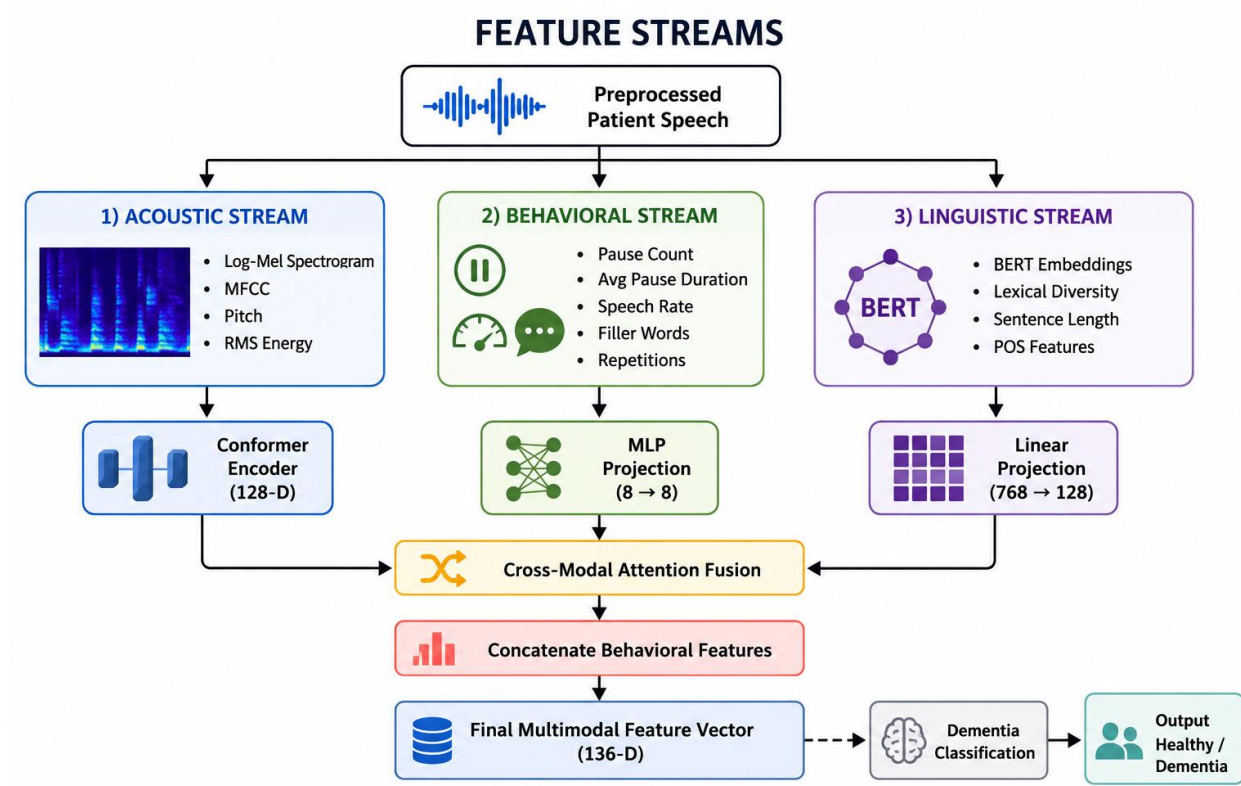


Fig. 2. Multimodal Feature Streams. Acoustic, behavioral, and linguistic speech features are extracted from preprocessed patient speech, projected into a shared latent space, and fused via cross-modal attention to generate a unified feature representation for dementia classification

Model & Loss

- Conformer (2 layers, 4 heads) on Mel-Spec → 128-dim. BERT → 128-dim. Bidirectional cross-modal attention → 128-dim + 8

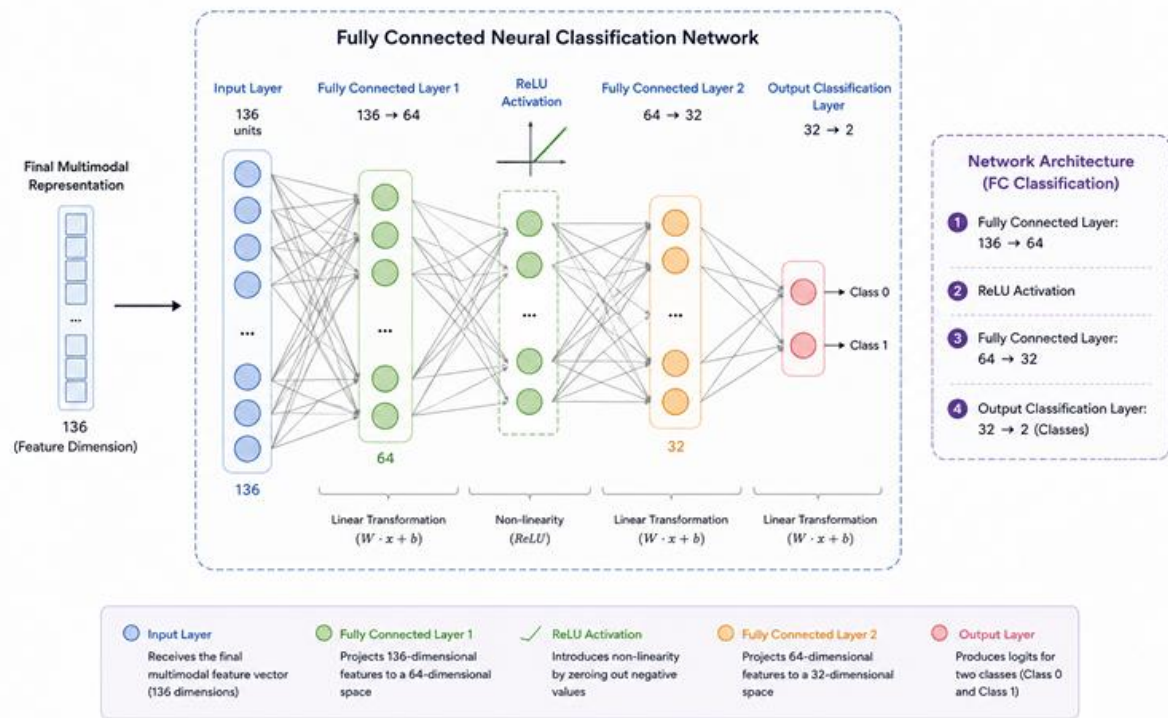


Fig. 2 Fully Connected Neural Classification Network

RESULTS

85.2%

Accuracy

90.2%

F1-Score

91.2%

AUC-ROC

Table 1 — Ablation Study

Model	Acc	F1	AUC
Full Model ★	85.2%	90.2%	91.2%
No Behavioral	77.5%	87.3%	78.8% ↓12.4
No Cross-Attn	77.5%	87.3%	85.1% ↓6.1
Text-Only	77.5%	87.3%	88.4%

Table 2 — Task-wise Performance

Speech Task	Acc	F1
Cookie Description	84.1%	90.5%
Sentence Repetition ★	84.7%	90.8%
Verbal Fluency	82.3%	89.2%

96.9% sensitivity

- 2,449/2,528 dementia cases correctly classified — only 79 false negatives. Ideal for screening.

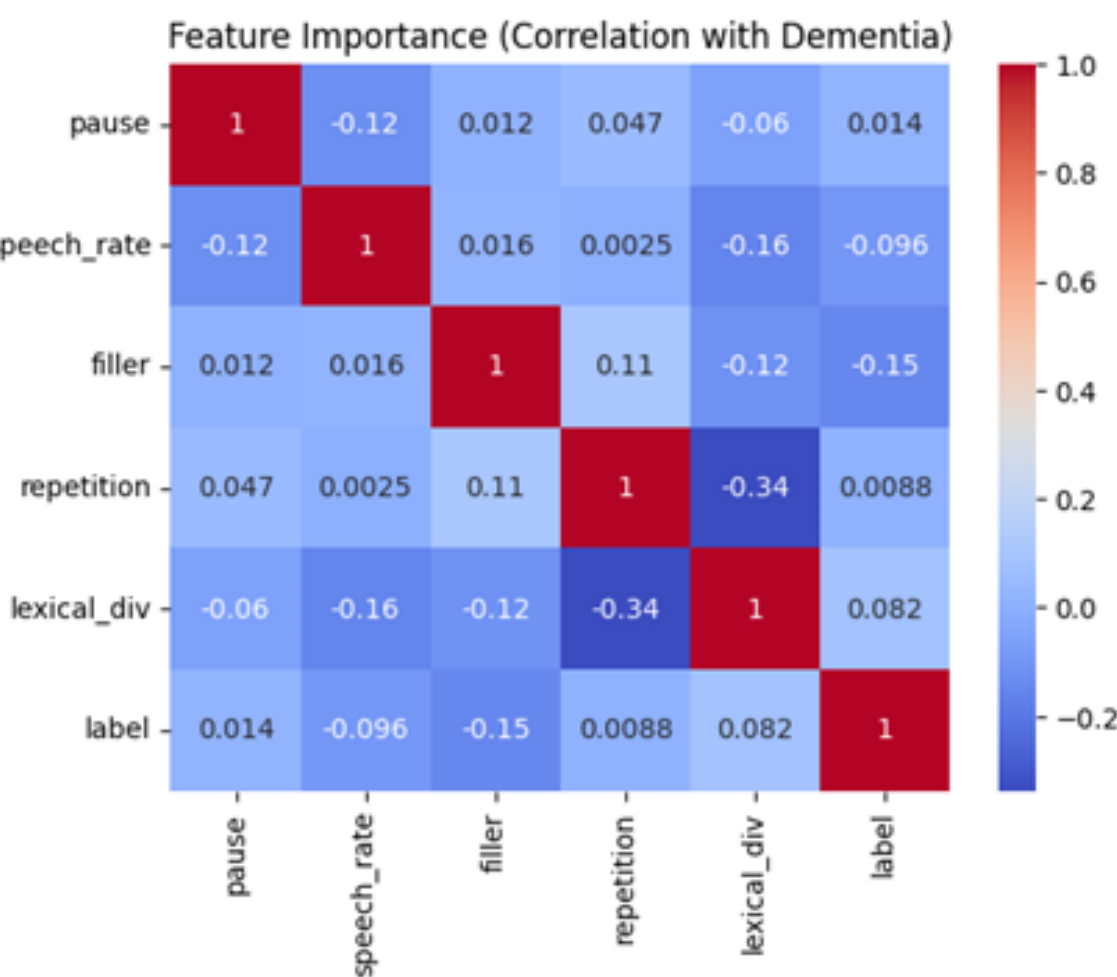


Fig. 4: Feature Importance (Correlation with dementia)

Key Findings

- No single feature strongly predicts dementia alone (all $|r| < 0.16$) — confirms multimodal fusion is essential.
- Behavioral features are the most discriminative modality (largest AUC drop when removed).
- Cross-modal attention significantly outperforms simple feature concatenation.

CONCLUSIONS

- Multimodal fusion of acoustic, linguistic, and behavioral speech features achieves clinically meaningful dementia detection on a small, imbalanced dataset.
- Behavioral cues (pauses, speech rate, fillers) are the single most valuable modality; cross-modal attention fusion significantly outperforms simple concatenation.
- Sentence Repetition gives the best task-level performance — it simultaneously stresses memory and language production.

Cognitive Drift Analysis (Novel)

- Post-classification module tracks temporal speech drift between chunk embeddings:
- $D_t = ||e_t - e_{t-1}|| / (||e_t|| + 10^{-6})$
- $C = \exp(-(\sigma d + \sigma p + \sigma r))$
- Dementia patients show rising drift ("cognitive fatigue") and breakdown points — adding clinical interpretability beyond a single label.

Limitations

- Severe class imbalance in 2 of 3 tasks; English-only; binary classification only (no subtypes).

Future Work

- Multilingual extension · Dementia subtype classification · MMSE severity regression · Longitudinal tracking · Real-time on-device deployment.

REFERENCES

- [1] Fraser et al. Linguistic features identify Alzheimer's disease in narrative speech. J. Alzheimer's Dis., 49(2), 407–422, 2016.
- [2] Jahan et al. Early dementia detection with speech analysis and ML. Discover Sustainability, 5(1), 65, 2024.
- [3] Becker et al. Natural history of Alzheimer's disease. Archives of Neurology, 51(6), 585–594, 1994.
- [4] Kumar et al. Dementia detection from speech using ML and DL. Sensors, 22(23), 9311, 2022.
- [5] Govindarajan et al. Dementia detection using Cogni-Wave. INDIACom, IEEE, 2025.
- [6] Ugwu & Oyeleke. Temporal-aware iterative speech model. arXiv:2510.00030, 2025.
- [7] Agbavor & Liang. Predicting dementia using LLMs. PLOS Digital Health, 1(12), 2022.
- [8] Sweidan et al. MMSE-calibrated few-shot prompting. arXiv:2509.19926, 2025.
- [9] Pahar et al. Multimodal feature fusion for cognitive decline. IEEE JBHI, 29(12), 2025.
- [10] Ehghaghi et al. Differentiating depression and dementia. W-NUT 2022.
- [11] Lopes da Cunha et al. Free speech analysis for AD & FTD. PLoS One, 19(6), 2024.
- [12] García et al. Speech markers of neurodegeneration: global equity. Brain, 146(12), 2023.
- [13] Kancharia et al. Corpus Callosum analysis for MCI. EMBC, IEEE, 2025.
- [14] Bhattacharya & Sinha. sMRI brain age in MCI via persistent homology. arXiv:2511.05520, 2025.
- [15] Martínez-Nicolás et al. Speech biomarkers for vascular dementia. Front. Neurosci., 16, 2022.
- [16] Walha et al. DL/ML for dementia detection in women. CMES, 142(3), 2959, 2025.
- [17] Mobatahej et al. Transformer-based DL for dementia: review. IEEE JBHI, 2025.
- [18] Volkmer et al. Communication needs in young-onset dementia. IJLCD, 59(4), 1553–1577, 2024.

CONTACT

G Vishwa Prakashini (Presenter)

bl.en.u4aie23007@bl.students.amrita.edu

Dr. Debanjali Bhattacharya (Corresponding)

b_debanjali@blr.amrita.edu