Early Detection of Dementia Using Multimodal Speech and Text Analysis

Research conducted by Royce Salah
Video Presentation Link: https://youtu.be/Rx-Usmjf6jE

INTRODUCTION

Dementia often goes undetected until late stages, limiting treatment options and quality of life [1,2]. Subtle linguistic and acoustic changes—such as reduced lexical diversity, increased pronoun use, disfluencies, and slower speech—are early cognitive markers but remain difficult to detect clinically [4,5,6,9]. This research asks: Can multimodal speech and text analysis, grounded in cognitive science, identify early signs of dementia up to 15 years before diagnosis?

Insights from semantic memory degradation, lexical retrieval failure, and discourse coherence theory [3,4,6] are combined with machine learning. Using the DementiaNet corpus of real-world interviews, this study extracts features such as lexical entropy, semantic drift [6], and acoustic prosody [10,11], aiming to develop lightweight, interpretable models capable of scalable early screening.

DESIGN

The **DementiaNet** dataset includes interviews from 84 individuals later diagnosed with dementia and 62 healthy controls, recorded 5–15 years pre-diagnosis.

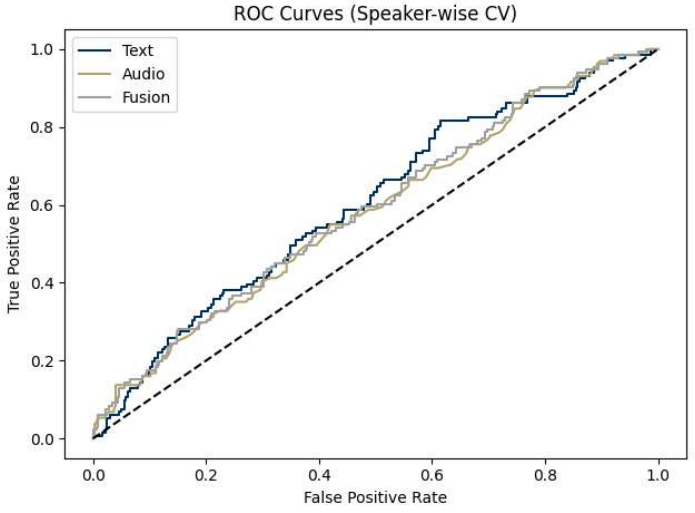
A **multimodal pipeline** extracted:

- **Lexical features**: semantic drift (MiniLM), lexical entropy (GPT-2), type-token ratio, pronoun ratio, sentence length
- **Acoustic features**: prosodic markers (pitch, jitter, pauses, speech rate) via openSMILE/eGeMAPSv02
- **Transcription**: Whisper ASR preserved natural speech variability

Separate models were trained per modality (logistic regression for lexical, random forest for acoustic), then combined via a meta-classifier. Evaluation used 10-fold speaker-wise Group-K Fold cross-validation to prevent data leakage.

The hypothesis posited that subtle speech and language changes serve as early markers of cognitive decline, with lexical and acoustic features predicting dementia/control status.

RESULTS



Model Performance
Fusion Model AUC: 0.605 ± 0.077
Text-Only AUC: 0.602 ± 0.106
Audio-Only AUC: 0.593 ± 0.073

Both the text and audio models outperformed the baseline, with the text-based model demonstrating slightly higher predictive power but greater variance. The audio model showed lower overall accuracy but more stable performance across folds. While the fusion model did not significantly exceed the performance of the individual classifiers, it offered a modest gain in precision by integrating complementary signals from both modalities.

CONCLUSION

All models surpassed the chance baseline (AUC = 0.5), confirming modest but consistent predictive capability.

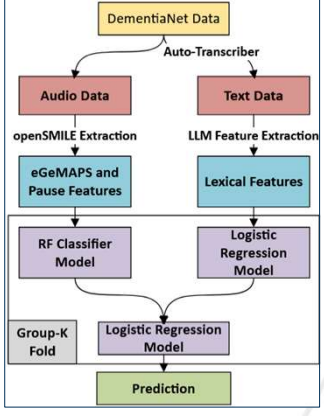
Feature Contributions

The five most predictive features, MATTR, pronoun ratio, semantic drift, type-token ratio, and lexical entropy, were all **lexical**. This reinforces prior cognitive research suggesting linguistic changes precede more overt behavioral symptoms in dementia.

Interpretation

Despite using lightweight models and a limited dataset, the pipeline reliably captured early indicators of cognitive decline. The lexical features show promise for integration into consumer-facing applications like smartphone-based screening tools.

Pipeline Visual



References

[1] Amjad et al., 2018 – Underdiagnosis
[2] Bradford et al., 2009 – Missed diagnosis
[3] Young et al., 2024 – Tau burden and speech
[4] "Language Markers of Dementia," Front. Aging Neurosci.
[5] Rentoumi et al., 2017 – Linguistic indicators
[6] Mendez et al., 2006 – Semantic drift, disorganized thought
[9] Bittner et al., 2022 – Pronoun use in preclinical AD
[10] Eyben et al., 2016 – GeMAPS prosodic features
[11] Ding et al., 2023 – Acoustic-brain volume link