EARLY DEMENTIA DETECTION USING MULTI-LINGUAL VOICE ANALYSIS WITH NLP TECHNIQUES

Madhavi RP
B.M.S College of Engineering
Bengaluru, India

Shantanu Shrivastav
B.M.S College of Engineering
Bengaluru, India

Sathuri Harshvardhan
B.M.S College of Engineering
Bengaluru, India

Jatin Bhatnagar
B.M.S. College of Engineering
Bengaluru, India

Shreya Mitawa
B.M.S. College of Engineering
Bengaluru, India

***Abstract*— Dementia remains one of the most critical global health challenges, particularly for underserved communities and aging populations. Current diagnostic tools are often constrained by language barriers, limited accessibility, and a lack of interpretability, even though early detection is vital for timely intervention.**

This work addresses these gaps by proposing a conceptual AI-driven framework that leverages multilingual voice analysis to identify early symptoms of dementia. The framework integrates machine learning (ML) and natural language processing (NLP) techniques to capture subtle variations in both acoustic features and linguistic markers.

The key contributions of this project are:

1. **Multilingual support in Hindi, Spanish, and English**
2. **Fusion of acoustic and linguistic biomarkers**
3. **Integration of explainable AI to enhance clinical trust**

The system is designed to operate on low-latency mobile devices, making it suitable for resource-limited and rural environments. Beyond therapeutic benefits, this framework can also help reduce healthcare disparities by providing affordable and accessible screening in regions lacking conventional diagnostic facilities.

Index Terms: Healthcare AI, explainable AI, voice biomarkers, natural language processing, dementia detection, multilingual analysis

INTRODUCTION

Dementia is a progressive neurological disorder that deteriorates over time, impairing memory, communication, and decision-making abilities. According to the World Health Organization, more than 55 million people globally live with dementia, with nearly 10 million new cases diagnosed annually [1]. The socioeconomic impact is substantial, with global costs exceeding $1 trillion per year. In India, the number of dementia patients is projected to rise sharply due to an aging population, placing significant pressure on the healthcare system.

Traditional diagnostic methods, such as the Mini-Mental State Examination (MMSE) and clinical brain imaging, are accurate but often inaccessible to patients in rural areas due to high costs, infrastructure requirements, and language or literacy constraints. Speech-based analysis provides a promising, non-invasive alternative, as linguistic and acoustic changes often appear years before formal diagnosis [2]. With the widespread use of smartphones and telemedicine platforms, voice-based diagnostics can be deployed at scale, offering a highly practical solution.

A. Motivation

The motivation for this work is based on three key observations:

- **Accessibility:** Voice samples can be collected using everyday devices such as smartphones and laptops, minimizing reliance on costly medical equipment.
- **Early Biomarkers:** Dementia patients often demonstrate reduced vocabulary, frequent pauses, and irregular intonation, which can be quantified through signal processing and NLP.
- **Trust and Explainability:** Clinicians are hesitant to adopt black-box AI models. Transparent, interpretable decision support is essential for clinical acceptance.

I. PROBLEM STATEMENT AND OBJECTIVES

Current methods for detecting dementia have drawbacks such as:

Language bias: English-only datasets are used to train the majority of models.

Disjointed characteristics: Research frequently examines either linguistic or acoustic characteristics, but not both at once.

Clinical adoption barriers: Many models are inappropriate for use in healthcare because they do not offer explanations.

The following are the project's goals: 1)To create a multilingual framework that can analyse speech samples in Hindi, English, and Spanish. 2)To combine linguistic biomarkers (such as lexical richness and syntactic depth) with acoustic biomarkers (such as MFCCs, jitter, and shimmer).

3) To use interpretability-enhancing explainable AI methods (SHAP, LIME).
4)To adapt the model to low-resource, real-time settings, like rural clinics.

II. SYSTEM DESIGN

A. Pipeline Overview

There are five main steps in the system pipeline:

1)Voice Input: Mobile devices or telemedicine platforms are used to collect samples of patients' speech. To get people to talk, they are asked to do things like describe pictures or read short passages.
2)Preprocessing: Noise from the environment is removed, audio signals are made equal, and speech is broken up into chunks that can be analysed.
3)Feature Extraction: We get acoustic and linguistic features. For instance, MFCCs show spectral properties, and the Type-Token Ratio shows how rich the vocabulary is.
4)Multi-Model Fusion: At the decision level, the outputs of CNN-based acoustic models and transformer-based NLP models are combined.
5) Explainable Output: The final prediction comes with feature importance values that help doctors understand how the model came to that conclusion.

B. Feature Engineering

Acoustic Features: Acoustic features show how something sounds when it is spoken. Prosodic: the pitch, loudness, and speed of speech. Spectral: MFCCs are used a lot in systems that recognize speech.
Voice Quality: Jitter and shimmer are examples of measures that show how unstable vocal fold vibrations are.
Linguistic features show what is said: Lexical: Type-Token Ratio and Brunet's Index for how words are used.
Syntactic: Average depth of the parse tree, number of subordinate clauses.
Semantic: how well sentences fit together and how similar word embeddings are.

III. DATASETS

There will be three datasets:

- **Dementia Bank**: Has audio samples and transcripts of English-speaking patients doing normal things.
- **ADReSS**: A dataset that is balanced and has standardized evaluation metrics for finding dementia.
- **Hindi-AD:** A dataset gathered locally, comprising 50 patients and 50 controls, concentrating on Hindi-speaking populations.

IV. TECHNICAL STACK

Table I summarizes the planned implementation stack.

Table 1: Proposed Technical Stack

Component	Technologies
Audio Processing	Librosa, PyAudioAnalysis
NLP Pipeline	spaCy, HuggingFace Transformers
ML Framework	PyTorch Lightning
Explainability	SHAP, LIME, Captum
Deployment	FastAPI, ONNX Runtime

V. APPLICATIONS AND IMPACT

Potential applications of this project include:

- **Clinical Screening:** Assisting neurologists in prioritizing patients for further testing.
- **Remote Healthcare:** Voice-based pre-screening for patients in rural areas via telemedicine.
- **Personalized Monitoring:** Longitudinal tracking of patients' cognitive health.
- **Policy Support:** Informing public health strategies and reducing healthcare inequalities.

VI. FUTURE WORKS

Future directions include:

- Training and validating models on multi-lingual datasets.
- Expanding to more Indian languages such as Kannada and Tamil.
- Integrating the framework into telehealth platforms.
- Conducting large-scale longitudinal studies to validate clinical utility.

VII. ACKNOWLEDGMENT

We thank NVIDIA for GPU donations and participating clinicians for domain expertise.

REFERENCES

[1] World Health Organization, "Dementia: A Public Health Priority," WHO Report, 2022.
[2] S. Luz, F. Haider, D. de la Fuente Garcia, "Detecting cognitive decline using speech: The ADReSS Challenge," INTERSPEECH, 2021.
[3] K. Fraser, J. Meltzer, K. Rudzicz, "Linguistic Features Identify Alzheimer's Disease in Narrative Speech," Journal of Alzheimer's Disease, vol. 49, no. 2, 2016.
[4] A. Balagopalan et al., "Comparing Acoustic and Linguistic Features for Alzheimer's Dementia Recognition," INTERSPEECH, 2020.
[5] D. Martinez, "Fusion of Acoustic and Textual Feature for Dementia Detection," ICASSP, 2021.