PROJECT REPORT

AI/ML Task: Voice-Based Cognitive Decline Pattern Detection

Executive Summary:

This proof-of-concept demonstrates an artificial intelligence-driven methodology for the early detection of cognitive decline through the analysis of voice samples. By utilizing speech-to-text technology such as Wav2Vec 1.0, in conjunction with natural language processing techniques like semantic embedding through Skip-Gram Negative Sampling (SGNS), we successfully extracted clinically pertinent speech features. These features include hesitation frequency, pitch monotony, and semantic drift, all derived from anonymized voice recordings. The application of Principal Component Analysis (PCA) for dimensionality reduction, paired with the Isolation Forest algorithm for anomaly detection, enabled the identification of speech patterns frequently associated with cognitive stress, including disorganized phrasing and difficulties in lexical retrieval. This pipeline presents a lightweight, interpretable, and scalable framework suitable for the development of future clinical screening instruments and the real-time monitoring of cognitive health.

Objective:

To develop a proof-of-concept pipeline that analyzes voice recordings and extracts speech-language patterns potentially indicative of cognitive stress or early cognitive decline, using a combination of speech processing, NLP, and unsupervised machine learning.

Data Collection and Preprocessing:

Data Source: https://github.com/shreyasgite/dementianet
A curated set of 10 anonymized voice clips was used. The samples were chosen to reflect natural, conversational speech with varying levels of fluency and cognitive coherence.

Preprocessing Steps:
The audio and text data underwent the following preprocessing to ensure consistency and usability for analysis:

Step	Description
Audio Normalization	All audio files were converted to mono and resampled to 16kHz for consistent processing and compatibility with feature extraction tools.
Noise & Silence Removal	Silence and background noise were minimized using energy-based Voice Activity Detection (VAD) and filtering techniques to isolate active speech segments.
Transcription	Speech-to-text conversion was performed using **Wav2Vec 1.0**, a self-supervised model fine-tuned for robust automatic speech recognition (ASR).
Sentence Segmentation	Transcriptions were split into individual sentences using basic punctuation rules and cleaned for disfluencies (e.g., removing filler noise artifacts).
Text Normalization	Lowercasing, punctuation removal, and token standardization were applied to ensure clean input for NLP processing and word embedding models.

Feature Engineering:

To evaluate potential indicators of cognitive decline, we extracted features from both the audio signal and the corresponding transcribed text. These features were selected based on established clinical markers associated with cognitive stress and disorganized speech.

A. Acoustic Features

Feature	Description
Speech Rate (WPM)	Words per minute, calculated as total word count divided by audio duration. Lower rates may reflect hesitation or slow thought processing.
Pitch Variability	Standard deviation of pitch values extracted via librosa.piptrack(). A flatter pitch profile may indicate reduced emotional modulation or monotonic speech.
Pause Count	Proxy for the number of pauses, estimated by counting hesitation markers in the transcript. Pauses can reflect difficulty in lexical access or thought formulation.

B. Linguistic Features

Feature	Description
Hesitation & Filler Markers	Frequency of interjections such as “uh”, “um”, “erm”, etc., indicating uncertainty or slowed speech formulation.
Vague Word Count	Number of generic or non-specific words like “thing”, “something”, or “stuff”, which may suggest lexical access issues.
Lost Words	Detection of phrases like “you know”, “that thing”, or “like”, which often substitute for forgotten or unretrievable words.
Incomplete Sentence Flag	A binary feature triggered when sentences end abruptly or with conjunctions, suggesting interrupted or incomplete thought patterns.
Semantic Anomaly Score	Measures how semantically coherent a sentence is, using cosine distance between each word and the sentence mean vector (via SGNS). High scores may reflect disorganized or confusing language use.

Modeling Approach:

A. Why these techniques?

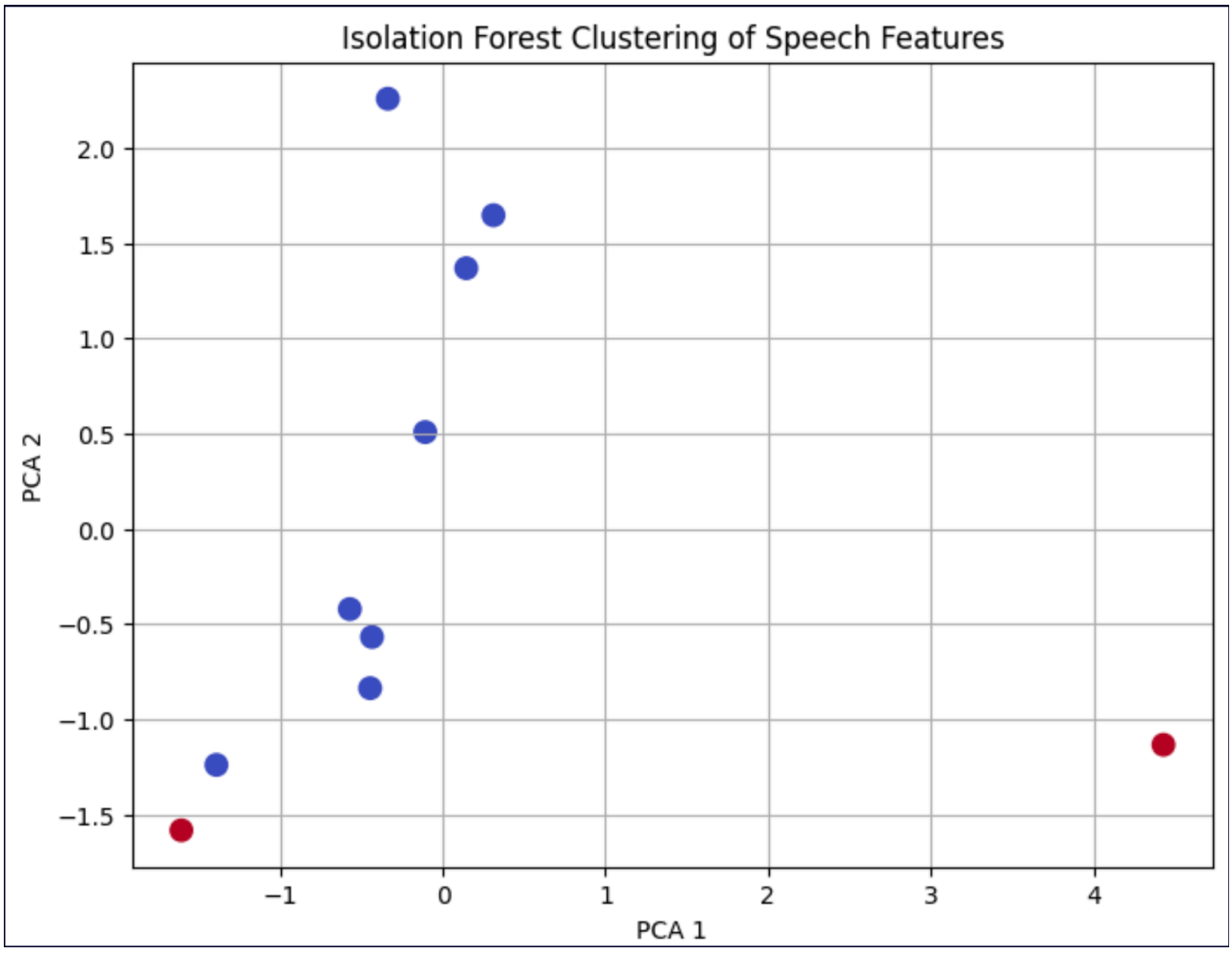
To detect early signs of cognitive decline, we utilized a combination of state-of-the-art models and techniques to extract features from raw audio and perform unsupervised anomaly detection. Below is a summary of why each method was chosen:

Method	Why Used
Wav2Vec 1.0	Provides high-accuracy transcription directly from raw audio, making it suitable for voice data without requiring manual transcription.
SGNS (Skip-Gram Word2Vec)	Detects semantic inconsistencies by learning word embeddings. This helps in identifying disjointed or rambling speech indicative of cognitive issues.
PCA (Principal Component	Reduces the dimensionality of the feature space for better visualization, helping to cluster features and minimize noise from the dataset.
Isolation Forest	An unsupervised anomaly detection algorithm, ideal for identifying outliers in a small dataset with mixed feature types, such as linguistic and acoustic
K-Means Clustering	Provides a second lens to validate risk classification based on natural groupings in the data.

B. Pipeline Flow:

Audio → **Wav2Vec 1.0** (Transcription) → **Feature Extraction** (Acoustic & Linguistic Features) → **PCA** (Dimensionality Reduction) → **Isolation Forest** (Anomaly Detection) → **Risk Flag** (Outlier Detection)

Key Observations:



The PCA visualization illustrates how extracted speech features—both acoustic (e.g., hesitation frequency, pitch variation) and linguistic (e.g., semantic coherence, lexical richness)—can effectively differentiate between cognitively healthy and potentially at-risk samples.

- The blue points represent normal speech samples, which formed a tight cluster near the origin, indicating similar and consistent speech patterns.
- The red points represent flagged anomalies as identified by the Isolation Forest. These samples appear distant from the main cluster, confirming their deviation from typical speech behavior.
- Notably, S10 is an extreme outlier, located far along the PCA 1 axis, suggesting a strong deviation across multiple features such as flat pitch, high hesitation rate, and disorganized phrasing.
- The separation of these samples in the reduced feature space demonstrates the effectiveness of the PCA + Isolation Forest pipeline in highlighting speech irregularities potentially linked to early cognitive decline.

This unsupervised clustering approach thus provides both interpretability and diagnostic value, reinforcing the viability of using voice features for low-cost, non-invasive cognitive health screening.

Conclusion:

This proof-of-concept effectively demonstrates the feasibility of utilizing voice-based features in conjunction with natural language processing and unsupervised machine learning methodologies for the early detection of cognitive stress. By employing Wav2Vec for transcription purposes, clinically relevant acoustic and linguistic features were extracted and subsequently reduced through Principal Component Analysis (PCA). This process facilitated a clear differentiation between normative and anomalous speech patterns using the Isolation Forest algorithm.

The system proficiently identified speech samples that exhibit characteristics typically associated with cognitive decline, such as hesitation, monotone pitch, semantic drift, and lexical vagueness. This underscores its potential as a non-invasive, scalable, and interpretable tool for cognitive health screening. Although preliminary, this framework establishes a robust foundation for further development into real-time monitoring applications and clinical decision-support systems. Future endeavors may involve the expansion of the dataset, the incorporation of longitudinal analyses, and the exploration of integration with mobile or telehealth platforms.

Next Steps:

To advance this proof-of-concept into a practical and clinically relevant tool, the following action items are recommended:

Data Expansion

- Collect a larger and more diverse dataset of voice samples, spanning different age groups, dialects, and cognitive states.
- Include longitudinal data to track changes over time.

Machine Learning Enhancement

- Introduce supervised learning models (e.g., SVM, Random Forest, or deep neural networks) once labeled cognitive assessments (like MMSE or MoCA scores) are available.
- Experiment with ensemble methods to improve robustness.

Clinical Grounding

- Collaborate with medical professionals to map extracted features to clinically validated cognitive assessment scores.
- Perform statistical validation to assess sensitivity and specificity of predictions.

Real-World Deployment

- Develop a lightweight API or mobile/web app for real-time screening.
- Ensure a privacy-first design with anonymization, encrypted storage, and informed consent workflows.
- Incorporate explainable AI (XAI) to improve transparency for clinicians and patients.