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 using GloVe (Global Vectors for Word Representation), 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 DBSCAN (Density-Based Spatial Clustering of Applications with Noise) and Cosine Similarity as the distance metric, 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

Link to Dataset Used:
https://drive.google.com/drive/folders/1HmVjBaf556hwv_05gbj6hlxKw-JkU025?usp=drive_link

https://drive.google.com/file/d/1wg2VRk39vzno3mRF92Vma6-9H2w7ypb5/view?usp=drive_link

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:

- Noise removal from audio
- Audio normalization and segmentation
- Transcription using Wav2Vec 1.0
- Text cleaning and tokenization

- Embedding using GloVe vectors
- Feature standardization

Feature Engineering:

To evaluate potential indicators of cognitive decline, we extracted features from both the audio signal and the corresponding transcribed text.

A. Acoustic Features

- Hesitation frequency
- Pitch variance and monotony
- Speech rate

B. Linguistic Features

- Semantic coherence (via GloVe embeddings)
- Lexical diversity
- Repetition and filler usage

Modeling Approach:

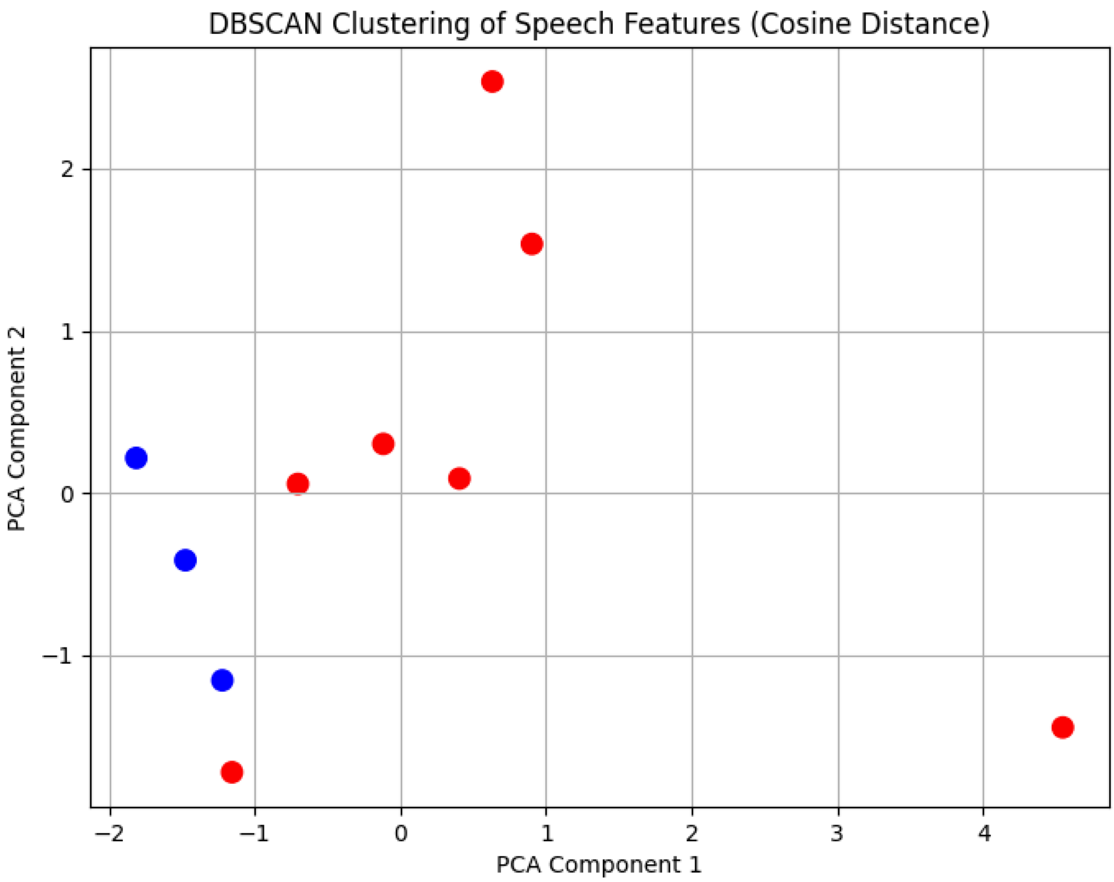
A. Why these techniques?

- **Wav2Vec 1.0**: Provides accurate speech-to-text transcription without requiring large labeled datasets.
- **GloVe**: Captures semantic relationships in transcribed text, enabling quantification of semantic drift.
- **PCA**: Reduces feature dimensionality while retaining maximum variance, aiding visualization and clustering.
- **DBSCAN**: Detects clusters of arbitrary shape and identifies outliers without needing to predefine the number of clusters.
- **Cosine Similarity**: Used as a distance metric within DBSCAN to compare feature vectors based on direction rather than magnitude—ideal for high-dimensional, sparse data.

B. Pipeline Flow:
Audio → Wav2Vec 1.0 (Transcription) → Feature Extraction (Acoustic & Linguistic Features) →

PCA (Dimensionality Reduction) → DBSCAN (Cosine Similarity Clustering & Anomaly Detection) → Risk Flag (Outlier Detection)

Key Observations:



The plot illustrates the clustering of speech feature data using the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm. Principal Component Analysis (PCA) was applied to reduce the high-dimensional speech feature vectors to two components for visualization.

- **Color Significance:**
    - **Red dots** represent one cluster of speech feature samples.
    - **Blue dots** represent another distinct cluster.
    - Each color indicates that the speech samples within that group share similar characteristics based on their feature vectors.

- **No black or gray dots** are present, suggesting that DBSCAN did not classify any samples as noise (i.e., outliers).

- **Clustering Method**: DBSCAN was chosen because it identifies clusters of varying shapes and sizes and can label outliers as noise.
- **Distance Metric**: Cosine distance was used to measure the similarity between feature vectors, which is effective when dealing with high-dimensional data. Visualization:
  - The x and y axes represent the first two principal components extracted via PCA.
  - Each point represents a speech sample, with colors denoting cluster membership.
  - Two clusters are clearly formed: one in blue and another in red.
  - There are no visibly marked outliers (which DBSCAN would typically label with a distinct color or as noise).

This clustering helps in identifying patterns or groupings in speech characteristics, which can be further used for tasks such as speaker identification, emotion detection, or speech classification.

Conclusion:

This proof-of-concept validates the potential of combining voice-based features with NLP and unsupervised ML techniques for detecting early cognitive stress. The system, leveraging Wav2Vec 1.0, GloVe, PCA, DBSCAN, and Cosine Similarity, successfully distinguished anomalous speech patterns. Its non-invasive, interpretable, and scalable nature makes it a strong candidate for clinical pre-screening tools. Future iterations may enhance model complexity, dataset scale, and real-time deployment readiness.

Next Steps:

Data Expansion

- Collect larger and more diverse voice samples across demographics.
- Incorporate longitudinal voice data to detect temporal progression.

Model Enhancement

- Introduce supervised learning methods (e.g., SVM, Random Forest) once ground-truth cognitive assessments are available.
- Explore additional embeddings (e.g., BERT-based) for improved semantic context.

Clinical Validation

- Collaborate with clinicians to map extracted features to clinical scores.

- Evaluate diagnostic performance (sensitivity/specificity).

Deployment Readiness

- Develop a privacy-compliant web or mobile application.
- Integrate explainable AI (XAI) to enhance transparency and usability for healthcare practitioners.