ML MODEL FOR EARLY COGNITIVE IMPAIRMENT DETECTION

-Aarushi Anand

This analysis primarily uses person-specific rule-based thresholds to detect cognitive impairment risk. Below are the key findings and methodological insights:

Most Insightful Features

The notebook extracts several speech-based features that are relevant for identifying potential cognitive impairment indicators. The most insightful among these include:

Pauses per minute: Long pauses (>1.2s) were counted and normalized over time, as increased frequency of such pauses is often linked to cognitive difficulties.

Hesitation rate: Hesitative utterances like "uh", "um", "hmm" were tracked as proxies for verbal disfluency.

Repetition rate: Both individual word and short phrase repetitions were flagged. Higher repetition frequency may reflect memory or speech formulation issues.

Speech rate (WPM): Deviations from a typical word-per-minute range may suggest cognitive or motor issues.

Pitch variability (standard deviation): Monotonic speech (low pitch variance) may be indicative of affective or neurological disorders.

Incomplete sentence rate: Reflects syntactic coherence and potential difficulties in expression.

These features are directly interpretable and align well with known clinical markers in neuropsychological assessments.

ML Methods Used and Rationale

Rule-Based Risk Classification:

The implemented approach uses an if-else scoring system to classify individuals into Low, Moderate, or High Risk. This rule-based system evaluates the extracted features against predefined thresholds:

- Score increases by 1 if pause rate > 5
- Score increases by 1 if repetition rate > 5
- Score increases by 1 if WPM < 100 or > 190
- Score increases by 1 if pitch standard deviation < 30

- Score increases by 1 if incomplete sentence rate > 50

Risk levels are determined as:

- Low Risk: score ≤ 1
- Moderate Risk: score 2-3
- High Risk: score > 3

This provides transparency and interpretability, which are beneficial in a clinical context.

Experimentation with SVM:

Although Support Vector Machine (SVM) classification was initially explored, it was ultimately removed from the implementation as it did not provide sufficiently personalized results for individual assessment. The rule-based approach was maintained as it allows for tailored person-specific thresholds and offers clearer interpretability in a clinical context.

Speech Processing Implementation:

The system uses AssemblyAI for speech transcription, leveraging its API for accurate word-level transcription with timestamps. The implementation:

- Tracks pauses between words (flagging those > 1.2s)
- Identifies repetitions of individual words and phrases (2-3 word sequences)
- Calculates speech rate in words per minute
- Uses librosa's piptrack for pitch analysis and standard deviation calculation

Sample visualizations of feature trends

Box Plot has been used to visualize the distribution of a key feature (e.g., mfcc_0, speech_rate) for each class (Normal vs. Risk)

Correlation Heatmap: Shows how features relate to each other and to the label.

Implementation Details

- **API Implementation:** The solution is packaged as a FastAPI application that accepts audio file uploads and returns cognitive risk assessment results
- **Audio Processing:** Uses soundfile and librosa for audio feature extraction, particularly for pitch variability analysis
- **Phrase Tracking:** Implements a sliding window approach to detect repetitions of 2-3 word phrases
- **Sentence Completion Analysis:** Examines sentence endings to identify incomplete thoughts

Speech Recognition Exploration

Several models were tried like Nemo (by Nvidia), Whisper, SpeechMatic, Krisp etc., but due to inefficiencies in the transcribed text they couldn't be used. The approach was finally boiled down to using AssemblyAI which gave the best results.

Other tools and models can be researched upon and fine-tuned for the same cause and tried in future.

Potential Next Steps Toward Clinical Robustness

a. Data Scaling

Increase dataset size: More speakers, age diversity, and validated clinical labels (diagnosed vs. control) are essential.

Balance demographics: Gender, accent, age, and language proficiency need to be controlled or accounted for.

b. Model Enhancement

Feature engineering: Include linguistic complexity (e.g., syntactic depth, lexical diversity) and acoustic biomarkers (e.g., jitter, shimmer).

Multimodal modeling: Combine audio features with transcript-based NLP features for a richer profile.

Deep learning models: Test CNNs/RNNs on spectrograms or use transformer-based embeddings (e.g., wav2vec) for improved representation.

c. Validation Strategy

Cross-validation: Employ stratified k-fold validation to ensure model generalizability.

Clinical benchmarking: Compare against neuropsychological assessments (e.g., MMSE, MoCA).

Human-in-the-loop testing: Involve clinicians for interpretability validation and iterative improvement.

d. Deployment Considerations

Privacy and security: Ensure compliance with HIPAA/GDPR for audio data.

Real-time processing: Optimize for latency and streaming inputs.

Explainability: Use SHAP or LIME to explain model predictions to clinicians.