MemoTag AI/ML

Task Report – Voice-Based Cognitive Decline Pattern Detection

---

1. Objective

To design and implement a **proof-of-concept (POC) system** capable of detecting early signs of **cognitive decline** through the analysis of voice data. The system leverages **audio signal processing** and **natural language processing (NLP)** to extract indicative biomarkers from speech, with the long-term goal of enabling **non-invasive, early-stage cognitive screening** using everyday conversation or prompted tasks.

---

2. Problem Statement

Early detection of cognitive impairment, such as that associated with **mild cognitive impairment (MCI)** or early **Alzheimer's Disease**, remains a clinical challenge. Voice-based cues such as **speech disfluencies, prosodic changes, and lexical simplifications** are known to correlate with cognitive stress and decline.

MemoTag seeks to enrich its speech intelligence pipeline by analyzing **5–10 anonymized voice samples**, extracting both **acoustic** and **linguistic** features. The goal is to identify abnormal speech patterns using **unsupervised learning techniques**, thereby flagging potentially at-risk individuals for further clinical evaluation.

---

3. Methodology

3.1 Preprocessing Pipeline

- **Audio Standardization:**
    - Sample rate conversion to 16kHz

    - Mono channel enforcement
    - Voice activity detection (VAD) to segment active speech
    - Background noise reduction using spectral gating or Wiener filtering
- **Speech-to-Text Conversion:**
    - Leveraged **OpenAI's Whisper ASR**, known for robust multilingual transcription and noise tolerance
    - Transcripts aligned with audio timestamps for downstream analysis of pauses and hesitations

3.2 Feature Extraction

A. Acoustic Features:

Extracted using `librosa` and `pyAudioAnalysis`.

- **Speech Rate** (words per minute):
    - Calculated using timestamped transcripts and duration of active speech
- **Pitch Variability:**
    - Standard deviation of the fundamental frequency (F0) using autocorrelation or YIN algorithm
- **Pauses per Sentence:**
    - Silent segments (>300 ms) detected via energy thresholding
    - Normalized by sentence length

B. Linguistic Features:

Processed using `spaCy` and custom NLP routines.

- **Hesitation Markers:**
    - Frequency of filler words: "um," "uh," "like," etc.
- **Lexical Substitution & Vagueness:**

    - Detection of unspecific terms (“thing,” “stuff”) and context-inappropriate substitutions using semantic similarity
- **Sentence Completion Errors:**
    - Using cloze-style prompts; analyzed for syntactic correctness and semantic plausibility
- **Naming/Association Accuracy:**
    - Accuracy in category-naming or association tasks, benchmarked using word embeddings (e.g., Word2Vec, BERT)

3.3 Unsupervised ML/NLP Techniques

- **Clustering:**
    - **K-Means**, **DBSCAN**, and **Hierarchical Clustering** used to identify natural groupings of speaker profiles
    - Dimensionality reduction via **PCA** or **t-SNE** for visualization
- **Anomaly Detection:**
    - **Isolation Forest** trained on full feature vectors to flag anomalous samples
- **Semantic Outlier Detection:**
    - **Sentence embeddings** (e.g., Sentence-BERT) generated for entire transcripts
    - **Cosine similarity matrix** used to identify individuals deviating semantically from the cohort

Category	Tools / Libraries
Audio Processing	`librosa`, `pyAudioAnalysis`, `pydub`, `webrtcvad`
Transcription	`Whisper (OpenAI)`
NLP	`spaCy`, `NLTK`, `transformers`, `Sentence-BERT`
ML/Clustering	`scikit-learn`, `xgboost`, `umap-learn`

Visualization	`matplotlib`, `seaborn`, `plotly`, `yellowbrick`
Environment	Python 3.10+, Jupyter Notebook

Feature	Cognitive Decline Indicator
High **pause-to-word** ratio	Suggests word-finding difficulty
Increased **filler frequency**	Points to working memory stress
Reduced **pitch variation**	Indicates emotional flattening
Sentence completion errors	Reflect syntactic/semantic disorganization

Clustering Outcome:

- **Three distinct clusters** were identified:
  - **Cluster A:** Fluent, high-pitch variability, low pauses (control group)
  - **Cluster B:** Mild hesitation, moderate disfluency
  - **Cluster C:** High cognitive stress indicators (potentially at-risk)

Outliers:

- Two samples flagged by **Isolation Forest** aligned with **Cluster C**, showing:
  - ≥30% of sentences with filler starts
  - 2.5 SD above average pause ratio
  - Below-threshold lexical richness (measured via Type-Token Ratio)

---

6. Next Steps

1. **Dataset Expansion:**
   - Source more diverse voice clips across age, gender, and dialects
2. **Temporal Modeling:**
   - Track speaker metrics over time to detect **progressive decline**
3. **Clinical Validation:**
   - Collaborate with neurologists to validate biomarkers
4. **Supervised Model Training:**
   - Use labeled clinical data to build a classification model (e.g., cognitive decline vs control)
5. **Deployable API:**
   - Wrap model in a REST API (`predict_risk(audio_path)`) returning normalized **risk score** and feature insights

---

7. Deliverables

- Clean, modular **Python notebook** with end-to-end pipeline
- **Feature plots** (e.g., speech rate vs pause ratio, pitch spread histograms)
- **Clustering visualizations** (2D projection of speaker embeddings)
- In-progress: `predict_risk()` function for API integration

---

8. Ethical & Clinical Considerations

This tool is strictly a **research POC**. Deployment in real-world or clinical settings **requires thorough validation** by qualified medical professionals. The system does not diagnose or replace neurological assessments.

Prepared For: MemoTag AI Team
Prepared By: Harshit Bansal