Conversational AI: Speech Processing and Synthesis
UCS749

Project File

AI-based diagnosis of Alzheimer’s disease

Submitted by :

Sarthak Kumar	102215231
Navansh Krishna Goswami	102215193
Manasvi Khokher	102215205
Mitali Gupta	102215054

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Thapar Institute of Engineering and Technology
Jan-May 2025

Introduction:

Alzheimer's disease is a progressive brain disorder that slowly destroys memory, thinking, and the ability to communicate. The most common form of dementia, it represents an emerging challenge to health care systems worldwide, especially in a world with a growing older population. Early detection is critical for effective management, but traditional methods like brain imaging and neuropsychological testing are often expensive, invasive, and often detect the disease only when the disease process is advanced.

In an interesting twist, one of the first indicators of Alzheimer's can manifest in someone's speech—pauses, word-finding issues, or variations in sentence structure. Such minor changes are too hard to detect with the naked ear but can be caught by Artificial Intelligence. AI-driven applications can look for speech patterns through natural language processing and machine learning to identify variations associated with cognitive decline.

This project deals with the potential for using AI to diagnose Alzheimer's disease through speech analysis for detecting differences. It delves into the technology involved in such systems, the nature of data employed, and how speech-driven AI tools might usher in a non-invasive, affordable, and cost-efficient method of early diagnosis.

Literature review:

Several studies have been conducted in the area of speech based diagnosis of neurological diseases, with various machine learning models using different architectures and features.

In [1] 1500 participants' data was collected who underwent evaluation at the Memory Clinic of Ace Alzheimer Center Barcelona over the course of about 1 year. After due preprocessing of the data obtained, composite scores created were from the NBACE battery which were used to determine the cognitive status of participants. 2 problems were addressed in the study. Firstly, the classification models were developed to differentiate between clinical phenotypes. Secondly, regression models were implemented to predict the cognitive composites. A combination of random forest, gradient boosting, SVM and K-nearest neighbour were used for classification. A vast collection of models was trained and performances were compared. The composites obtained showed strong discriminatory abilities between SCD(subjective cognitive decline) and ADD (Alzheimer's disease dementia) individuals. For evaluation, F1 scores, Precision, Sensitivity, Specificity and Balanced Accuracy were used. The best performing models were "SCD vs cognitive impairment" and "SCD vs ADD" which exhibited high accuracy metrics.

In another significant research [2], the authors developed an Adaptive Fourier Decomposition-based Frequency Modulation (AFM) model that is designed to extract speech biomarkers from dysarthric speech. The AFM model creates a rich time-frequency representation through the decomposition of speech into mono-components that reflect high amplitude and frequency features. Classifiers like LightGBM and XGBoost were trained with the TORGO dataset, whose features included bandwidth, spectral flatness, and energy ratios. The high accuracy of the models in classification means that the AFM model's characteristics can be extended for the identification of neuromotor speech impairments due to neurological disorders.

In [3] a research on how neurological diseases affect speech, researchers contrasted the acoustic, perceptual, and physiological impact of diseases such as Parkinson's disease, ALS, MS, Huntington's disease, and cerebellar disorders. Data were gathered through diverse clinical assessments and acoustic measurements. After appropriate processing, significant characteristics of speech such as pitch, loudness, jitter, shimmer, and speech rate were obtained for quantifying the impairments. Two primary aims were discussed in the study: Most significantly, distinct speech traits were first associated with distinct disorders-decreased vocal effort in Parkinson's, or tense unstable voice in MS. Another significant, objective instrumentation was used for assessing outcomes from treatment as well as previously unheard of detection of the disease. Different analysis acoustic techniques were subsequently used in speech quality comparisons. Overall, the findings showed that voice analysis could be a potent diagnostic as well as management tool for neurological speech disorders.

[4] mentions scientists are beginning to employ speech analysis as an easy, noninvasive method for assisting in detecting brain disorders. In one instance, they videotaped the tone of 83 individuals with neurological conditions such as Parkinson's disease, strokes, and memory disorder, having them sustain the sound "ah." They also recorded 53 comparison subjects who were healthy. They employed a special program called PRAAT to quantify 16 characteristics of the voice, such as how stable the sound was (referred to as jitter and shimmer) and how much background noise the voice contained. They discovered that 13 of these voice characteristics were distinctly different in sick versus healthy individuals. Jitter and shimmer were the most helpful to identify differences. When they used computer models to attempt to distinguish between patients and non-patients, one model (RBF) was better (87.5%) than another model (MLP), which achieved 83.33%.
The research indicates that quantifying such aspects of speech as jitter, shimmer, and noise can enable doctors to detect neurological issues. However, the research had some restrictions, such as involving only a few individuals and not utilizing typical speech (such as conversation). The scientists recommend that future research should involve more individuals and greater varieties of speech and sound information to enhance results.

Another significant research study [5] mentions, Physicians and researchers are discovering that speech analysis may be a useful and painless method of identifying brain disorders. In one study, researchers had 83 individuals with diseases such as Parkinson's disease, stroke, and age-related memory problems hold the sound "ah." They also had 53 healthy individuals for comparison. They analyzed these using a program called PRAAT, and they examined 16 various characteristics in the voice—like jitter (tiny variations in pitch), shimmer (tiny variations in volume), and levels of background noise. They found that 13 of those traits varied between the control group and the patients, and jitter and shimmer varied the most. They tested two types of computer models for distinguishing between patient and normal groups. The RBF network fared better than the other (Multilayer Perceptron or MLP), with about 87.5% accuracy. The research shows that jitter, shimmer, and noise in the voice of a person can be useful to identify neurological disorders. However, there were some limitations—such as using fewer individuals in the experiment and analyzing only one type of speech sound. In the future, researchers recommend using longer samples of speech, more individuals, and several voice qualities so as to make the method more precise.

Dataset:

For the purpose of this project, the dataset “**Parkinson's Speech with Multiple Types of Sound Recordings**” [6] has been used from the “UC Irvine Machine Learning Repository”.

Link to dataset

The training dataset contains 1040 data points consisting of recordings from 20 Parkinson’s disease subjects and 20 healthy subjects and with multiple types of sound recordings (26 voice samples including sustained vowels, numbers, words and short sentences) from each subject are taken.

The dataset has recordings from 20 PWP (14 male and 6 female) and 20 healthy individuals (10 male and 10 female). For each voice sampled, it also contains 26 features derived from linear and time frequency analyses along with UPDRS (Unified Parkinson’s Disease Rating Score) score which is a clinical measure assessed by a physician to determine the severity of Parkinson’s symptoms.

For the test dataset, 28 PD patients are asked to say only the sustained vowels 'a' and 'o' three times respectively which makes a total of 168 recordings.

A problem in this dataset was that the test dataset only contained samples from PWP. This leads to misleading performance results unless accounted for.

Methodology:

Since the test dataset contains samples belonging to a single class label, we first combine the test and train datasets. Then, we randomly split the dataset into training and test datasets. The parameter "stratify" was used to ensure the split of the data maintains the same distribution of the class label in the training and test data.

Before building the ML model, values in the dataset were first scaled using the Scikit-learn library.

Looking at the dataset, we had to build a model for binary classification. For this purpose, we used a number of **ensemble learning** techniques to make predictions using both a Voting Classifier and a stacking classifier.
For the voting classifier, a combination of **SVM**, **KNN** and **Random Forest** was used for this purpose with probabilistic voting for each of the components.

Another model was trained using **Gradient Boosting** (XGBoost) whose results were similar to the custom voting classifier.

Additionally, a stacking classifier was trained using Random Forest, Gradient Boosting and Support Vector Machine (SVM).

Finally, a shallow neural network was trained to perform binary classification on the dataset. PReLU was used as the activation function which treats the slope of negative inputs as a learnable parameter. Additionally, to prevent overfitting, dropout was performed with a dropout rate of 0.3. The loss function used was 'BCEwithlogitsloss' which is useful for simple binary classification tasks.

To optimize results, each ensemble model was wrapped in **GridSearchCV** to perform **hyperparameter tuning**. The list of parameters was different for each model although the metric was the same, i.e. F-1 Score.

The results from each model were then compared based on Accuracy, F-1 Score and AUC-ROC.

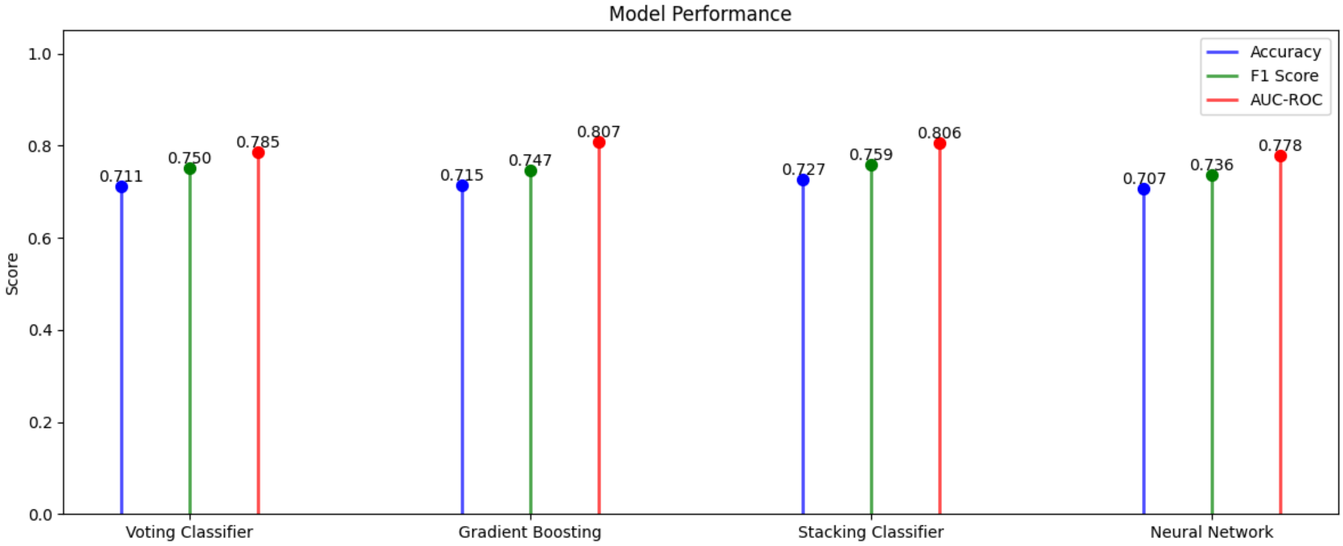
Code: Link to notebook

Results:

Since we had a fairly balanced dataset after preprocessing, accuracy would have sufficed. But even so, since we were using ensemble learning, we calculated metrics like F-1 Score and ROC-AUC (Receiver Operating Characteristic - Area Under the Curve) also.

$Accuracy \ = Number\ of\ Total\ Predictions\ /\ Total\ number\ of\ predictions$
$F1\ score \ = \ 2 \times Precision \times Recall/(Precision \ + \ Recall)$
ROC-AUC = Measures the ability of the model to distinguish between classes. The AUC value represents the probability that the model will rank a random positive instance higher than a random negative instance.

Below image shows the performance of various models based on different performance metrics.



Conclusion:

In this project, we developed a series of machine learning models to correctly predict if a person had Alzheimer's disease or not based on speech patterns. By exploring two types of ensemble learning techniques and a simple neural network, we were able to understand the differences and nuances between different one-shot classifiers and a shallow neural network.
Although XGBoost outperformed our custom voting classifier, we were able to leverage the strengths of different techniques and combine various techniques to then build a stacking classifier.
The neural network, although a completely different architecture, performed similar to the ensemble models.

This project showcased how various techniques can be used to accurately diagnose neurological disorders in people using simple and non-invasive techniques, with neural networks being one of the more scalable solutions in case of larger and more complex datasets.

Future Works:

The dataset used here was very limited in terms of data points and types of speech captured. The models deployed in this project might give better results if trained on a larger and more comprehensive dataset.

Furthermore, deep learning models like CNN's may perform better by capturing more features and revealing more information about speech features.

Finally, light-weight, real-time systems could be developed which use aforementioned ML architectures to allow people to perform basic diagnosis from the comfort of their homes.

References:

[1] García-Gutiérrez, F., Alegret, M., Marquié, M. et al. Unveiling the sound of the cognitive status: Machine Learning-based speech analysis in the Alzheimer's disease spectrum. Alz Res Therapy 16, 26 (2024). https://doi.org/10.1186/s13195-024-01394-y

[2] Shabber, S. M., & Sumesh, E. P. (2024). AFM signal model for dysarthric speech classification using speech biomarkers. *Frontiers in Human Neuroscience*, *18*, 1346297. https://doi.org/10.3389/fnhum.2024.1346297

[3]Smith, M., & Ramig, L. O. (n.d.). *Neurological disorders and the voice*. Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts; University of Colorado Health Sciences Center; University of Colorado-Boulder.

[4]Vma Rani K and Mallikarjun S. Holi, Analysis of Speech Characteristics of Neurological Diseases and their Classification.

[5]Nosirova Umida Abdusattarovna, The Relationship Between Neurological Disorders and Speech Impairments — Autism Spectrum Disorders and Speech Development.

[6] Kursun, O., Sakar, B., Isenkul, M., Sakar, C., Sertbas, A., & Gurgen, F. (2013). Parkinson's Speech with Multiple Types of Sound Recordings [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5NC8M.