Parkinson's Disease Detection by Voice Analysis

AFRINA AKHTER ANIKA(1606150), MORTUZA MINHAJ CHOWDHURY(1606157), SUDIPTA CHANDRA SARKER(1606162)
ANAS EBNA KALAM(1606169), MD SHAHEDUL HASAN(1606179)

***Abstract*— Parkinson's disease is a disorder of central nervous system. It is estimated that 90 percent of people with Parkinson's disease suffer from speech and voice disorders [1]. In this paper various features have been extracted from the voice signals of healthy people and people suffering from Parkinson's disease. With those extracted features we train different ML classifier model like KNN, SVM, RANDOM FOREST etc. We explore different feature extraction approaches like PRAAT, VGGish, MFCC, RESNET50 and compare their performances. Among all feature extraction approaches, VGGish with KNN classifier shows maximum accuracy for predicting PD (96.226% for nearest neighbor =2 & 98.113% for nearest neighbor =3).**

Keywords—RESNET50, VGGish,PRAAT,KNN, SVM

INTRODUCTION

Parkinson's disease (PD) is the second most common age-related neurodegenerative disorder (after Alzheimer's), affecting about 7 to 10 million people worldwide [2]. The gradual decaying of the neurons that produce a chemical called dopamine causes abnormal brain activities that result in PD symptoms [3]. This disease affects patients' quality of life, makes social interaction more difficult for them, and worsens their financial condition with extravagant medical expenses [4]. PD causes several motor and non-motor symptoms, which gradually become prominent after different disease progression stages. One of the secondary motor symptoms that people with PD may experience is the change in their speech quality or difficulty speaking in worse cases [5]. Currently, there is no single, definitive test to diagnose PD. Instead, PD is diagnosed through a combination of clinical criteria, which include the measurement of the deterioration of two or more motor symptoms over time. One of the most common scales for the assessment and tracking of these impairments is the Unified Parkinson's Disease Rating Scale (UPDRS). Although UPDRS-based measurements along with other clinical tests have significant predictive power and can diagnose PD with up to 90% accuracy, clinicians need to follow patients for up to 5 years before making a definitive diagnosis [6]. But, early detection of PD is essential as the treatments such as levodopa/carbidopa are more effective if administered in the early stages of the disease [7]. The analysis of patient voice data offers one potential route to streamline the PD diagnosis process.

Over the past decade, several studies have illustrated that speech-based machine learning (ML) classification models can be employed to detect PD. The majority of these studies have relied on groups of handpicked features that are often extracted using clinical acoustic analysis software like PRAAT to train ML classifiers. But, their reliance on manually prespecified feature sets may limit performance and generalizability. In this research we conduct an analysis that compares classification performance achieved using several different feature extraction approaches. We compare the utility of using Convolutional Neural Network generated audio embeddings to that of handcrafted feature sets for training voice-based ML classification models. We explore the diagnostic performance improvements that can be achieved from leveraging these CNN approaches, including VGGish embeddings, MFCC, RESNET50 as compared to previous signal processing techniques.

METHODOLOGY

On speech data captured on a smartphone, we compare the outcomes of four alternative feature extraction algorithms for building binary PD diagnosis models. PRAAT is a voice Analytical Software. We have extracted 14 features using PRAAT.

I. Features extracted using PRAAT

1. Fo (Hz): Mean fundamental frequency.
2. Fhi (Hz): maximum fundamental frequency.
3. Flo (HZ): minimum fundamental frequency.
4. Jitter (%): This is the average absolute difference between consecutive periods of fundamental frequency, divided by the average period (expressed as a percentage)
5. Jitter (ABS): This is the average absolute difference between consecutive periods of fundamental frequency, in microseconds (μs)
6. Jitter (RAP): This is the Relative Average Perturbation, the average absolute difference between a period of fundamental frequency and the average of it and its two neighbors, divided by the average period.

7. Jitter (DDP): This is the average absolute difference between consecutive differences between consecutive periods, divided by the average period
8. Shimmer: This is the average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude
9. Shimmer (dB): This is the average absolute base-10 logarithm of the difference between the amplitudes of consecutive periods
10. Shimmer (APQ3): This is the three-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of its neighbors, divided by the average amplitude.
11. Shimmer (APQ5): This is the five-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of it and its four closest neighbors, divided by the average amplitude
12. Shimmer (DDA): This is the average absolute difference between consecutive differences between the amplitudes of consecutive periods.
13. HNR: harmonics to noise ratio.
14. NHR: noise to harmonics ratio.

There are 42 participants in the dataset, 21 of them are healthy and the other 21 are Parkinson's patients. We took a four-second sample. There are a total of 199 samples, with 115 samples from healthy patients. Audio data for this research was obtained from the Mobile Device Voice Recordings dataset released by King's College London (MDVR-KCL) [8].

II. Extra Feature Extraction for improving result

a. Preprocessing

We have used MATLAB code to extract additional audio characteristics from our audio files in order to get a better outcome. We did this by first zero padding and then framing our 4-second audio file into 200-millisecond pieces. We employed a hanning window for each frame to reduce spectral flux.

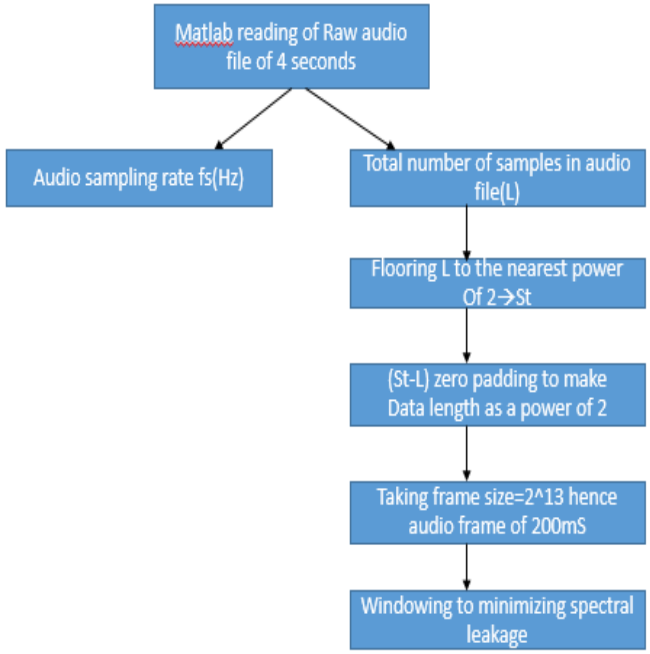

Fig. 1. algorithm of Preprocessing

b. Envelope Analysis

From windowed frames we extracted maximum values from each frame and stored them in a maximum frame value matrix. To focus more on how the values are changing in respect to maximum value of new matrix, we normalized our matrix by dividing all elements by the maximum value of new matrix. From this normalized matrix we found average and average deviation.

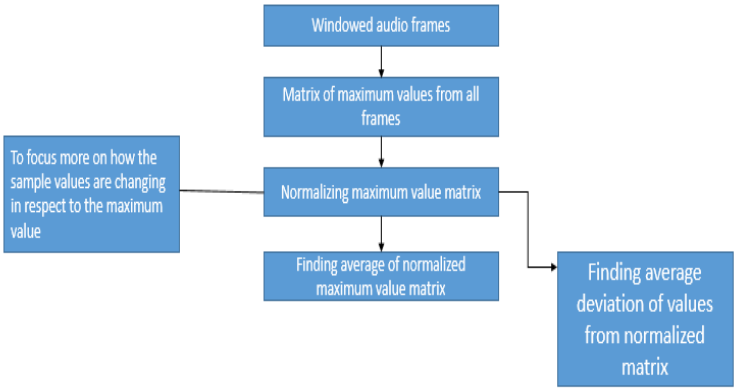

Fig. 2. algorithm of Envelope Analysis

c. Root Mean Square Energy

Root mean square energy is an important audio feature for audio analysis as it changes rapidly for different kind of audio files. From normalized windowed audio frame maximum value matrix, we extracted the square matrix and hence found normalized RMS.

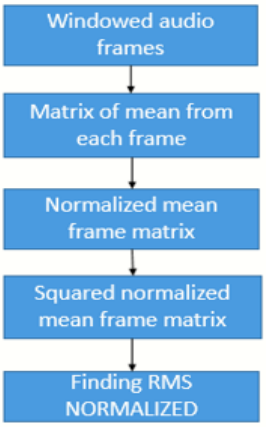

Fig. 3. Algorithm of Root Mean Square Energy

d. Zero Crossing Rate

Finally, we used the zerocrossing rate to determine the erraticity of audio recordings. Every four-second audio file was broken into four frames of one second each. We obtained four distinct zero crossing rates by calculating the number of sign changes in each frame and dividing it by the frame size. After that, we computed their average.

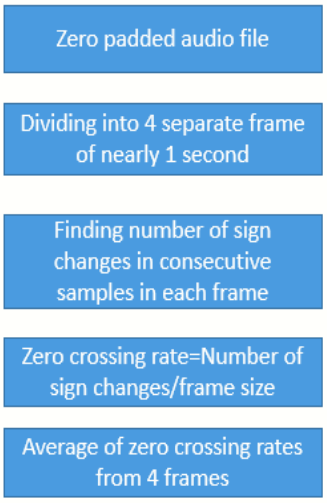

Fig. 4. Algorithm of Zero Crossing Rate

III. MFCC

A signal's mel frequency cepstral coefficients (MFCCs) are a small group of characteristics that succinctly define the overall shape of a spectral envelope.

IV. VGGish

It is a pretrained convolutional neural network based on the well-known VGG image categorization networks. VGGish turns audio input data into a 128-D embedding with semantic meaning that can be fed into a classification model.

V. RESNET50*:*

It is a 50-layer deep conventional pretrained neural network. More than a million photos from the Imagenet collection were used to train it. With a Maxpool and an average Pool Layer, it contains 48 convolution layers.

VI. Transfer Learning:

When components of a pre-trained machine learning model are reused in a new machine learning model, this is known as transfer learning. It is the process of fine-tuning the network by training the pre-trained model with additional data.
After feature extraction, all data were standardized prior to training each classifier. Each model was trained on an evenly balanced training set that contained 70% of the voice clips. To avoid the leaking of patient-specific information, each training set patient's clips belonged to just one-fold. A held-out set containing 20% of the data was used to assess performance.

RESULT

By extracting features by PRAAT, KNN and Logistic regression have showed highest accuracy in test data(a). To improve the accuracy we have extracted Preprocessing, Envelope Analysis, Root Mean Square Energy, Zero Crossing Rate and MFCC by MATLAB and added with the PRAAT features. Then the accuracy improves(b).

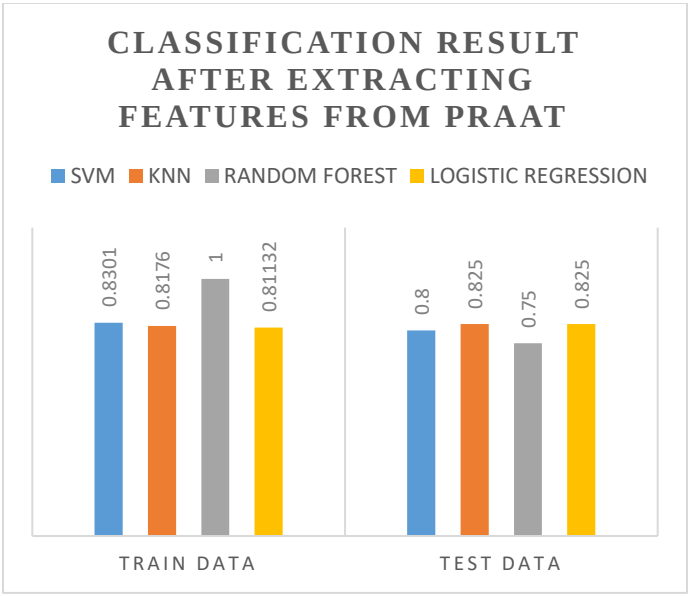

Fig. 5. Classification result from PRAAT

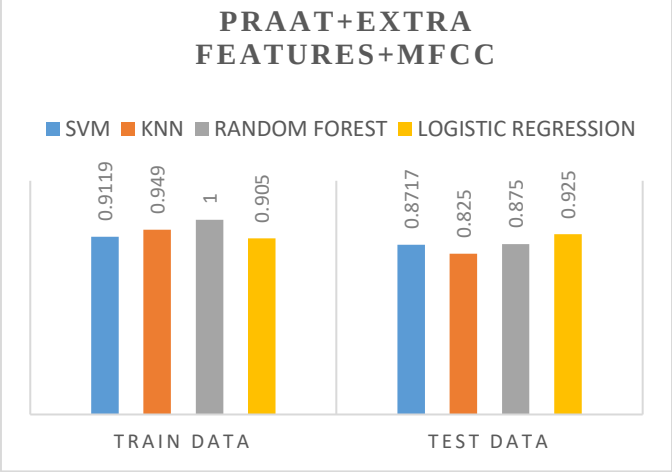

Fig. 6. Classification result from PRAAT+Extra featues+MFCC

VGGish embeddings were the most accurate in predicting the patient among the four feature extraction approaches used. In all classifiers, VGGish and PRAAT performed better than the others. Transfer learning was added to increase the accuracy of RESNET50, and it improved the accuracy. The accuracy increased to 85.71 percent, although not to the level of VGGish.

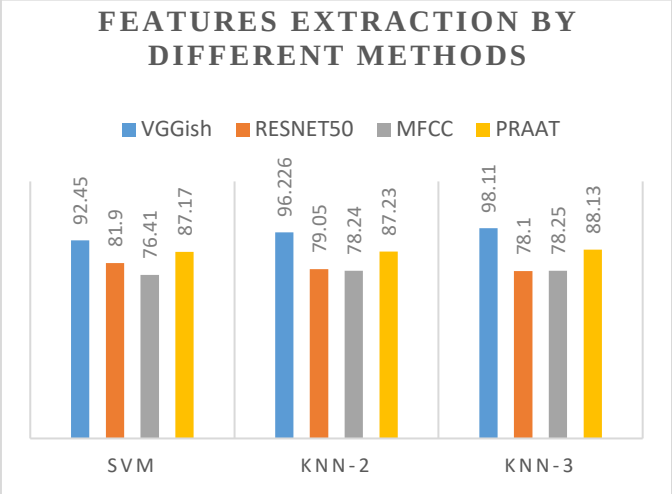

Fig. 6. Classification result from different features extraction method

Conclusion

Through the KNN pipeline the VGGish feature extraction outperformed other feature representations in predicting PD based speech abnormalities. The VGGish KNN pipeline is highly competitive with clinical UPDRS III-18 labels in detecting PD. While these clinical labels were determined after review of the overall 1 to 4-minute file, our classifiers achieved similar performance with just 4 seconds of input. A VGGish embeddings-based pipeline provides several advantages. The pipeline eliminates the need for an expert handpicked feature set to inform PD detection. While there exists an overall set of established features for dysphonia measurements, there is no established subset for PD diagnosis. The VGGish strategy offers an accurate and easily applicable alternative. Further, the VGGish strategy robustly differentiated standard speech clips and does not require audio input that contains long vowel sounds or other specific vocal tones, which were often required for achieving high accuracy in previous classification approaches. As for future work, hardware implementation can be done by introducing the cloud storage through smartphones. The remote monitoring of PD patients can be performed by analyzing their voice records collected through their smartphone. This collected database can be used upon their consent for population studies on the incidence of Parkinson's, which are essential to scientists' understanding of its history, progression, and risk factors. In this way, this system can help healthcare experts design strategies to meet patients' needs, especially for rural areas where access to a neurologist is minimal.

Acknowledgment

It is a great pleasure for us to express unfettered gratification, sincere appreciation and profound respect to our course instructor Shahed Ahmed & Talha Ibn Mahmud sir.

References

[1] R. A. Shirvan and E. Tahami, "Voice analysis for detecting Parkinson's disease using genetic algorithm and KNN classification method," 2011 18th Iranian Conference of Biomedical Engineering (ICBME), 2011, pp. 278-283, doi:10.1109/ICBME.2011.6168572.

[2] Parkinson's Disease Statistics. https ://parkinsonsnewstoday.com/parkinsons-disease-statistics /

[3] Men More Likely to Get Parkinson's Disease? https ://www.webmd.com/parkinsons-disease/news/20040 317/men-more-likely-to-get-parkinsons-disease

[4] Bovolenta T, Azevedo S, Saba R, Borges V, Ferraz H, Felicio A (2017) Average

[5] annual cost of Parkinson's disease in São Paulo, Brazil, with a focus on

[6] disease-related motor symptoms. Clin Intervent Aging 12:2095–2108

[7] Symptoms—Speech Difficulties or Changes. https://parkinsonsdisea

[8] D. Trivedi H. Jaeger and M. Stadtschnitzer. 2019. Mobile Device Voice Recordings at King's College London (MDVR-KCL) from both early and advanced Parkinson's disease patients and healthy controls. https://doi.org/10.5281/zenodo.2867216 Data set.

[9] S. Kurada and A. Kurada, "Poster: VGGish Embeddings Based Audio Classifiers to Improve Parkinson's Disease Diagnosis," 2020 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), 2020, pp. 9-11, doi: 10.1145/3384420.3431775.