Non-Invasive Prediction of Parkinson's Disease

submitted in partial fulfillment of the requirement
for the award of the Degree of

Bachelor of Technology
in
Computer Engineering

by

Jinang Gandhi (2017130019)
Aumkar Gadekar (2017130023)
Tania Rajabally (2017130047)

under the guidance of

Dr. Preetida Vinayakray-Jani

Department of Computer Engineering
Bharatiya Vidya Bhavan's
Sardar Patel Institute of Technology
(Autonomous Institute Affiliated to University of Mumbai)
Munshi Nagar, Andheri-West, Mumbai-400058
University of Mumbai
October 2020

Certificate

This is to certify that the Project entitled "NON INVASIVE DETECTION OF PARKINSON'S DISEASE" has been completed to our satisfaction by Mr. Jinang Gandhi, Mr. Aumkar Gadekar and Ms. Tania Rajabally under the guidance of Dr Preetida Vinayakray-Jani for the award of Degree of Bachelor of Technology in Computer Engineering from University of Mumbai.

Certified by

Dr Preetida Vinakyakray-Jani
Project Guide

Dr. Anant V. Nimkar
Head of Department

Dr. Bhalchandra Chaudhari
Principal

Department of Computer Engineering
Bharatiya Vidya Bhavan's
Sardar Patel Institute of Technology
(Autonomous Institute Affiliated to University of Mumbai)
Munshi Nagar, Andheri(W), Mumbai-400058
University of Mumbai
April 2019

Project Approval Certificate

This is to certify that the Project entitled “NON INVASIVE DETECTION OF PARKINSON’S DISEASE” by Mr. Jinang Gandhi, Mr. Aumkar Gadekar and Ms. Tania Rajabally is found to be satisfactory and is approved for the award of Degree of Bachelor of Technology in Computer Engineering from University of Mumbai.

External Examiner	**Internal Examiner**
(signature)	**(signature)**
Name:	**Name:**
Date:	**Date:**

Seal of the Institute

Statement by the Candidates

We wish to state that the work embodied in this thesis titled "NON INVASIVE DETECTION OF PARKINSON'S DISEASE" forms our own contribution to the work carried out under the guidance of Dr Preetida Vinayakrai-Jani at the Sardar Patel Institute of Technology. We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission.

Name and Signature:
1.Jinang Gandhi
2.Aumkar Gadekar
3.Tania Rajabally

Acknowledgments

It is really a pleasure to acknowledge the help and support that has gone to in making this thesis. I express my sincere gratitude to my guide Dr Preetida Vinayakray-Jani for her invaluable guidance. I also thank her for encouraging me to work on a topic of such importance in today's time. Without her encouragement this work would not be a reality. With the freedom she provided, I really enjoyed working under her. Her constant support and guidance encouraged me to give my best.

I would also like to thank our co-guide Prof Dayanand Ambawde. His valuable insights helped us to carry out our research in the right direction.

I thank my examiners, Prof. Jyoti Ramteke for her word of advice.

I thank HOD and staff of Computer Engineering Department for giving me all the facilities required to carry out this research work.

I would like to thank all my family members and well wishers for their constant encouragement for all these years, without which I could not have completed this work.

Contents

List of Figures

List of Tables

List of Abbreviations

PD	Parkinson's Disease
SST	Static Spiral Test
DST	Dynamic Spiral Test
CNN	Convolutional Neural Networks
MNIST	Modified National Institute of Standards and Technology
LSTM	Long Short Term Memory
SVM	Support Vector Machine
UPDRS	Unified Parkinson's Disease Rating Scale
ML	Machine Learning
ResNet	Residual Neural Network
ReLU	Rectified Linear Unit
AC	Mean Autocorrelation
NTH	Mean Noise-to-Harmonics ratio
HTN	Mean Harmonics-to-Noise ratio
SMOTE	Synthetic Minority Oversampling Technique
KNN	K-Nearest Neighbors
XGBoost	Extreme Gradient Boosting
LGBM	Light Gradient Boosting Machine

Abstract

Parkinson's is a disease affecting 700000 plus people in India. It primarily affects the nervous system and has many symptoms. These symptoms manifest in the form of lack of motor skills, verbal and mental abilities, affected speech and memory. The disease has no cure and the best remedy is treatment to relieve the patient and reduce the severity of the symptoms. The early diagnosis can help in the treatment of the patient before the symptoms are exacerbated. The lack of easy access to medical facilities and doctors is a major deterrent to early diagnosis of Parkinson's.

Therefore, we propose here a solution in the form of a mobile application that allows people to test themselves for symptoms of Parkinson's anywhere, anytime. The solution uses a two-way approach to detect the probability of the patient having Parkinson's. The authors posit image processing and machine learning techniques to detect anomalies in speech and tracing patterns.

Chapter 1

Introduction

This work is concerned with the easy detection of Parkinson's disease. The scope of the problem as follows:

- To increase the reach amongst the people and remote areas as it is based on mobile system.
- To solve the issue of language barrier for speech tests by implementing a phonetic model.
- To use widely available commercial grade smartphones; avoiding the use of any other custom hardware.
- The application also acts as a source of data.

Parkinson's disease is a nervous system disorder which starts with barely noticeable symptoms and gradually develops into major movement disorders. It is the second most prevalent neurodegenerative disorder. Parkinsonian syndrome is caused due to damage of the nerve cells in the brain which causes dopamine levels to drop. Symptoms range from tremors and slowed movement, loss of memory, cognitive disorders, to changes in speech and writing. It consists of 5 stages with increasing severity and can lead to a complete lack of motion.

Recognising the symptoms at an early stage is of prime importance. Identifying this disease is very difficult as there is no laboratory test to diagnose with certainty and the symptoms are common to various disorders. Hence, accurate prediction is required for effective action. There is no treatment of Parkinson's disease but if medication is begun at an early stage, it is more efficacious in relieving the symptoms. In a lot of patients, it goes undetected because they simply fail to visit doctors. Often, neurologists specializing in Parkinson's may not be available for consultation. While a doctor is definitely needed for an official diagnosis of Parkinson's, advances in computer vision and artificial intelligence make it possible to predict the presence of a disease with reasonable accuracy.

To aid in the early diagnosis, numerous tests have been devised which look for biological markers to estimate if a person has Parkinson's.A large number of current prediction systems for Parkinson's disease are based on a single test. Or then require special hardware like motion sensors or special smart tablets with accessories. The drawback of having single tests for prediction is that a person may perform poorly or below average on a certain test due to physical deformity or some accident rather than it being a symptom of Parkinson's. Also, the symptoms are common to multiple diseases and if verified by a single test, can lead to misdiagnosis. Moreover, PD is more prevalent in the elderly and hence using custom hardware will require some amount of assistance.

The research in this paper proposes a user friendly, portable, non-invasive system to predict the prevalence of Parkinson's disease based on the onset of its early symptoms. To solve the above mentioned problems, the proposed system will take into consideration two distinct tests for predicting whether the individual has Parkinson's. One of the most prevalent early observed symptom is tremors. It becomes very difficult for a person having tremors to draw or trace lines. Previous research has demonstrated a sharp contrast between tracing made by healthy people and those having loss of motor skills as experienced by PD patients. However, since tremors are also observed in other motor-skill related diseases, and often simply due to old age, this can be prone to mis-classification. Therefore, to consolidate the results, a speech test is also conducted. Since Parkinson's also affects a person's control over their voice, PD patients suffer from dysphonia, which manifests in abnormalities in measures such as jitter, shimmer, voice breaks, and other properties of sound. The proposed system, which is implemented on smartphones, will incorporate these tests in the form of an android application, The application accepts input in the form of tracing of common shapes, and voice recordings of the vowel phonation 'aaa'. The two inputs are analysed and fed to two machine learning models. The outputs are combined to give an overall score which serves as a probability of the subject having Parkinson's.

1.1 Motivation

The evolution of technology has brought about mergers with other domains. Technology in combination with the medical sector is of utmost importance. This is because of the significance of the medical industry and the problem solving nature of technology. The union of such kind makes it easier to treat illnesses and save lives.

- An estimated 70 lakh to 1 crore people worldwide have Parkinson's disease, with 7 lakh + people in India alone.
- A lot of people in our country lack access to basic medical facilities or qualified doctors.
- With the ongoing pandemic; the accessibility to medical facilities is reduced further.
- For a disease like Parkinson's with no cure, early detection is needed so treatment for alleviating the symptoms can be given.
- A solution which is easily accessible to the people, easily affordable and which allows people to test themselves for the symptoms without any external help / hardware is required.

1.2 Objectives

- We aim to develop a non invasive system that would detect Parkinson's disease.
- To research extensively on previous work and ascertain which features are best discriminators of healthy people and subjects with Parkinson's.
- To employ a multi-modal approach to pick on various biomarkers, covering more symptoms.
- To provide a convenient, user friendly and portable system mechanism.
- To deploy the application on Appstore and Playstore.

1.3 Problem Statement

Our problem statement divides into two parts :

1. To conduct a thorough comparative analysis of existing technologies for detecting symptoms of Parkinson's disease. Identify the gap and scope of research in this domain.
2. Design a non invasive, multi modal system for prediction of Parkinson's disease in the form of a portable, user friendly mobile application.

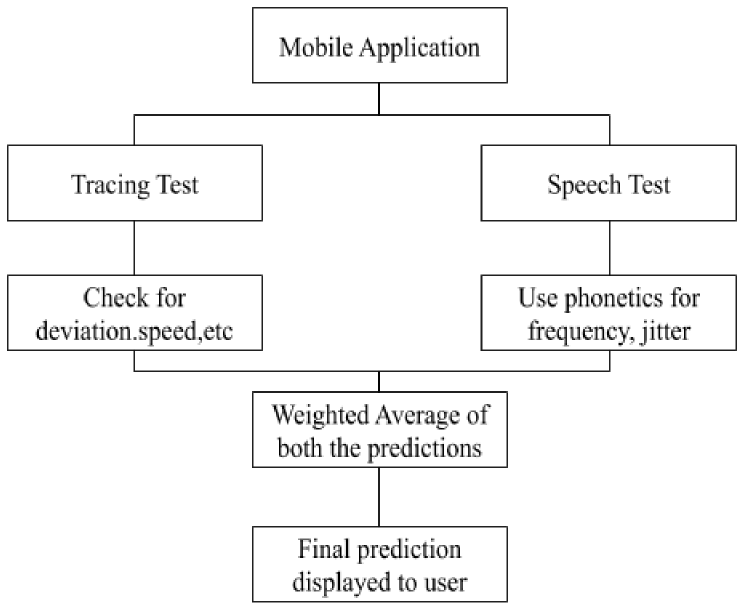


Figure 1.1: Problem Statement

1.4 Contributions

- Most of the existing mobile applications focus on the therapy of Parkinson's Disease rather than its detection. Our system will provide with early detection of the disease.
- Robust, user-friendly and one of a kind application which will be affordable to the masses.
- Will consider various symptoms to provide accurate analysis.

1.5 Layout of the Report

A brief chapter by chapter overview is presented here.

Chapter 2: A literature review of different tests carried out for the detection of Parkinson's disease is presented.

Chapter 3: Design and working of the system will be described in this chapter.

Chapter 4: In this chapter, the two tests that will be carried out and the models used will be presented.

Chapter 5: This section highlights the comparisons within the various models and explains the results of the best model obtained.

Chapter 6: This chapter will closes the archives of the system with the help of a brief summary on how the system works and what are its general aims.

Chapter 7: This chapter discusses the future scope of the system and the updates and evolutions that can be done over time.

Chapter 8: Conclusions and discussion on future course of research work.

Chapter 2

Literature Survey

The papers surveyed used two common methods for detecting symptoms of PD, namely, impairment of motor movements by detection of tremors, and speech impairment.

2.1 Impairment of motor movements

Checking for tremors in one of the most common methods adopted. Some papers adopted the method of checking deviation from a trace whereas others concentrated on the handwriting of the user. It is important to determine which parameters are most useful for classification from tremors. P. Drotar et al. analysed the prediction potential of various handwriting and tracing tasks to determine which perform as the best classifiers for Parkinson's Disease [7]. Their work concluded that the best classification accuracy is achieved by evaluating the patients across 4 different tasks and 14 features extracted from those tests. The work of Jan Mucha et al. proposed the use of fractional derivatives on features derived from the handwriting dataset [14]. The 3 most discriminative features identified were : vertical acceleration, vertical jerk, and vertical velocity.

One method cascades Chi2 model for feature extraction from images and AdaBoost model for classification [3]. Whereas another approach [9] uses a special graphics tablet where speed, acceleration are recorded. It introduces a new test: Dynamic Spiral Test which tests memory and tremors simultaneously. This paper shows that correlation of acceleration between SST and DST is an important metric for classification. Acceleration readings for SST and DST test are highly correlated incase of Control group; whereas they are much less correlated in case of Parkinson's Disease patients.

An interesting implementation [10] increases accuracy while reducing the amount of input features. Similar accuracy is observed when the CNN model is trained with actual optical images and extracted feature images. Building on this, another system [15] implements Transfer Learning on Handwriting images and yields 98.28 % accuracy. It uses a fine-tuned AlexNet model on ImageNet and MNIST datasets as well as AlexNet Freeze model on ImageNet and MNIST datasets. The study concludes that best accuracy is observed for Spiral trace images. We infer that customization with transfer learning can yield more accurate results.

A different approach was adopted by [1] where the user has to install an application which checks key hold time and latency. 27 features were extracted from this and the model provided an accuracy of 94 %.

2.2 Speech impairment

Speech changes are frequently observed during the onset of Parkinson's disease. This included abnormal vocal sounds or dysphonia, as well difficult in articulating. Dysphonic symptoms include reduced loudness, rough voice, vocal tremors and deviations in pitch.

2.2.1 Datasets

A problem encountered in the study was the lack of large scale datasets. Although numerous studies have been conducted on detecting Parkinson's through speech impairment, most available datasets suffer from one or both of the following two problems :

- Small size of samples : Numerous datasets have sample size of less than 1000 or even 500, which makes creating a robust classifier difficult.
- Lack of reliability of Data : Certain datasets enable participants of the study to submit their voice recordings over the phone, from anywhere. This data is often not verified, and also susceptible to background noise.

In addition to this, certain features of speech differ based on the dialect, language, accent, gender, age and other parameters. As such, it is difficult to obtain a database from a diverse demographic of participants. Four prominent voice-based datasets were analysed, and the observations noted below.

1) Oxford Parkinson's Disease Detection Dataset

Max little et al, in their work on identifying suitable features of dysphonia as discriminators of Parkinson's, worked on a small database of 195 samples[13]. This dataset has biomedical voice measurements from 31 people, out of which 23 have Parkinson's disease. The recordings themselves are not made public, rather they have been processed to give certain features for identifying dysphonia. The features are both non-linear and linear measures of frequency, amplitude,variations in sound, noise and harmonics. The variation in frequency, jitter and the variation in amplitude have both been identified as important features for discriminating PD patients from healthy controls. Extensive research on that dataset has resulted in methods with over 99% accuracy, but the drawback is the extremely small sample size.

2) Parkinson Speech Dataset with Multiple Types of Sound Recordings

The dataset was recording at the Department of Neurology in Cerrahpasa Faculty of Medicine, Istanbul University, by Saker et al in their work in this field[16]. Recordings of 40 subjects, 20 with PD and 20 without have been used to generate 1040 samples. The participants were asked to give numerous vocal recordings, from phonations of vowels, to words and sentences. The data has been passed through a PRAAT script to generate 26 features. Some of these overlap with the features in

Little et al's work, in addition to new measurements. Again, these include various measurements of jitter and shimmer, pitch and period, voice breaks and pulses in the recordings. The records have also been supplemented with UPDRS ((Unified Parkinson's Disease Rating Scale) score which is determined by expert physician.

3) Italian Parkinson's voice and Speech dataset

G. Dimauru and F. Girardi conducted a study of 65 Italian people, 28 with Parkinson's and 37 healthy controls[8]. The recordings are available on IEEE Dataport. The subjects were made to sound various vowels, words and sentences. The total sample size is close to 800. Also available is the gender and age of the participant. The healthy controls have been split into two categories for further analysis, old people without PD, and young participants. The dataset is relatively unexplored and has few citations. The original authors built a speech to text system and tested it on the dataset, since the clarity of speech of PD patients is lower. It was found that the system failed to identify the speech on the recordings of PD patients.

4) mPower Dataset

The mPower dataset present on synapse.org is by far the most extensive dataset of its kind. Over 65,000 records by more than 6000 participants were collected over mobile-phone voice recordings.[5] The participants were also asked to fill a survey with personal details, as well as any medical history, behavior tendencies, and history of diagnosis and medication. This dataset overcomes the problem of small sample size, of previous datasets. However, since the entire data is self verified by participants, and not recorded in the presence of researchers, there numerous bad records present. Therefore, significant pre-processing is needed to extract only the useful samples.

2.2.2 Previous research work

Abnormality in vocal features or dysphonia was the focus of research of significant papers surveyed. The authors of the paper in [17] worked on the Istanbul dataset of 26 acoustic features. K means and decision tree was then applied to obtain an accuracy between 88 % to 94 %. The same dataset was also in this paper [2] which include frequency(jitter), pulse, amplitude, voicing, harmonicity, pitch parameters. Extreme Machine Learning was then used to obtain 81.55 % accuracy on an independent dataset.

In [13], working on the Oxford Dataset, phonations of vowels are used for prediction as these are unlikely to differ with dialects and language. The user is supposed to hold the pitch for a single vowel as long and as constant as possible. 91.40 % accuracy was obtained using SVM. The paper identifies pitch period entropy (PPE) as an important feature. This research is corroborated by the work in paper [6] which also identifies non linear parameters in addition to PPE being important. It performs a thorough analsysis of several dysphonic features to identify 16 dominant features, and achieves an accuracy of 89%.

A measure of variation in fluctuation and the recurrence probability is used for classification of voice samples in [12]. This approach is applicable to any type of voice disorder. It obtained an accuracy of 91.8 % which is better than traditional methods.

Certain other papers worked on spectral features instead of acoustic measures. In [18], the authors used the voice recordings to generate spectral images of the sound data, which were then processed using CNN-Encoders, to give an overall accuracy of 91.17%. The system proposed in [20] converts the sampled voice sample into spectrogram image via Joint Time Fourier Transform. It implements a CNN model and shows superior results on spectrogram image as compared to Time domain image. It was also observed that LSTM may not be useful in phonetic voice samples.

A similar approach [11] was adopted where vowels were used to generate time series by Markov process. It uses voice sample classification for any disorder. It also shows that non linear models work better on time series and speech as compared to linear models. A complex but highly accurate novel approach for voice sample analysis is explored here.

2.3 Combined Approaches

Apart from these individual tests, there are some implementations which carry out multiple tests to predict the outcome. A detailed survey [21] was carried out to study the existing solutions and the drawbacks associated with them. Impaired gait and balance and sleep disorder, one of the earliest symptoms, was tested using the gyroscope and accelerometer of a smartphone. Vocal impairments were observed through the pitch, jitter, shimmer and repetition in the speech of the user. Handwriting analysis, tracing as well as tapping a finger on the screen was used to check for tremors and speed. This paper emphasized on checking for non motor disorders like mood disorders which generally develop before motor disorders. This can be done via a gyroscope as well as emotion detection from speech. The authors inferred that tests for multiple symptoms aid in effective classification.

This pilot study [4] explores classification for Parkinson's Disease via the data collected from customer grade smartphones. A multitest smartphone application is developed in this system. This application consists of five tests: Phonetic voice test, Posture test, Gait test (walking test), Finger tapping test and Reaction time test. A high sensitivity of 96.2 %, and high specificity of 96.9 % were observed. Random Forest classifier is used to output UPDRS score in the scale of 11 to 34. This paper shows the effectiveness of data collected from smartphones for Parkinson's Disease classification. Its future scope emphasises on large scale assessments and passive monitoring tests as they have shown high accuracy.

2.4 Observations

The current existing systems check for a single test. The symptoms of Parkinson's disease are common to multiple diseases and hence single tests can result in misdiagnosis. Therefore, rather than checking for a single symptom, different types (pertaining to different domains) of symptoms should be tested and the final prediction should be given on the basis of results of all the individual tests.

Various implementations also require special hardware such as sensors, tablets with stylus, graphics tablet or applications that run only on a computer. This may not be easily available or convenient. Elderly people may not be familiar with computer based systems. Hence these solutions do not cater to a large number of people including the elderly who are the prime victims of Parkinson's disease.

Some systems show comparatively higher accuracy using different approaches. However most of the time these methods are quite complex and hence require a lot of computation and preprocessing of data. Even if the model is hosted on a server; data extraction for feeding into the model is of highly computational nature; which would not be suitable for a smartphone application.

Few of the voice based systems are language specific and hence have limited usage. The model is trained on sentences in a particular language and extracts features relevant to that language. Using a more generalised way without any language barriers will make the application more accessible.

The pilot studies carried out for mobile applications were on limited data, of approximately around 20 people. Data is very important for training the model. The larger the dataset, better can the model be trained to generalize well and to sustain high accuracy even on previously unseen data.

Chapter 3

Design

This section elaborates the entire chronology of the processes involved alongside insights into the algorithm. Fig. 3.1 shows the working of the system.

The application consists of a two-part test, the first being a tracing test while the second, the speech test. The tests can be given in any order one after the other, but both must be given in order to generate a prediction.

1. Trace test :

   The trace test offers two options to the user: either they can trace the image on the phone screen, or they can trace the printed image. In the initial case, the figure to be traced appears on the screen, and the user's trace is captured to generate the image. In the second case, the user traces the image on a printed piece of paper having the shape to be traced. A picture of the traced image is clicked and uploaded to the application.

   This image is transmitted to the server, where it is fed into the ML-Model. This module separates the user's trace from the original trace object. It extracts the relevant features present in the person's writing and measures the dissimilarity between these and the image given for tracing. Based on the computations using this dissimilarity, it generates a score.

2. Speech test :

   The second test is the speech test. The user is prompted to record a stretched vowel sound for 10 seconds. The app records the speech. The recorded signal is then uploaded to the server. It feeds the speech into the PRAAT script, which extracts features like jitter, shimmer, frequency. These features from the speech are fed into a machine learning model. The model can discriminate between the normal speech of a healthy person and the slurred speech of someone suffering from Parkinson's. The model generates a score based on the values.

3. Combined results :

   The two scores generated from the two models are combined by a weighted average.

   This gives us the final output which tells us how likely it is that the person suffers from Parkinsons. The user can then consult a doctor accordingly.

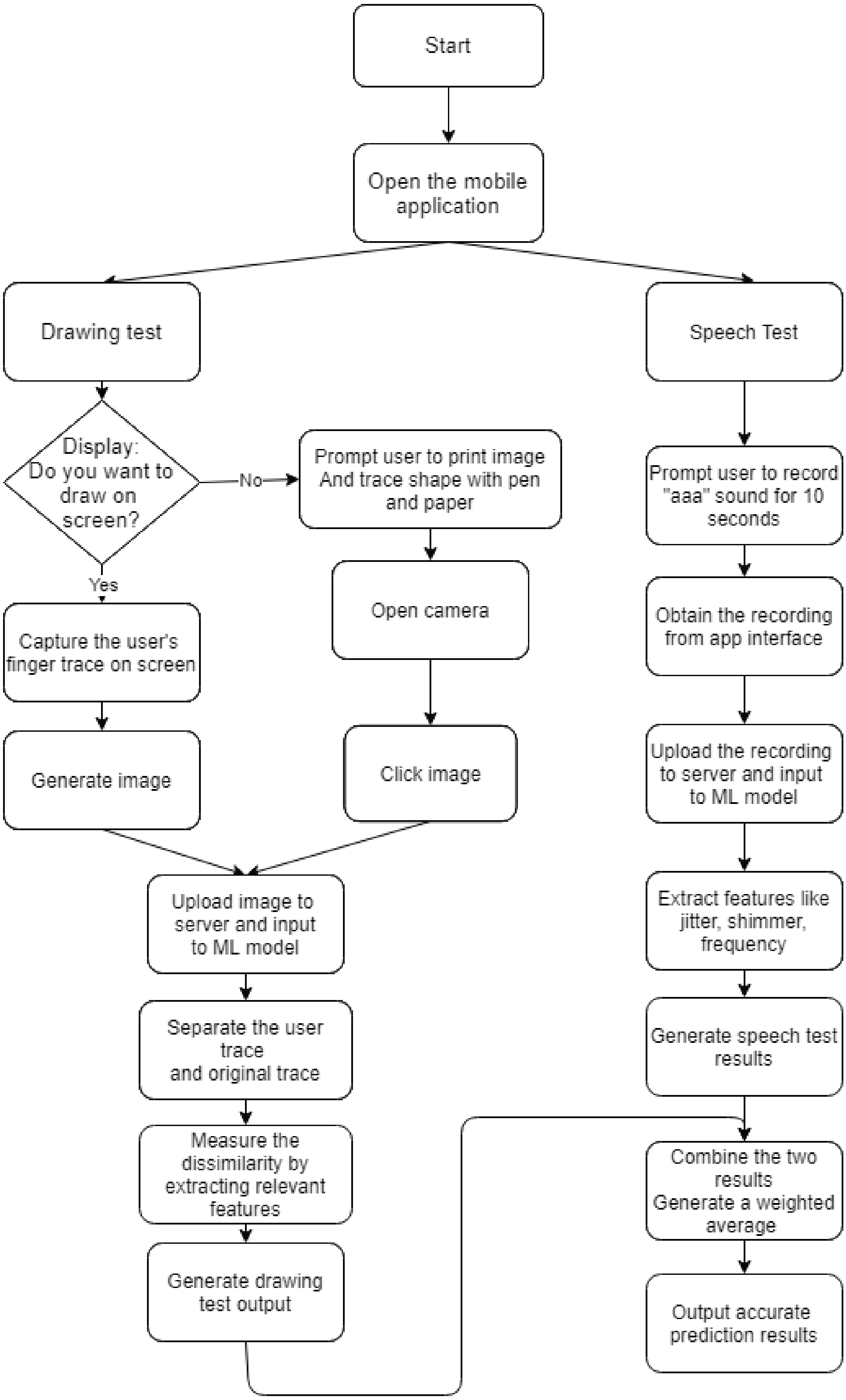


Figure 3.1: Working of the application.

Chapter 4

Implementation

As discussed, the proposed system is a mobile application which will have two modules.

1. Trace Test
   This module will detect and classify the extent of tremors in an individual and assign a score for this test. For this test, the user will have to trace a laid-out meander trace image. The deviations from the trace and the nature of the user trace are key factors in determining the test score.

2. Speech Test
   The purpose of this test is to extract features of a voice sample of an individual and feed it to a Machine Learning model which will output a score indicating the likeliness of that sample belonging to PD group. To standardize language and accent constraints; only vowel sounds are considered. User will be asked to record a 10 second sample making "aaa" sound. The features required will be extracted from this voice sample.

Lastly, on the basis of scores for both the tests; a weighted average will be calculated. And this would be the final score which would serve as a result; indicating how likely is the candidate to have Parkinson's disease.

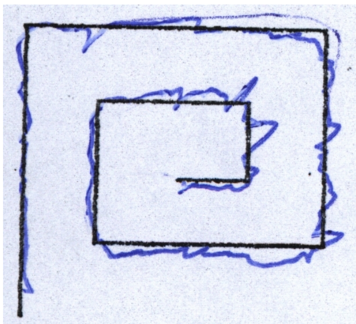
4.1 Trace Test

4.1.1 Dataset

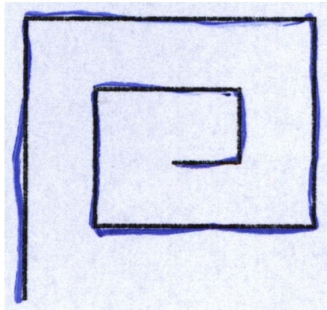
The Meander images of the NewHandPD dataset are used for the model training. It consist of images for different trace shapes and signals of the PD and Control group participants. Shapes like Spiral, Circle and Meander are used as the trace. Out of which; the Meander trace was found to have more separable data. It contains around 264 images with nearly even distribution between the gender of the participants.

4.1.2 Model for Trace test

A Convolutional Neural Network based model was used in this module. Different models such as Inceptionv3, Xception, ResNet, VGG16, VGG19 and MobileNet were used to fit the data using transfer learning. All of these models were pre-trained



(a) Tracing Of PD Patient



(b) Tracing of Healthy control

Figure 4.1: Sample images of the Meander NewHandPD dataset

on the ImageNet dataset. The ImageNet dataset consists of more than 14 million images and contains more than 1000 classes of images. Custom layers were added to these models by trial and intuition.

Since the size of the dataset was fairly small, the study incorporated augmentation of images. The images in the dataset were flipped horizontally and vertically to generate more test images. Owning to the symmetry of the trace; the flipping of the images does not affect the image class. Using augmentation; approximately 2000+ images were generated.

All these images were resized to 224x224 pixels for the VGG16 and VGG19 models and resized to 299x299 pixels for the rest. These preprocessed, augmented images were used to train the pretrained CNN models via transfer learning. The custom layers added on the top of the layers of the pretrained model are trained to fit the data.

4.1.3 Results

The NewHandPD dataset consisted of Spiral as well as Meander images. Separate CNN models were trained on both the trace types. On an average; models trained on the meander trace showed higher accuracy.The table 4.1 shows the accuracy obtained on spiral images. Models were trained with and without augmentation. Similar techniques were adapted for the meander trace. The table 4.2 shows the accuracy obtained on meander images.

From amongst all different model architectures and hyper-parameters, VGG19 model with 15 epochs and augmented data on meander images showed the highest accuracy and low false negatives. VGG19 model consists of a set of 26 layers having different functions such as GlobalAveragePooling, Dropout Layers, Dense layers and so on. We have added custom 3 custom layers which are as below:

- Dense Layer (1024 nodes) - ReLU activation
- Dense Layer (512 nodes) - ReLU activation
- Dense Layer (1 node) - Sigmoid activation

The last layer was the output layer. The score output by the model indicates the likeliness of prevalence of Parkinson's disease in a candidate; by considering only the Trace Test.

For the compiling of the ML model, Adam optimizer was used, and loss was calculated by considering Binary Cross-Entropy Loss function.

Model	Epochs	Augmentation	Training Accuracy	Testing Accuracy
MobileNet	15	Yes	85.88	81.08
Inception	15	Yes	84.73	79.73
Inception	15	No	94	83.23
Xception	15	Yes	92.75	81.08
VGG19	15	Yes	89.93	74.32
VGG19	15	No	98.21	83.01
VGG16	15	Yes	92.75	83.78

Table 4.1: Experimental Results for Trace Test on Spiral Images

Model	Epochs	Augmentation	Training Accuracy	Testing Accuracy
Inception	15	No	78.67	81.13
VGG19	15	Yes	99.18	97.39
VGG16	15	No	98	94.34
ResNet	15	No	98.58	90.57

Table 4.2: Experimental Results for Trace Test on Meander Images

4.2 Speech Test

4.2.1 MPower Dataset Description

There are numerous data-sets available with voice recordings of Parkinson's patients against healthy controls, which were explored in this study. A major issue with most of these data sources is the small sample size. Most data-sets available range from 200 to 800 samples. Therefore, the study worked on the mPower Voice dataset, part of the mPower mobile Parkinson's Disease study which is by far the largest such dataset. The dataset contains more than 65,000 recordings of volunteers, both healthy as well as those suffering from Parkinson's recording the "aaa" sound for 10 seconds. The recordings were accompanied by supplementary data provided by the participants. A lot of this supplementary data is incomplete for numerous columns. The columns identified as being important to the classification include :

- Time of medication : Whether the participant has taken medicines for Parkinson's just before the recording, the just after, at some other time, or not applicable.
- Healthcare-provider : The doctor that the patient is consulting, can be a general physician, a neurologist, PD specialist, or none.
- The year of diagnosis and year of onsent of the disease

- Professional diagnosis : Whether the participant has been officially diagnosed as having Parkinson's disease by a medical professional. It is this column which has been used for labelling the target variable, with patients have a professional diagnosis being in the 'PD' class, with the rest in the control group.

- Other information such as gender, age, race, previous medical history etc,

4.2.2 Processing

Owing to the large number of samples in the dataset and thus its size, approximately 3000 samples were considered for this module. The mPower data set is open for contribution, i.e. anyone can submit a recording from their phone and all information is self verified. This is unlike other datasets in which the patients were recorded in quiet, controlled conditions in laboratories or studios while being monitored by researchers. Due to the self-verified nature of the data, it is likely that some records are corrupted, incorrectly labelled, or have excess background noise.

Since the noise in all recordings is different, a generalized noise classifier is difficult which can filter out background noise. Therefore, the voice samples were manually heard and those having significant disturbance, or blank recordings were removed. After this, it was possible that recordings could be incorrectly labelled. Since the participants in the study had been asked to fill a questionnaire, the details were used to further eliminate ambiguous entries. The work of [19], made a distinction between the symptoms of patients who had taken their medicine just before the recording and those who hadn't. Since it is likely that the symptoms would be diminished in those patients who had the medicine just prior to recording, these records were dropped. After this, the 'healthcare-provider' column was observed. It was found that there were numerous records where the person was visiting a neurologist or PD. specialist, but had not been officially diagnosed. Since it is likely that the person was showing symptoms of PD, but had simply not had an official diagnosis, such ambiguous records were also dropped, to avoid false negatives.

The recordings had to be converted to features for training the classifier. A Praat script was used to extract a total of 26 features from the voice samples. These 26 features have been proven as being successful for recognizing dysphonia and thus Parkinson's in previous studies.[16] The features include the Jitter, Shimmer, Pitch, Degree of Voice breaks and others. Praat is a phonetics software used for audio analysis. praat-parselmouth; a python library for Praat, was used to extract the features from the data samples.

The data obtained after extraction was as follows:

- Column 1: Voice id

- Column 2-26: Features

  - Features 1-5: Jitter (local),Jitter (local, absolute),Jitter (rap),Jitter (ppq5),Jitter (ddp),
  - Features 6-11: Shimmer (local),Shimmer (local, dB),Shimmer (apq3),Shimmer (apq5), Shimmer (apq11),Shimmer (dda),
  - Features 12-14: AC,NTH,HTN,

- Features 15-19: Median pitch,Mean pitch,Standard deviation,Minimum pitch,Maximum pitch,
- Features 20-23: Number of pulses,Number of periods,Mean period,Standard deviation of period,
- Features 24-26: Fraction of locally unvoiced frames,Number of voice breaks,Degree of voice breaks

- Column 27: Class information

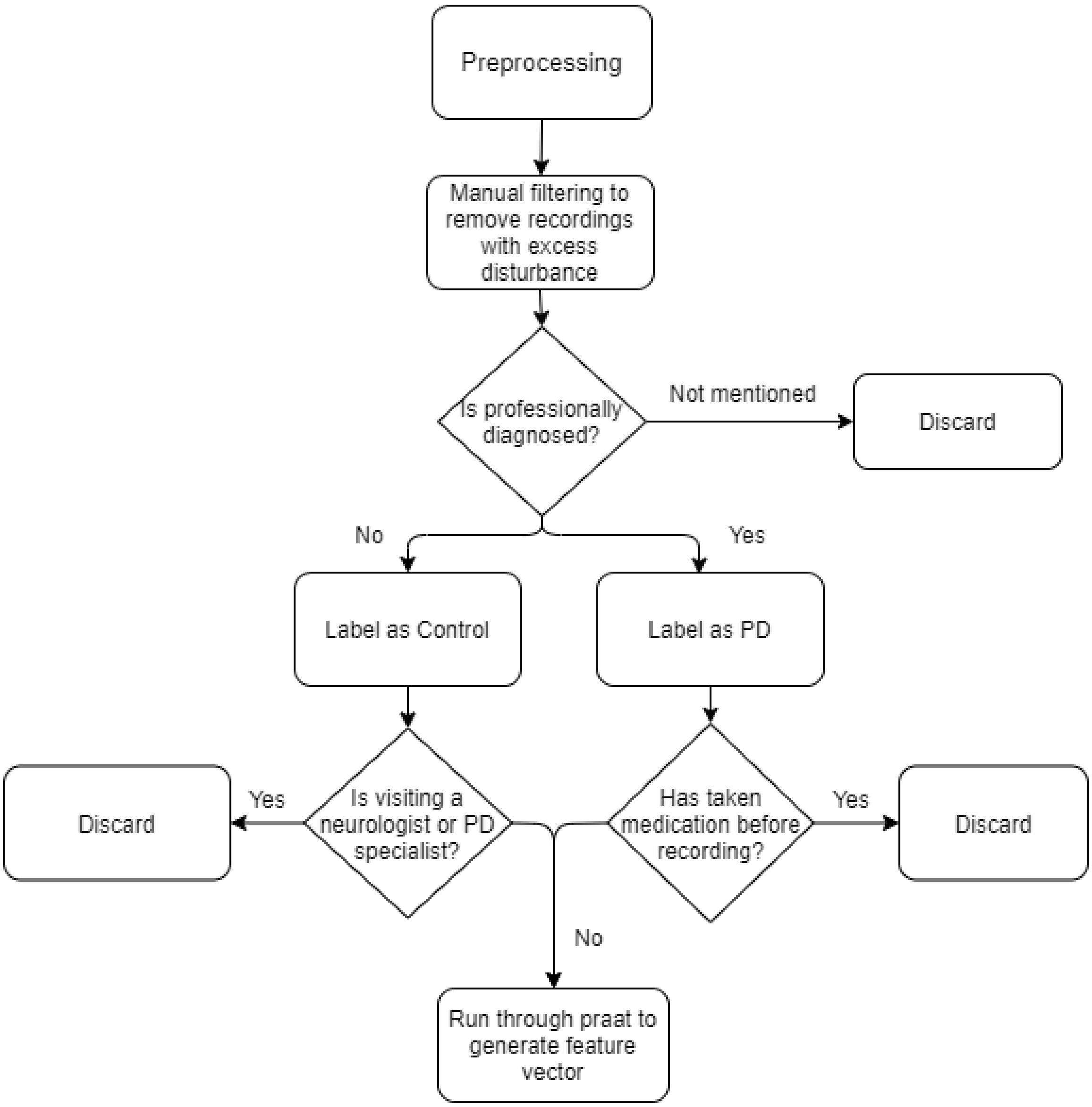


Figure 4.2: Preprocessing steps

4.2.3 Feature Analysis

The feature obtained were analysed to look for feature redundancy, similarity and correlation.

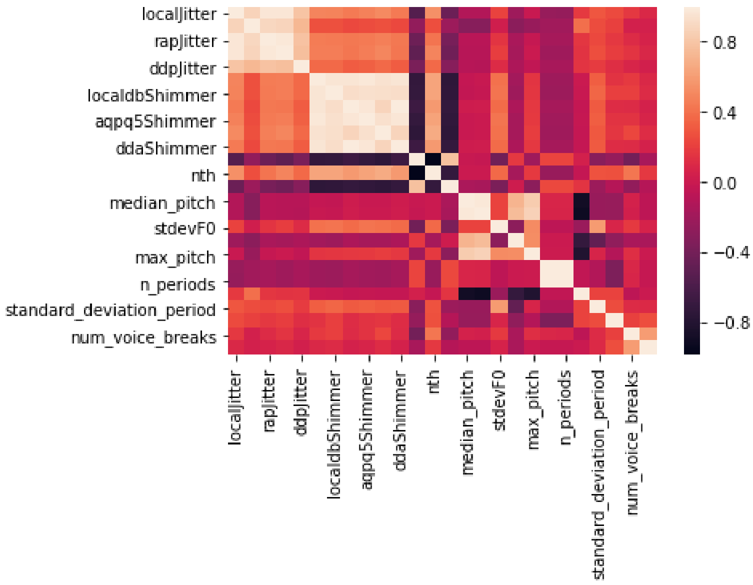


Figure 4.3: Correlation matrix

A correlation graph between the features shows which features are highly related to each other. The four different types of shimmer values measure have significance correlation with one another. Highly correlated features have more or less the same effect on the target.

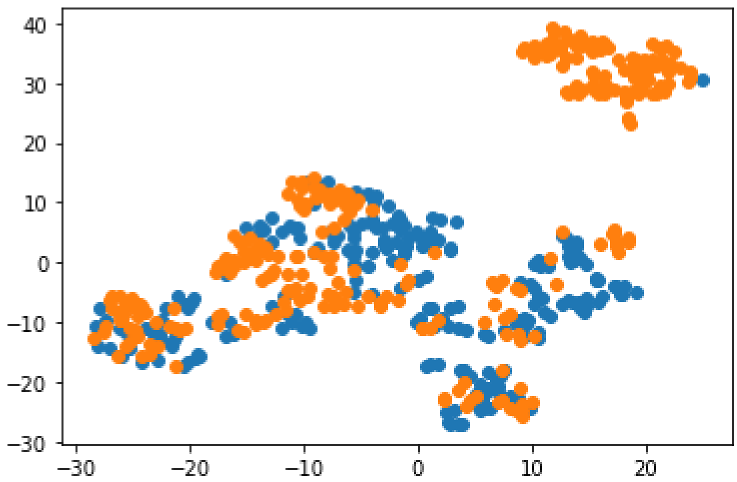


Figure 4.4: 2-D visualization using TSNE

Due to the large number of features, it is not possible to visualize the points as 26-dimensional vectors. Therefore, t-distributed stochastic neighbor embedding (TSNE) was used to map the data from 26 dimensions to low 2-dimensional space, which allows visualization. The plot shows the data points of the same category clustering together.

4.2.4 Smote for imbalance

Since in the pre-processing section, a lot of ambiguous or incorrect entries were removed, the number of final samples dropped to 2000. In these, 500 were records of PD Patients, and 1500 of healthy control. As a result, the dataset was now imbalanced. Therefore, this research implements Synthetic Minority Oversampling Technique, or SMOTE.

SMOTE is used for synthesizing new data points of the minority class from existing data, to overcome class imbalance. SMOTE works by drawing a line closest to the existing samples of the minority data class, and then generating new points around that line.

4.2.5 Experimental results

Refer table 4.3 for recorded accuracies of different machine learning algorithms.

Table 4.3: Experimental Results for Speech Test

Model	**Training Accuracy**	**Testing Accuracy**
Random Forest	100	90
SVM (Linear)	68	62
SVM (RBF)	62	55
SVM (Sigmoid)	61	55
Naive Bayes	61	55
KNN	86	78
Decision Tree	100	79
XGBoost	90	71
LGBM	100	90
CatBoost	99	89

The best performing algorithm for speech test is LGBM classifer, followed by Random forest, both of which give 100% accuracy on the training set and 90% on the test set.

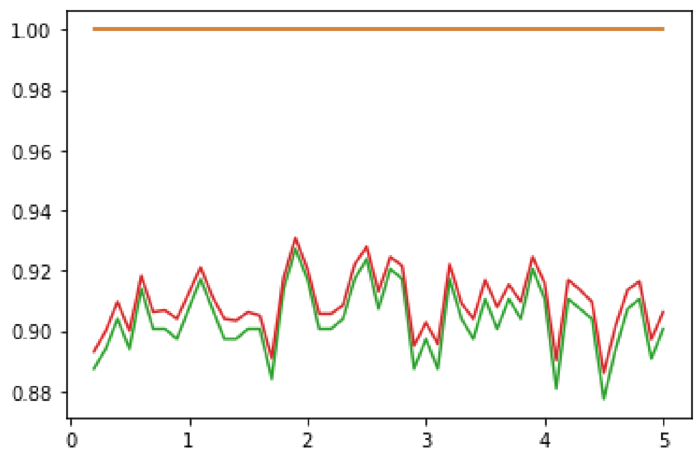


Figure 4.5: Accuracy variance with weights of the two classes

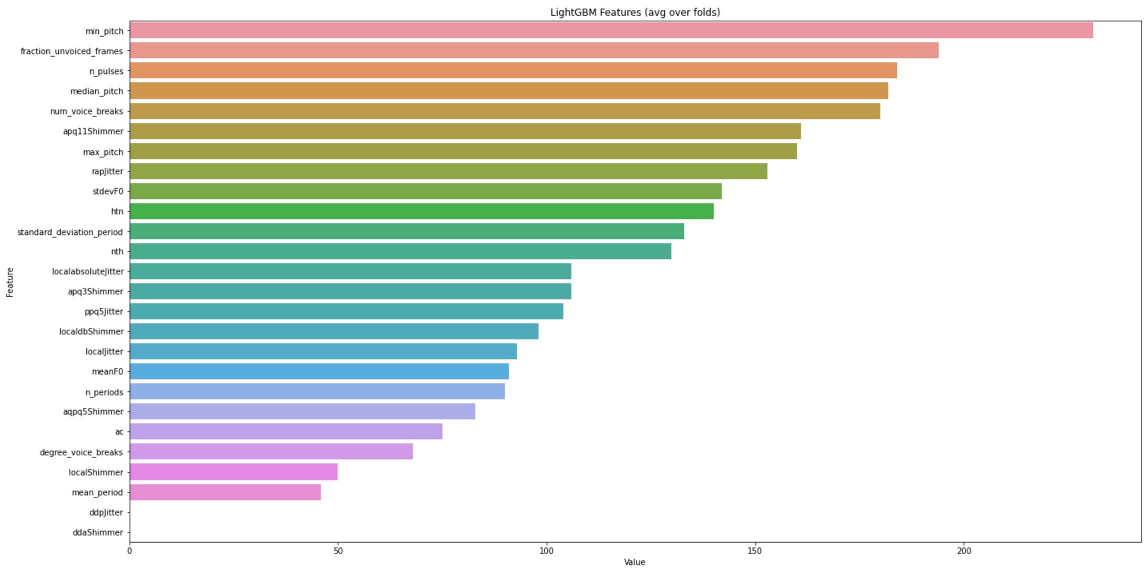


Figure 4.6: Feature Importance for LGBM Model.

4.3 System Architecture

The proposed system is developed for mobile phones using Flutter, which makes the system compatible with Android and iOS based operating systems. This mobile application, as discussed before contains two separate modules for the Trace and the Sound tests. The two tests are to be given in succession. The user input (trace image for trace test and voice recording for speech test) is sent from the mobile application to the server which hosts the machine learning model. The data is first preprocessed, before being input into the models. The two models for tracing and speech will compute scores and which are combined into a weighted average. Weights are assigned according to the accuracy of the model. This final score is output to the user. Higher values indicating higher chances of prevalence of Parkinson's Disease.

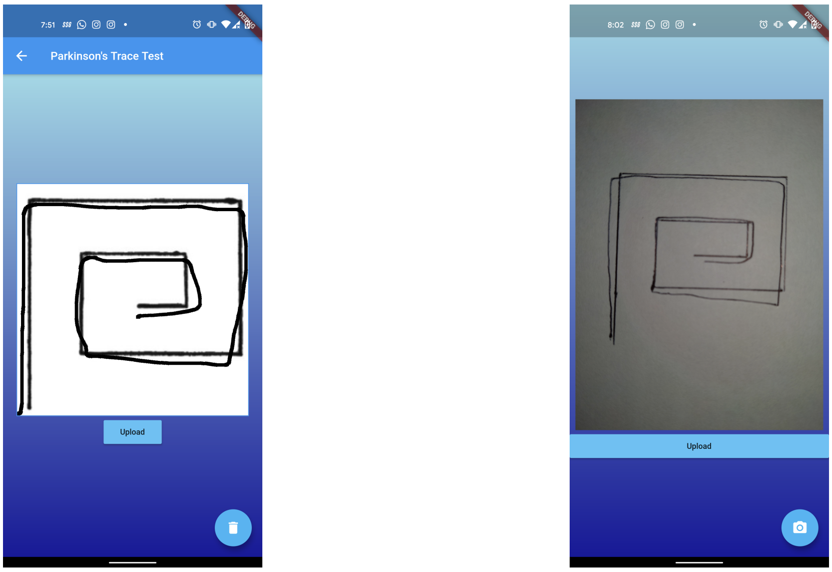
4.3.1 Obtaining data for Trace test

The image can be input in the following two ways :

- The user draws the trace on screen
  Here, the meander trace will be displayed on the mobile application screen. This outline has to be traced using a stylus device pen. The image containing the user trace along with the original outline would be transferred to the server.

- User traces the outline on paper
  In this case, the user trace will be obtained by the using pen and paper. Given a printout of the standard meander outline, the user will then trace the outline. This image would be uploaded to the application.

]

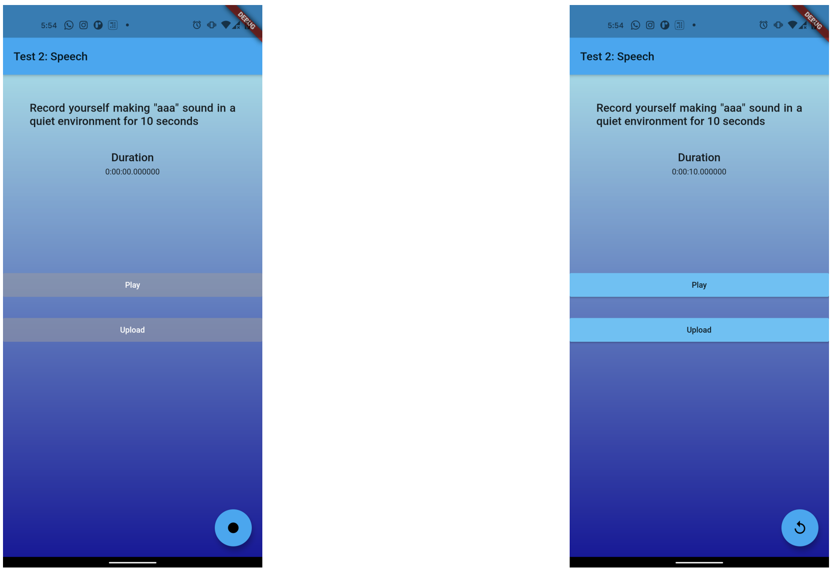


(a) Drawing the trace on the screen (b) Tracing the outline on paper

Figure 4.7: Trace Test

4.3.2 Obtaining data for Speech Test

Voice sample of the user is obtained via the application interface itself. An interactive screen is used to obtain a 10 second recording the user saying the vowel 'aaaa'. The user is expected to record this sound in a quiet environment for accurate predictions.



(a) User supposed to record 'aaa' sound for 10 seconds (b) User can hear the recording before uploading

Figure 4.8: Speech Test

4.3.3 Processing and output

The input data is sent to the back-end hosted on flask. The tracing image is resized and entered into the VGG-transfer learning model, to generate a prediction. For the speech test, the audio file is sent to the Server, where the 26 features of dysphonia are extracted. Then, this feature vector is input the machine learning model which generates another prediction. The scores from the two predictions is combined, and pushed back into the app from the server.

The weighted average for the final output is calculated as follows:

$$O_f = \frac{\sum_{i=1}^{2}(A_i * O_i)}{\sum_{i=1}^{2}(A_i)}$$

where:

- A_i is the accuracy of the i^{th} test
- O_i is the output of the i^{th} test
- O_f is the final weighted Output

The models with higher accuracy are given higher weights. Hence the outputs of the models are multiplied by a weight factor; i.e. the accuracy of that model.

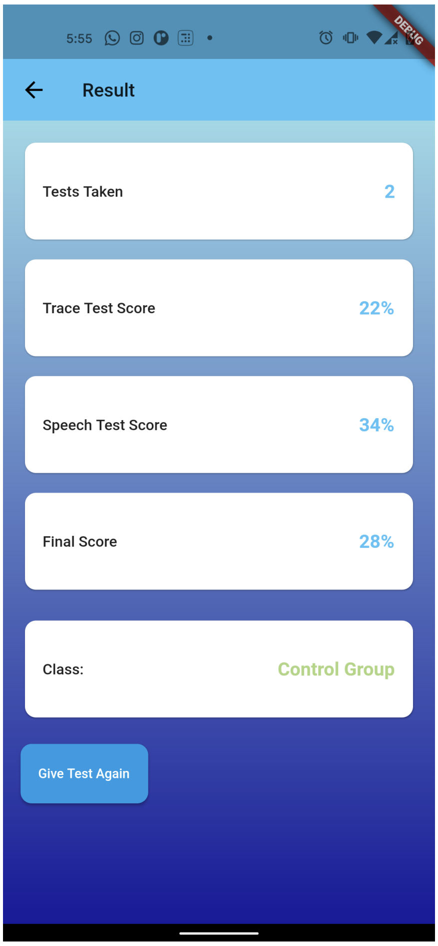


Figure 4.9: Final Result displayed to the user.

Chapter 5

Results and Discussion

The tracing module achieves an optimum accuracy when using the VGG19 architecture.

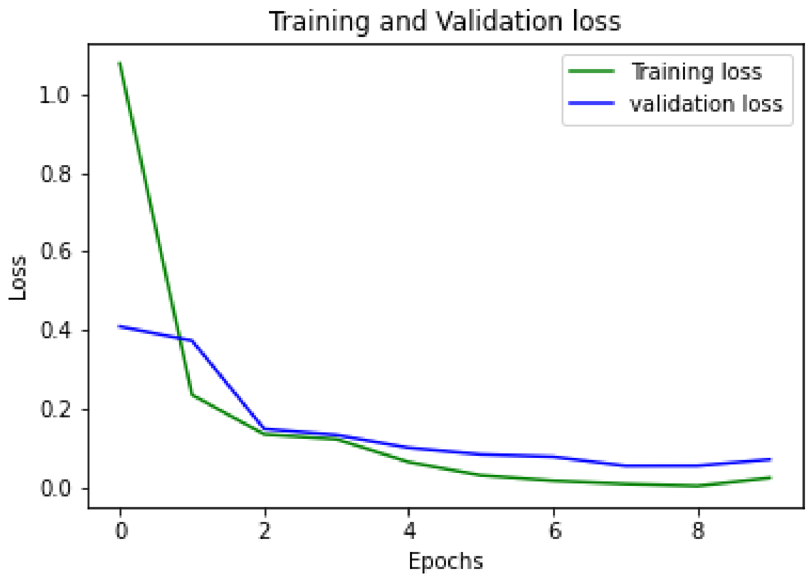


Figure 5.1: Training and Validation Loss for Trace Test

The model achieves an accuracy of 97.39% on the test dataset. Moreover, False Acceptance Ratio (FAR) is defined as the Fraud rate or the probability of a healthy individual being labelled as PD patient. Whereas, False Rejection ratio (FRR) is defined as the the probability of a PD patient being labelled incorrectly as a healthy individual. FAR of 2.41% and FRR of 2.85% is obtained for the Tracing module.

		Predicted		
		PD	Control	Total
Actual	PD	136	4	140
	Control	4	162	166
	Total	140	166	306

The best performing model for the speech test is the LGBM-Classifier, which has an accuracy of 90% on the speech test.

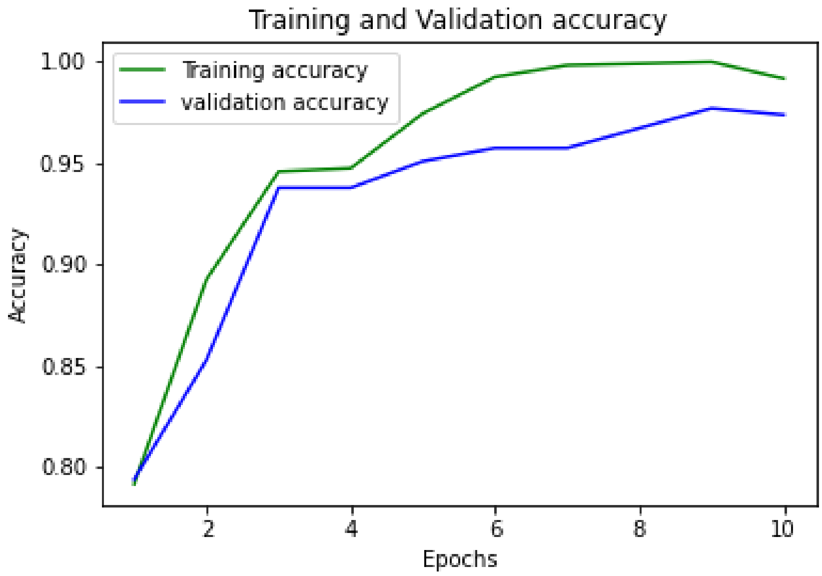


Figure 5.2: Training and Validation Accuracy for Trace Test

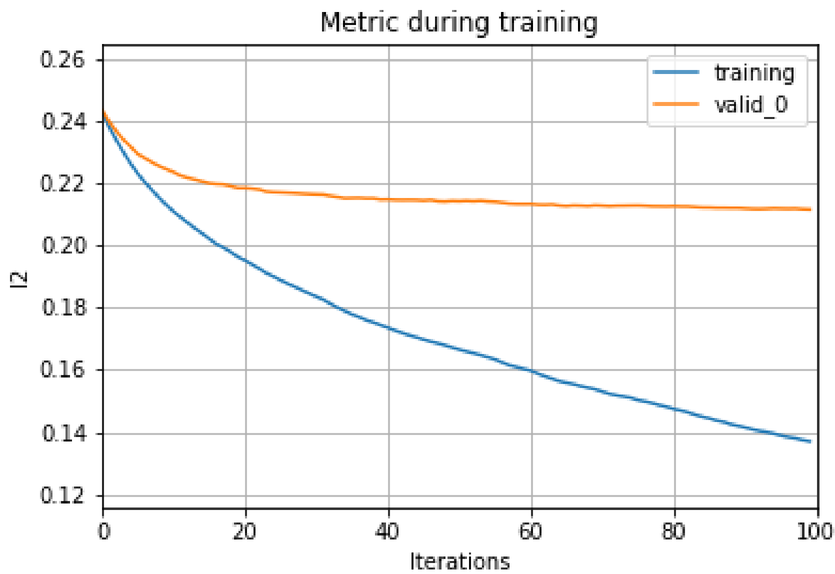


Figure 5.3: L2 Regularization graph for Speech Test.

		Predicted		
		PD	Control	Total
Actual	PD	278	22	300
	Control	39	264	303
	Total	317	286	603

Here the FRR is 6.29% and FAR is 11.29%.

5.1 Limitations

In case of datasets for speech tests; there were linguistic barriers in some cases, whereas limited number of records in others. The mPower dataset on the other hand is a vast collection of records. However, this dataset was a part of a public study; and apart from having possible mislabelled records, it also has noisy audio samples, blanks, and other irrelevant data. This would hamper the model fitting and extensive manual filtering was required. Hence, it is imperative that a larger dataset be developed along with proper monitoring of the data quality.

A lot of research has been done for Parkinson's disease detection using standalone tests. However, the same is not true in case of multiple tests. And as a result; a dataset which stores records for multiple tests for each patient is rare if not unavailable. To the authors' best knowledge, there is no common dataset having records for drawing as well as speech tests. The availability of such a dataset would help in further evaluating this research and many such others.

Chapter 6

Conclusions

Symptoms of Parkinson's being very common tend to go unnoticed till major development takes place. Since there is no cure of PD, detecting it at an early stage is very important. Our research aims to aid in this process. Speech disorders and tremors are the most common symptoms observed among PD patients. Thus, Speech and Tremor tests can help in early detection of the disease. The use of smartphones for accurate PD symptoms analysis and detection will be more accessible to the people hence allowing the widespread use of our application. Multiple tests give better validation and hence can give a better prediction as to whether a person has developed Parkinson's disease or not.

As the research increases in the field of Parkinson's Disease detection, more importance should be given to increasing the domain of the tests and developing a multi-modal approach for the same. This research explored the feasibility of the said approach and implemented the same. This work would help other researchers in exploring along the same lines and thus help develop a larger system, including more tests and better results.

Chapter 7

Future Scope

- We can use the data acquired from the app, after verification, to train the model and increase its efficiency and accuracy. Thus the data acquired from our application can be used for further research and analysis.
- More tests can be added that would check for non motor symptoms which generally develop before motor symptoms.
- Data collected from the app can be used for training the models for both the tests with data from a single patient. This will also help in combined output evaluation.

Chapter 8

Research Publication

- Jinang Gandhi, Aumkar Gadekar, Tania Rajabally, Dr Preetida Jani, Dayanand Ambawade, Detection of Parkinsons Disease via a Multi-Modal Approach, 12th Internation Conference on Computing, Communication and Networking Technologies (ICCCNT), July 2021.

Bibliography

[1] Warwick Adams. High-accuracy detection of early parkinson's disease using multiple characteristics of finger movement while typing. *PLOS ONE*, 12:e0188226, 11 2017.

[2] A. Agarwal, S. Chandrayan, and S. S. Sahu. Prediction of parkinson's disease using speech signal with extreme learning machine. In *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, pages 3776–3779, 2016.

[3] L. Ali, C. Zhu, N. A. Golilarz, A. Javeed, M. Zhou, and Y. Liu. Reliable parkinson's disease detection by analyzing handwritten drawings: Construction of an unbiased cascaded learning system based on feature selection and adaptive boosting model. *IEEE Access*, 7:116480–116489, 2019.

[4] Siddharth Arora, V. Bse, Sean Donohue, Kevin Biglan, E.r Dorsey, and M.A. Little. Detecting and monitoring the symptoms of parkinson's disease using smartphones: A pilot study. *Parkinsonism & Related Disorders*, 21, 03 2015.

[5] Brian Bot, Christine Suver, Elias Chaibub Neto, Michael Kellen, Arno Klein, J Christopher Bare, Megan Doerr, Abhishek Pratap, John Wilbanks, E. Dorsey, Stephen Friend, and Andrew Trister. The mpower study, parkinson disease mobile data collected using researchkit. *Scientific Data*, 3:160011, 03 2016.

[6] S. Chandrayan, A. Agarwal, M. Arif, and S. S. Sahu. Selection of dominant voice features for accurate detection of parkinson's disease. In *2017 Third International Conference on Biosignals, Images and Instrumentation (ICBSII)*, pages 1–4, 2017.

[7] P. Drotár, J. Mekyska, Z. Smékal, I. Rektorová, L. Masarová, and M. Faundez-Zanuy. Prediction potential of different handwriting tasks for diagnosis of parkinson's. In *2013 E-Health and Bioengineering Conference (EHB)*, pages 1–4, 2013.

[8] Giovanni Dimauro; Francesco Girardi. Italian parkinson's voice and speech, 2019.

[9] Muhammed Isenkul, Betul Sakar, and O. Kursun. Improved spiral test using digitized graphics tablet for monitoring parkinson's disease. 05 2014.

[10] P. Khatamino, İ. Cantürk, and L. Özyılmaz. A deep learning-cnn based system for medical diagnosis: An application on parkinson's disease handwriting drawings. In *2018 6th International Conference on Control Engineering Information Technology (CEIT)*, pages 1–6, 2018.

[11] M Little, Patrick Mcsharry, Irene Moroz, and Stephen Roberts. Testing the assumptions of linear prediction analysis in normal vowels. *The Journal of the Acoustical Society of America*, 119:549–58, 02 2006.

[12] Max Little, Patrick Mcsharry, Stephen Roberts, Declan Costello, and Irene Moroz. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *Biomedical engineering online*, 6:23, 02 2007.

[13] M. A. Little *, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig. Suitability of dysphonia measurements for telemonitoring of parkinson's disease. *IEEE Transactions on Biomedical Engineering*, 56(4):1015–1022, 2009.

[14] J. Mucha, J. Mekyska, M. Faundez-Zanuy, K. Lopez-De-Ipina, V. Zvoncak, Z. Galaz, T. Kiska, Z. Smekal, L. Brabenec, and I. Rektorova. Advanced parkinson's disease dysgraphia analysis based on fractional derivatives of online handwriting. In *2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, pages 1–6, 2018.

[15] Amina Naseer, Monail Rani, Saeeda Naz, Muhammad Razzak, and Guandong Xu. Refining parkinson's neurological disorder identification through deep transfer learning. *Neural Computing and Applications*, 00:00, 03 2019.

[16] B. E. Sakar, M. E. Isenkul, C. O. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, and O. Kursun. Collection and analysis of a parkinson speech dataset with multiple types of sound recordings. *IEEE Journal of Biomedical and Health Informatics*, 17(4):828–834, 2013.

[17] S. R. Sonu, V. Prakash, R. Ranjan, and K. Saritha. Prediction of parkinson's disease using data mining. In *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, pages 1082–1085, 2017.

[18] M. Wodzinski, A. Skalski, D. Hemmerling, J. R. Orozco-Arroyave, and E. Nöth. Deep learning approach to parkinson's disease detection using voice recordings and convolutional neural network dedicated to image classification. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 717–720, 2019.

[19] T. J. Wroge, Y. Özkanca, C. Demiroglu, D. Si, D. C. Atkins, and R. H. Ghomi. Parkinson's disease diagnosis using machine learning and voice. In *2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pages 1–7, 2018.

[20] H. Zhang, A. Wang, D. Li, and W. Xu. Deepvoice: A voiceprint-based mobile health framework for parkinson's disease identification. In *2018 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, pages 214–217, 2018.

[21] Hanbin Zhang, Chen Song, Aditya Singh Rathore, Mingchun Huang, Yuan Zhang, and Wenyao Xu. mhealth technologies towards parkinson's disease detection and monitoring in daily life: A comprehensive review. *IEEE reviews in biomedical engineering*, May 2020.