SAMAN JAVED
01-134131-080
HOORAIN ALI
01-134131-033

Early Prediction of Parkinson's Disease through Computerized Analysis of Handwriting

Bachelor of Science in Computer Science

Supervisor: Dr. Imran Siddiqi

Department of Computer Science
Bahria University, Islamabad

November 2016

Certificate

We accept the work contained in the report titled "Early Prediction of Parkinson's Disease through Computerized Analysis of Handwriting", written by Ms. Saman Javed AND Ms. Hoorain Ali as a confirmation to the required standard for the partial fulfillment of the degree of Bachelor of Science in Computer Science.

Approved by:

Supervisor: Dr. Imran Siddiqi (Associate Professor)

____

Internal Examiner: Mr Ghulam Ali Mirza (Assistant Professor)

____

External Examiner: Dr Muhammad Sajjad Khan (Associate Professor)

____

Project Coordinator: Dr . Arif Ur Rehman (Associate Professor)

____

Head of the Department: Dr. Faisal Bashir (Associate Professor)

____

November 21st, 2016

Abstract

Parkinson disease (PD) is a neurological disorder that influences movement of muscles causing disabled posture, rigidity and tremors. People suffering from PD face trouble in sleeping, walking and sitting. Generally, potential PD subjects are examined by an expert medical practitioner who employs the clinical symptoms like stiffness and slowing movements to detect the presence or absence of the disease, only after the disease had progressed considerably. Research on PD has revealed that analysis of handwriting and speech can serve as an effective early warning for Parkinson. With the advancements in image analysis and pattern classification, the manual analysis of these handwritten samples is being replaced with computerized analysis and automated prediction systems. This project presents a system that exploits online features of handwriting to predict PD in subjects. Features considered in our study include writing speed, pen pressure and pen-up/pen-down times. The features from PD patients and control subjects are used to train a support vector machine classifier. Evaluations on a benchmark database of online handwriting samples realized promising classification rates.

Acknowledgments

In the name of ALLAH the most gracious the most merciful. First of all we would like to thank ALLAH (SWT) for his countless blessings. Then we would like to thank our parents and family members who always been encouraging and motivating us to complete our project in a manner that satisfies our academic requirements. They have always inspired us to tackle each and every task with confidence, full devotion and dedication. Furthermore, we are very grateful to our supervisor Dr. Imran Siddiqi who has given us the chance to work on such an amazing project and for providing professional guidance to complete this project in successful manner. Their support, professional guidance and advice throughout the project is highly appreciated. May ALMIGHTY bless them.

SAMAN JAVED AND HORRAIN ALI
Bahria University Islamabad, Pakistan

November 2016

Contents

List of Figures

List of Tables

Acronyms and Abbreviations

PD	Parkinson's Disease
PaHaW	Parkinson's Disease Handwriting Database
SVM	Support Vector Machine

Chapter 1

Introduction

1.1 Overview

Parkinson disease (PD) is a neurological disorder that influences movement of muscles causing disabled posture, rigidity and tremors. People suffering from PD face trouble in sleeping, walking and sitting. In earlier ages, the gadget used for Parkinson diagnosis was to be examined by an expert medical practitioner who used the clinical symptoms like stiffness and slowing movements, only after the disease had progressed considerably. SPECT scans are also employed for detection of PD where a radioactive tracer is infused into the patient to image the brain. In addition to these traditional medical examinations, research on PD has revealed that analysis of handwriting [1, 2, 3] and speech [4] can serve as an effective early warning for Parkinson. Speech changes are indicative of PD as the tongue does not move as far and as quick as it should. This results in subtle changes in the speech patterns and monitoring these changing patterns can be employed to predict PD or the progression of disease over time. Likewise, 'Micrographia' or small handwriting is also known to be an early symptom of PD. This allows analyzing the handwritings of potential subjects to predict the presence or absence of PD .

With the advancements in image analysis and pattern classification, the manual analysis of these handwritten samples is being replaced with computerized analysis and automated prediction systems [5]. In addition to the size of handwritten text, other attributes like pen pressure and pen up/pen down times are also indicative of the disease. Pressure is linked to the grip on the writing instrument while pen-up time reflects the time subject is taking to plan the subsequent writing actions; more time being indicative of reduced cognitive ability. These features, however, require online handwriting samples captured on digital devices capable of recording this information. It should be noted that the objective of such automated systems is to facilitate and not to replace the human experts. These systems could be used as a filter so that only subjects suspected to have PD are referred to the expert practitioner. A number of the research efforts have been carried out at different institutions revealing the major differences of writing among the healthy and diseased persons. Inspired by the strong correlation between handwriting and PD, we intend to develop a computerized system to assist the physicians and neurologist in early prediction of PD through analysis of online handwriting. The key objectives of the proposed study are presented in the following section.

1.2 Objectives

The goal of this project is to automate the early detection of Parkinson disease using online handwriting analysis through pattern classification techniques. This will assist professionals in diagnosing PD at early stages by a simple and objective tool and will ultimately lead to the better handling of the disease.

1.3 Problem Description

The proposed project aims at developing a tool allowing computational analysis of online handwriting as an early warning of PD. This early prediction of disease could lead to treatment sessions instead of wasting time in detection. Early detection of PD has become the need of age as in current era no such system exists for early diagnosis and disease can only be entertained when it is in later stages . We aspire to develop a system that will require subjects to copy given templates on a digitizing tablet. Features extracted from these online writing samples will be employed to predict the presence of PD.

1.4 Proposed Methodology

The proposed methodology primarily rests on image analysis and pattern classification techniques to classify a given online handwriting sample into one of the two classes, PD and healthy. As discussed earlier, one of the most prominent indicatives is micrographia where the writing appears cramped and the writing size tends to be smaller than normal. In most cases, the subject starts writing in regular size which gradually decreases as the writing progresses. The progression of size can be used as a feature to test for micrographia.

Generally, in addition to copying the textual templates, subjects in handwriting based PD test are also required to draw a spiral. While the normal person is likely to produce a smooth spiral, the one drawn by a Parkinson patient normally is jerky reflecting the hand tremor. Analysing the contours of these drawings, therefore, can serve as a useful feature.

In addition to the size and smoothness of drawn shapes, online information of handwriting provides useful insights in PD investigations. Features like pen pressure and pen-up time will be considered in our study. The extracted features from samples of patients and control subjects will be fed to a learning algorithm to learn to discriminate between the two classes. During classification, features extracted from the queried subject will be fed to learned classifier to assign the output label, healthy or potential PD patient. Classifiers like Support Vector Machine (SVM) and Artificial Neural Network (ANN) will be explored for this purpose. An overview of the proposed scheme is summarized in Figure 1.1.

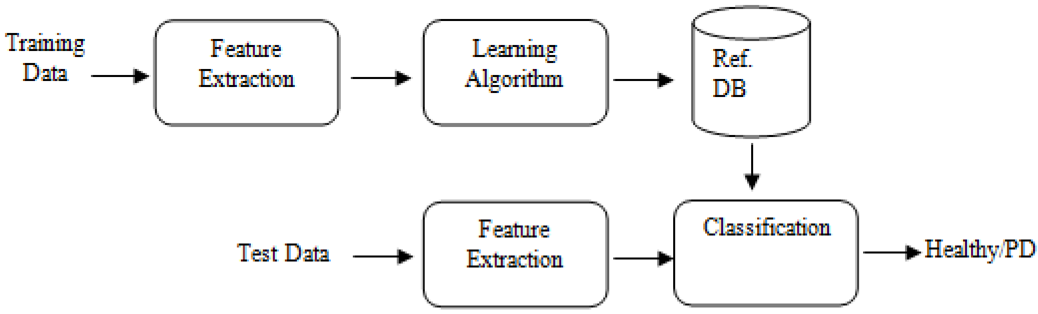

Figure 1.1: Overview of the proposed system

The system will be trained and evaluated on the Parkinson's Disease Handwriting Database (PaHaW) [6, 7] The database has been acquired in licensed agreement with the Movement Disorders Center at the Department of Neurology, Masaryk University and St. Anne's University Hospital in Brno, Czech Republic. The data base comprises online handwriting samples with pressure and pen-up/pen-down information collected from 37 patients and 38 control subjects. Each group has approximately same distribution of male and female subjects. The content of the template to be copied comprises a spiral and eight lines of text as illustrated in Figure 1.2.

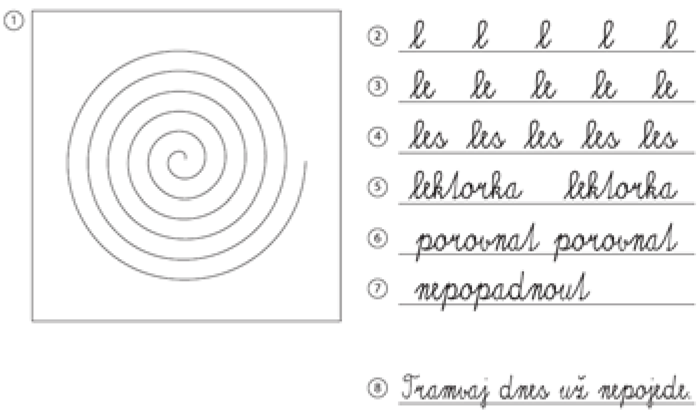

Figure 1.2: Template of the PA Haw database copied by the subjects

Like other studies in the literature, we will employ 70 percent of data for training and 30 percent for evaluation. In addition to the PaHaW database, we also intend to develop a module to capture online handwriting and predict the PD for new subjects. The data will be captured using the Wacom Intuos Pro digitizer shown in Figure 1.3. The SDK of the digitizer will be employed to capture the pen trajectory, pressure and pen-up, pen-down states. The recorded data will then be classified using the already trained classifier.

Figure 1.3: Wacom digitizer to capture online handwriting

1.5 Project Scope

The developed system will rely on online data for prediction of PD. The system will be mainly trained and evaluated on the PaHaW database for which the ground truth information is available. For live sessions with subjects, each subject will be asked to copy the given template including the Archimedean spiral and text with loops. The Wacom tablet captures the dynamic features including pen trajectory (x,y coordinates), the time taken, how much pen was tilted, what was the elevation level, Button status which is binary value 0 for in air pen movement and 1 for on surface movement.

1.6 Feasibility Study

The biggest challenge in the project is availability of an appropriate database. While collection of samples from control (healthy) subjects is straight forward, collecting data from Parkinson patients is challenging. Fortunately, the database of writing samples from Parkinson patients has already been acquired. Another potential issue could be integration of the Wacom digitizer SDK in the system to allow capturing online handwriting for new subjects.

1.7 Solution Application Areas

The intended automated system will facilitate the practitioners in an efficient decision making of Parkinson patients. Non-medical personnel can be trained to use the system and provide details to the experts for further assessment enabling them to keep their focus on the patients and treatment instead of taking time for detection and diagnosis.

1.8 Organization of Document

This report is organized as follows. The next chapter will encompass significant contributions towards the automation of early prediction of Parkinson Disease. Chapter 3 will discuss requirement specifications of our project followed by system design in Chapter 4. Implementation of the system is presented in Chapter 5 followed by the experimental results in Chapter 6. Finally, Chapter 7 concludes the document.

Chapter 2

Literature Review

Parkinson Disease is a neural defect that affects the functionlity of brain which influences the movement of muscles, speech and causes disabled posture, rigidity and tremors. PD is found in a significant portion of world's population and impacts on approximately 1 percent of those over 50 years of age. PD is a progressive neurogenerative disease as it is found more in men as compared to women in late ages. PD causes partial or complete passivation of motor reflexes, speech and posture [5]. It is usually caused of the death of neurons thus the brain messages are not transmitted properly resulting in slowed moments and rigidity. Approximately 90 percent of people with Parkinson have speech disorders. The disease can not be diagnosed at the very start as it shows resemblance to other medical diseases. Clinically the disease is detected through the symptoms of bradykinesia (slowing of physical movement), rigidity, tremors or postural imbalance. Voice data mining is considered to be a good methodology for the disease detection as well [8].

Many researches have concluded that handwriting can be used as a significant tool for the disease diagonsis as it is a unique personal feature. Our system aims to work on the most prominent feature of handwriting called 'Micrographia'. It is symptomized with the minimzed size of handwriting. Studies also show that measurement of kinematic aspects of movement like speed, acceleration and stroke duration are helpful in disease prediction. Our system records the handwriting of the patient on a tablet, a spiral and seven lines of text involving loops. These loops and spiral are slightly deformed by patients with tremor. The deformation can be used as a feature for diagnostics [9].

A study carried out at the University of Haifa, Israel [10], conforms o the idea proposed that handwriting can be used as powerful tool for the disease diagnosis .The study carried out involved 40 subjects out of which 20 were diseased and 20 were healthy. Subjects were required to pen down their address with a pressure sensitive pen on a paper attached to Wacom digitizer. For every task the mean pressure and velocity was recorded. Spatial and temporal characteristics were measured for every stroke. While automating the results some features of handwriting were considered such as (length, height and width of handwriting). Following conclusions were drawn from this study.

- Subjects suffering from Parkinson exerted more pressure as that of healthy ones.
- Crumpled hand writing was spotted of Parkinson's subjects.
- PD patients took more than usual time for writing.

Authors reported a correct identification rate of 97.5 percent with the studied features.

In another study [6], the authors exploit features like on-surface movement, in-air movement and pressure for prediction of PD from online handwritings. Extracted sample features were used to train the Support vector Machine(SVM). The study realized a correct classification rate of 89 percent on a database comprising writing samples of 37 patients and 38 control subjects. The same study was extended [7] to compare the performance of three classifiers, KNN, ensemble Adaboost and SVM. Among these, SVM reported the highest classification rates. Likewise, authors in employed a Biosensor smart pen (BiSP) to capture handwriting samples of 48 healthy adults and 38 PD patients sharing the same demographics. Features like pen pressure, finger pressure on pen tip and acceleration etc. were extracted. Using 70 percent of writings in training and 30 percent in test with an SVM classifier realized correct classification rate of more than 92 percent.

Figure 2.1 and Figure 2.2 show sample handwritings of in air movement and on surface movement of PD patient and healthy control subject respectively [11]. The sample handwriting of PD patient is jerky and deformed while the healthy person tends to write smoothly.

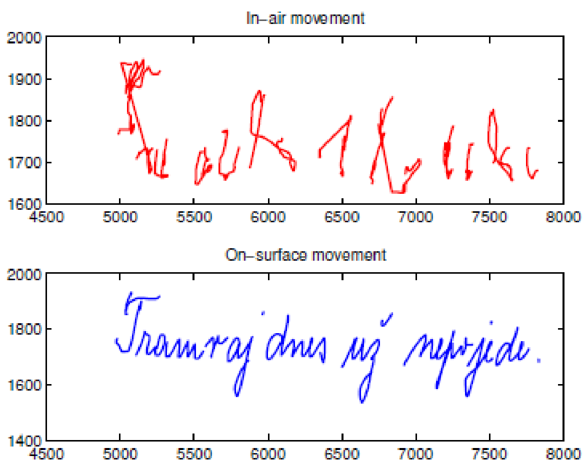

Figure 2.1: Sample of handwriting of PD Patient [11]

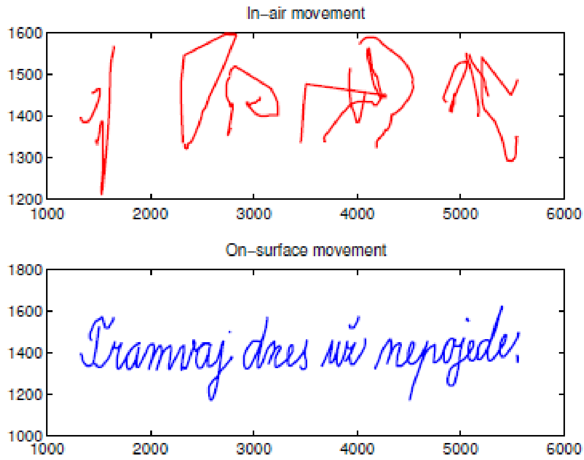

Figure 2.2: Sample of Handwriting of Control Subject
[11]

In [7], Wacom digitizer is employed for data acquisition. The Wacom digitizer is capable of recording several dynamic features like pressure exerted by the pen, the tilt, elevation level, as well as recording the movements represented by the x,y coordinates which tell the button status (binary 0/1) where 0 represents when the pen is above the surface and 1 represents pen is touching the surface. The coordinates help us in evaluation of several kinematic features such as height, velocity acceleration, stroke length, pressure exerted per unit time etc. The kinematic features used are listed in Figure 2.3.

Feature	(s)/(v)	Description
stroke speed	v	trajectory during stroke divided by stroke duration
speed	s	trajectory during handwriting divided by handwriting duration
velocity	v	rate at which the position of a pen changes with time
acceleration	v	rate at which the velocity of a pen changes with time
jerk	v	rate at which the acceleration of a pen changes with time
horizontal velocity/acceleration/jerk	v	velocity/acceleration/jerk in horizontal direction
vertical velocity/acceleration/jerk	v	velocity/acceleration/jerk in vertical direction
number of changes in velocity direction (NCV)	s	the mean number of local extrema of velocity [15]
number of changes in acceleration direction (NCA)	s	the mean number of local extrema of acceleration [15]
relative NCV	s	NCV relative to writing duration
relative NCA	s	NCA relative to writing duration
in-air time	s	time spent in-air during writing
on-surface time	s	time spent on-surface during writing
normalised in-air time	s	time spent in-air during writing normalised by whole writing duration
normalised on-surface time	s	time spent on-surface during writing normalised by whole writing duration
in-air/on-surface ration	s	ratio of time spent in-air/on-surface

Figure 2.3: Kinematic Handwriting Features
[7]

The features are normalized before classification to have zero mean and a standard deviation of one. Different kinematic and pressure features with largest relevance to class

label are sorted. Displayed are medians and standard deviations (std) for healthy controls and Parkinson's disease group (Figure 2.4) [7].
Classification was carried out using a support vector machine and the realized results are summarized in Figure 2.5. To obtain more accurate results, authors also compared different classifiers have been compared like the AdaBoost classfier and K-NN classifier [7]. The kinematic and pressure features are shown in Figure 2.6 and Figure 2.7 respectively. The realized results validated the idea that handwriting is a effective tool for diagnosis of PD providing classification performance of around 80 percent [11].

feature, stat. functional, task number	$\lvert\rho\rvert$	PD median (std)	H median (std)
stroke speed, std, task 8	0.39	0.45 (0.88)	-0.47 (0.97)
relative NCP, task 8	0.37	-0.22 (0.63)	-0.06 (1.16)
horizontal velocity, std , task 8	0.35	0.20 (0.99)	-0.33 (0.89)
ρ_{vel}, 99th percentile, task 2	0.35	-0.57 (0.86)	0.37(1.02)
horizontal velocity, 99th percentile, task 8	0.33	0.4 (0.95)	-0.44 (0.95)
R^{fall}_{time}, median, task 8	0.33	-0.42 (0.57)	0.11(1.22)
$\rho^{rise}_{horizontal,acc}$, 1st percentile , task 8	0.33	-0.17 (0.83)	-0.37 (1.06)
relative NCP, –, task 6	0.33	-0.40 (0.56)	-0.1 (1.23)
R^{rise}_{press}, 99th percentile - 1st percentile, task 3	0.33	-0.33(0.61)	0.1 (1.2)
$\rho^{main}_{vertical,vel}$, 99th percentile - 1st percentile , task 2	0.32	-0.47 (0.75)	-0.11(1.12)
R^{rise}_{press}, std, task 3	0.32	-0.29(0.6)	0.06 (1.21)
horizontal jerk, std, task 8	0.32	0.23 (1.0)	-0.41 (0.9)
horizontal velocity , 99th percentile 99 - 1st percentile , task 8	0.32	0.22 (1.01)	-0.48 (0.89)
horizontal jerk, 99th percentile , task 8	0.32	0.29 (0.99)	-0.25 (0.91)
R^{fall}_{time} ,median , task 3	0.31	-0.17(0.65)	-0.1 (1.19)
velocity, median, task 8	0.31	0.2(0.9)	-0.35 (1.0)
horizontal velocity(rising edge) , std, task 8		0.26 (0.98)	-0.44(0.94)
horizontal jerk, percentile 99th - percentile 1st, task 8	0.31	0.26 (0.97)	-0.33 (0.95)
$\rho_{vertical,vel}$, std, task 3	0.3	-0.6 (0.93)	0.23 (0.99)
velocity (rising edge), mean, task 8	0.3	0.16 (0.98)	-0.41 (0.93)

Figure 2.4: Description of calculated features
[7]

Evaluated task	P_{acc} pressure features	P_{acc} kinematic features	P_{acc} kinematic and pressure
1 (Archimedean spiral)	62.8	-	62.8
2 (letter l)	72.1	69.2	72.3
3 (bigram le)	71.5	72.5	71.0
4 (word les)	66.4	-	66.4
5 (word lektorka)	66.9	65.1	65.2
6 (word porovnat)	74.2	64.9	73.3
7 (word nepopadnout)	66.8	66.4	67.6
8 (sentence)	73.2	74.9	76.5
overall	82.5	75.4	81.3

Figure 2.5: Classification accuracies of different handwriting tasks for kinematic and pressure features using SVM

[7]

classifier	P_{acc} [%]	P_{spe} [%]	P_{sen}[%]
SVM	81.3	80.9	87.4
AdaBoost	78.9	79.2	82.4
K-NN	71.7	70.8	78.5

Figure 2.6: Comparison of different classifiers for Kinematic and pressure features

[7]

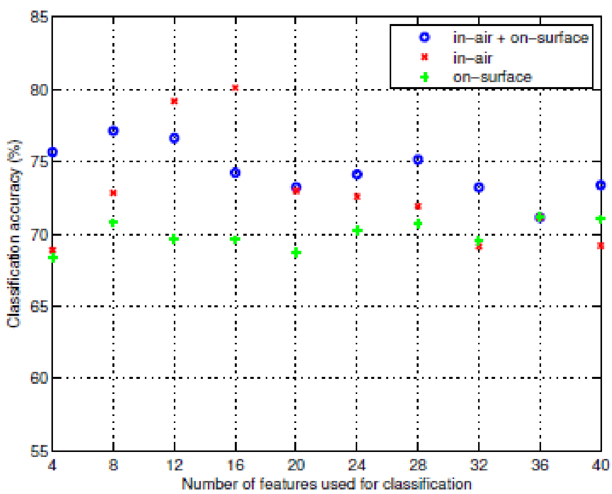

Figure 2.7: Classification accuracy of SVM for different modalities

[11]

After having presented the notable contributions to prediction of PD from handwriting, we discuss the requirement specifications in the next chapter.

Chapter 3

Requirement Specifications

The proposed project aims to provide an automated tool for the early prediction of Parkinson disease in elderly people through computerized analysis of handwriting. User interacting with the system is required to write down few sentences, As discussed earlier one of the most prominent indicatives is micrographia where the writing appears cramped and the writing size tends to be smaller than normal. In most cases, the subject starts writing in regular size which gradually decreases as the writing progresses (Figure 3.1). The size progression of size can be used as a feature to test for micrographia.

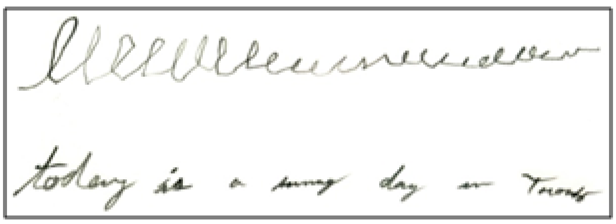

Figure 3.1: Handwriting of a Parkinson patient - Writing size begins as normal and gradually reduces

Generally, In addition to copying the textual templates, subjects in handwriting based PD test are also required to draw a spiral. While the normal person is likely to produce a smooth spiral, the one drawn by a Parkinson patient normally is jerky reflecting the hand tremor (Figure 3.2). Analysing the contours of these drawings, therefore, can serve as a useful feature.

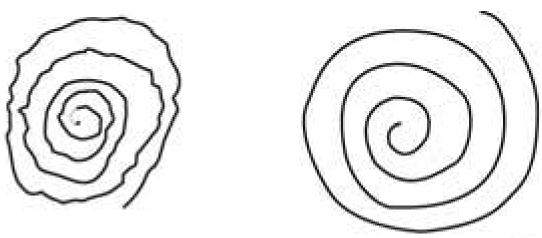

Figure 3.2: PD patient - Normal person

The proposed project is a desktop application. The purpose of the proposed project is to develop an efficient, reliable and accurate diagnostic tool which will reduce the work load of experts by automating the test. The intended automated system will help the practitioners in an efficient analysis of Parkinson patients. Non-medical personnel can be trained to use the system and provide details to the experts for further assessment enabling them to keep their focus on the patients and treatment instead of taking time for detection and diagnosis.

3.0.1 Limitations

The developed system will rely on online data for prediction of PD. The system will be mainly trained and evaluated on the PaHaW database for which the ground truth information is available. For live sessions with subjects, each subject will be asked to copy the given template including the Archimedean spiral and text with loops. The Wacom tablet captures the dynamic features including pen trajectory (x,y coordinates), the time taken, how much pen was tilted, what was the elevation level, Button status which is binary value 0 for in air pen movement and 1 for on surface movement.

3.1 Requirement Specifications

Requirement specifications are presented in terms of functional and non-functional requirements.

3.1.1 Functional Requirements

The functional requirements of the system are as follows.

3.1.1.1 User Requirements: Expert

Following will be the functionality provided by the system to the expert.

- Start the application.
- Train the classifier.
- Connect the Device(Wacom)to the laptop and start taking readings.
- Ask the User to draw the Archimedian spiral and text with loops.
- Classify the person as diseased or healthy.
- Close the application.

3.1.1.2 User Requirements: Patients

Following will be the functionality provided by the system to the patient.

- Start the application.
- Draw the respective Archimedian spiral and text with loops as displayed on the screen.
- View the classification result.
- Close the application.

3.1.2 Non-Functional Requirements

Following will be the non-functional requirements provided by the system.

- **Performance:** The system should be efficient and able to classify the person as a diseased or healthy without a noticeable delay.
- **Reliability:** The application should reliably classify and recognize the Archimedian spiral and the text with loops.
- **Security:** The system does not have any specific security requirements.
- **Availability:** The system does not require any Internet connectivity and will be available once the application has been installed on a device.
- **Maintenance:** The maintenance of the application will be carried out by the developers, if required.

3.1.3 User Categories

There are two types of users involved in the project illustrative in Figure 3.3.These include patient and the expert.

3.1.4 Use Cases

Use cases are a powerful tool for describing scope and interaction of the users with the system. The use cases in the developed system are presented in the following.

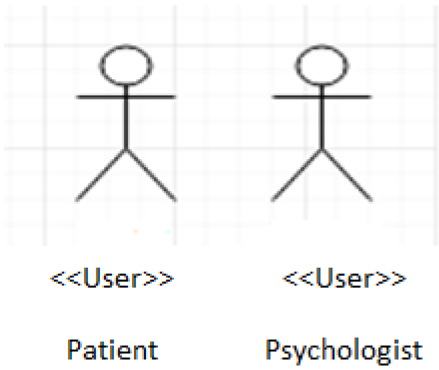

Figure 3.3: User Categories

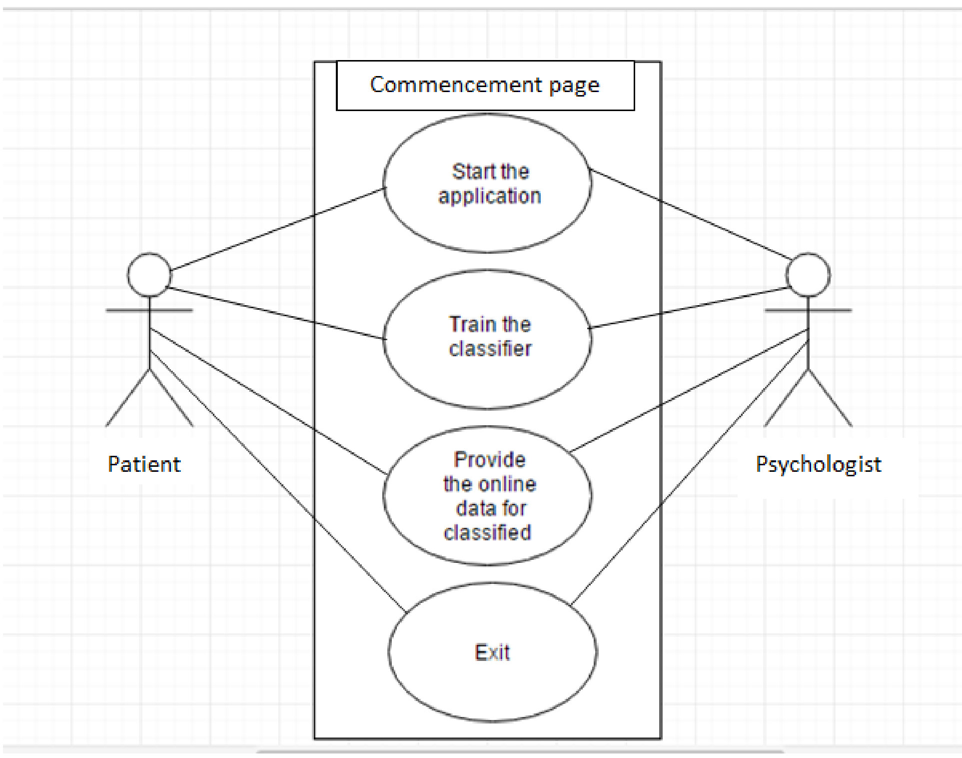

Figure 3.4: Use case at the system Commencement page

3.1.4.1 UC1: Commencement Page

Table 3.1: Commencement page

Title	Commencement page
Actors	Patient and Psychologist
Roles	Patient and Psychologist must Start the application and can proceed to training the data set or exit
Description	Patient can start the test
Pre-conditions	System should be running
Post-conditions	Starting the application should effectively lead to training session

3.1.5 UC2: Training Data-set

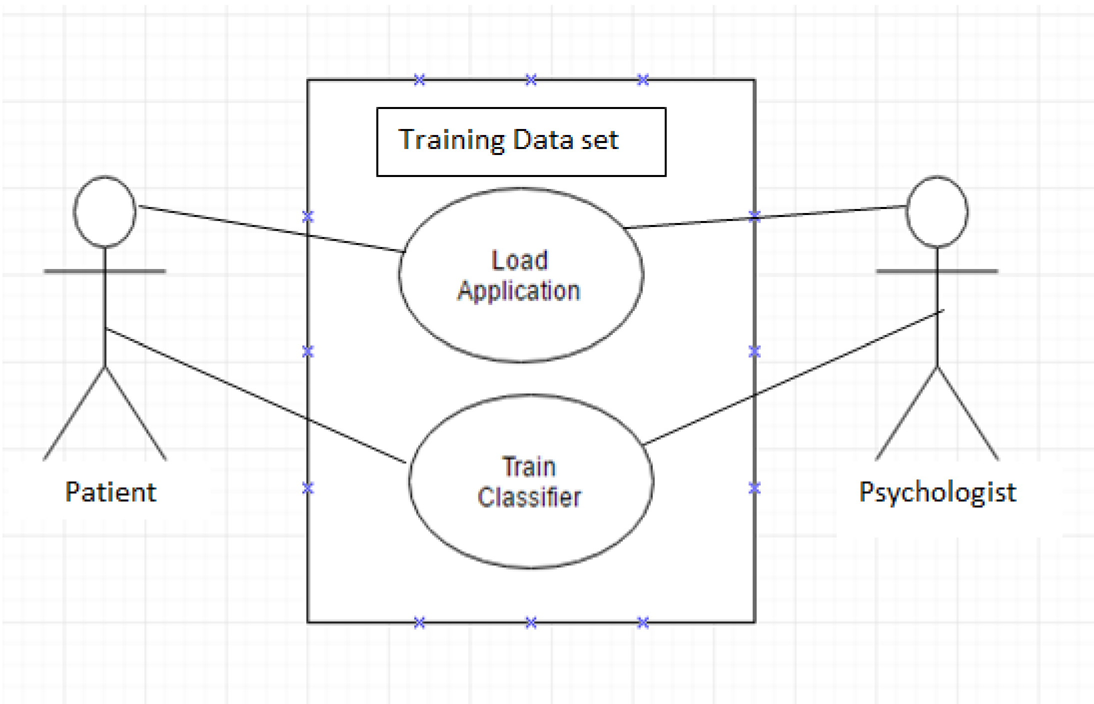

Figure 3.5: Training Data-set

Table 3.2: Training Data-set

Title	Training Data-set
Actors	Patient and Psychologist
Roles	Patient and Psychologist must load the application and train the classifier.
Description	System must start performing the training
Pre-conditions	System should be running
Post-conditions	Subject can perform the training and then provide offline or online data for classification

3.1.6 UC3: Offline Classification

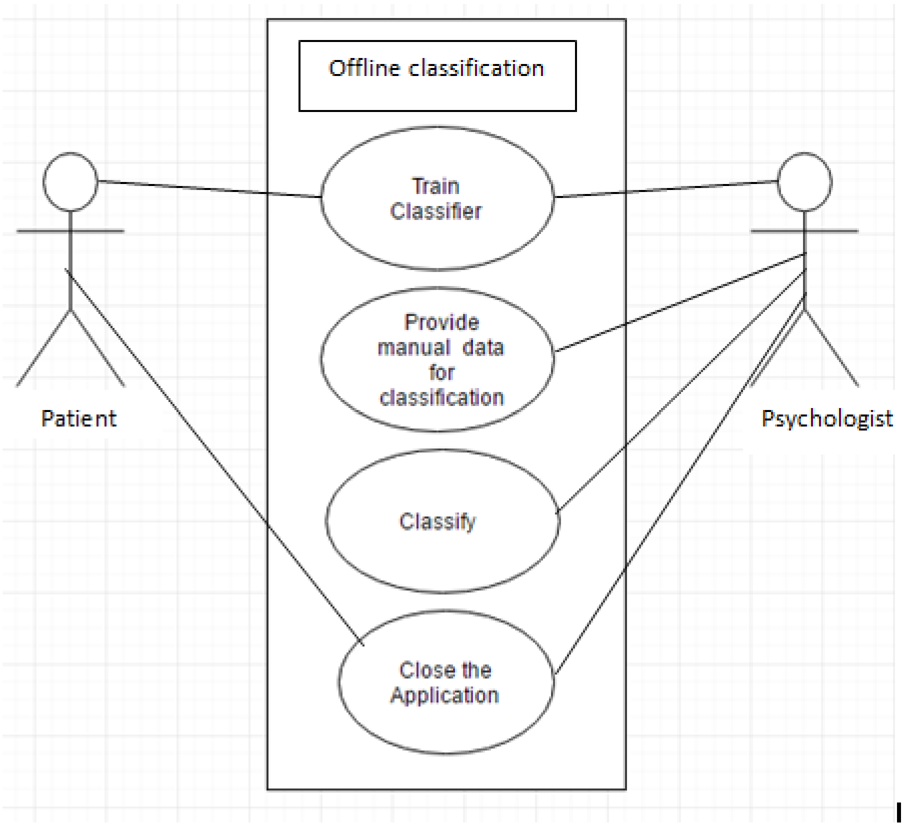

Figure 3.6: Offline Classification

Table 3.3: Offline Data Classification

Title	Offline Data Classification
Actors	Psychologist
Roles	Psychologist should be able to train classifier and then provide manual data for the classification
Description	Application must allow for manual input/offline input of data
Pre-conditions	System should be well trained
Post-conditions	Offline data acceptance and required classification

3.1.6.1 UC4: Online Classification

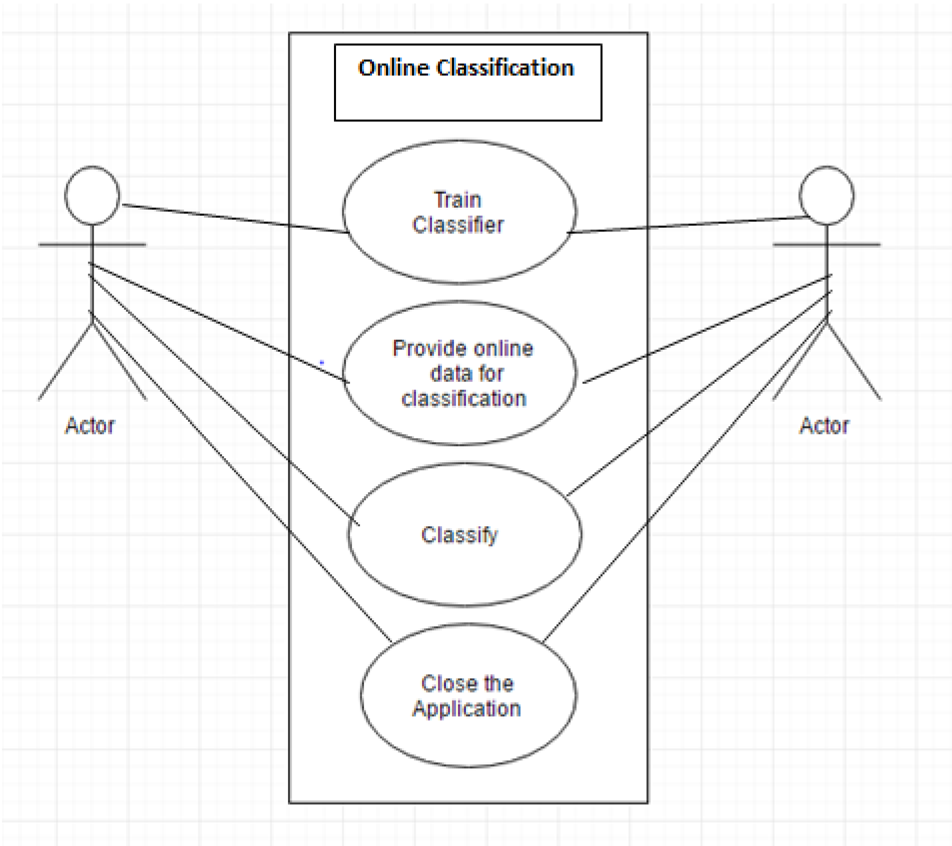

Figure 3.7: Online Classification

Table 3.4: Online Classification

Title	Online Classification
Actors	Patient and psychologist
Roles	Patient provides online data for classification while the psychologist views the classification result and works on further disease cure
Description	This use case represents online data classification
Pre-conditions	System should be running, well trained and allows for online data classification .
Post-conditions	Classification of a subject as diseased or healthy to be displayed

Chapter 4

Design

4.1 System Architecture

In our system, online handwriting data will be input to the system. The system is trained and evaluated on the Parkinson Disease Handwriting Database (PaHaW). Features extracted from these online writing samples are employed to predict the presence of PD. The system uses SVM in order to predict PD based on extracted features of handwriting. The application uses Accord.net library for extracting features from the samples of patients and control subjects which are fed to SVM classifier to discriminate between two classes PD and healthy. In addition to the PaHaW database, our system has a module to capture online handwriting and predict the PD for new subjects. The data will be captured using the Wacom Intuos Pro digitizer. Data captured from Wacom Intuos Pro digitizer will used as test data. The overall system architecture of the application is described in Figure 4.1.

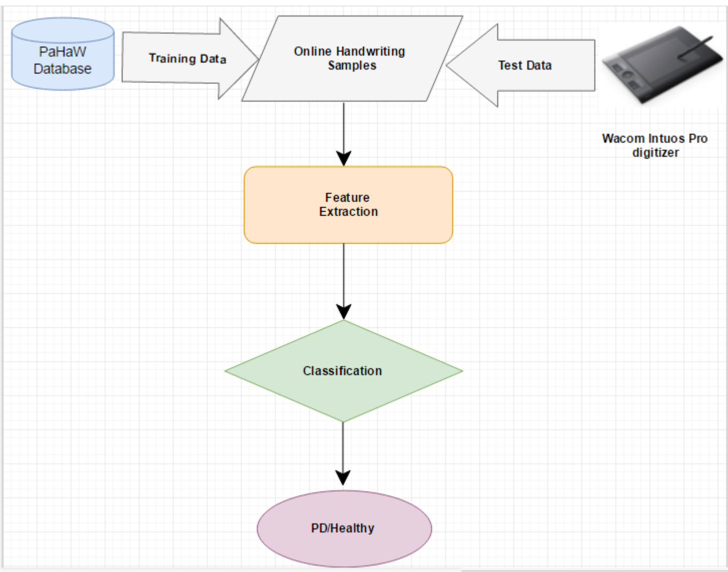

Figure 4.1: Architecture of overall system

4.2 Design Constraints

4.2.1 Software Requirements

- A computer with Windows Operating system
- Wacom Intuos Pro digitizer Software

4.2.2 Development Environment Requirements

- Microsoft Visual Studio 2013
- Software Development Kit Wacom Intuos Pro digitizer
- Accord.net Library

4.2.3 Programmimg Language

- C# Programming Language

4.3 Design Methodology

An overview of the intended system with major modules is presented in Figure 4.2. The design methodology primarily rests on image analysis and pattern classification techniques to classify a given online handwriting sample into one of the two classes, PD and healthy. The system will be trained and evaluated on the Parkinson Disease Handwriting Database (PaHaW). At first we will use 70 percent of data for training and 30 percent for evaluation. Later, the system will be tested with new samples obtained from the Wacom Digitizer.

The first module is to get training data from PaHaW database. Next, in order to train the classifier we employ features like stroke speed, acceleration, pressure etc. There is a prominent difference in the features of a PD Patient and the healthy person. The writing of PD patient appears cramped and the writing size tends to be smaller than normal. While the normal person is likely to produce a smooth spiral, the one drawn by a Parkinson patient normally is jerky reflecting the hand tremor. The features are fed to the SVM classifier and classifier is trained based on these features. Once the classifier is trained, test data is given to the system that is the remaining 30 percent in the database. During classification, features of the queried subject are fed to the classifier to assign the output label, healthy or potential PD patient.

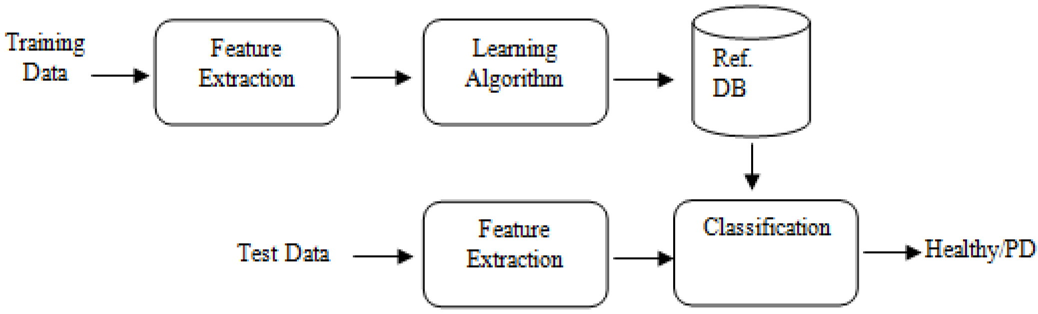

Figure 4.2: Methodology

4.3.1 PaHaW Database

Our system is using Parkinson Disease Handwriting Database (PaHaW). The PaHaW database contains handwriting samples from 37 Parkinson's patients' 19men/18 women and 38 healthy controls 20 men/18 women. The database comprises of folders of patient and healthy controls. Each folder of the subject contains eight files of svc file structure. The eight files represent the eight tasks performed by the subjects in which they were asked to copy a template consisting of a spiral and seven lines of text. The first line in svc file structure represents the number of sample in each file and the n-the line represents Y coordinate, X coordinate, time stamp, button state, azimuth, altitude, pressure.

For the ease of users, the device is covered with sheet of paper and subject is requested to draw spiral and write 7 lines of text. Intuos Wacom Digitizer was used for data acquisition. All the subject had to complete the handwriting task following a template at any speed. Digitized signals were obtained during the movements. While person writes on the surface is called on surface movement and when the pen is slightly above the surface is called in air movement. The signals started to record as soon the pen touched the surface and recording was finish when the task was completed. x-coordinate, y-coordinate, time stamp, button status, pressure, tilt and elevation are the features recorded by the tablet. The button status is binary feature 0 when pen is in upstate and 1 for pen down state. The database also has PAHAW excel file which contains the characteristic of each subject. The characteristics are summarized in Figure 4.3.

	A	B	C	D	E	F	G	H	I	J
1	**ID**	**Nationality**	**Sex**	**Disease**	**PD status**	**Age**	**Dominant hand**	**LED**	**UPDRS III**	**Length of PD**
2	00001	Czech	F	PD	ON	68	R	1115	2	6
3	00002	Czech	F	PD	ON	78	R	2110	2	8
4	00003	Czech	F	PD	ON	69	R	1556.6	2	7
5	00004	Czech	F	PD	ON	79	R	1691	2	12

Figure 4.3: Characteristics of Parkinson Handwriting Database

4.4 High Level Design

The high level design including logical design and process design are explained in this section.

4.4.1 Conceptual/Logical Design

Figure 4.4 shows an overview of the logical or conceptual design of the system in terms of package diagram.

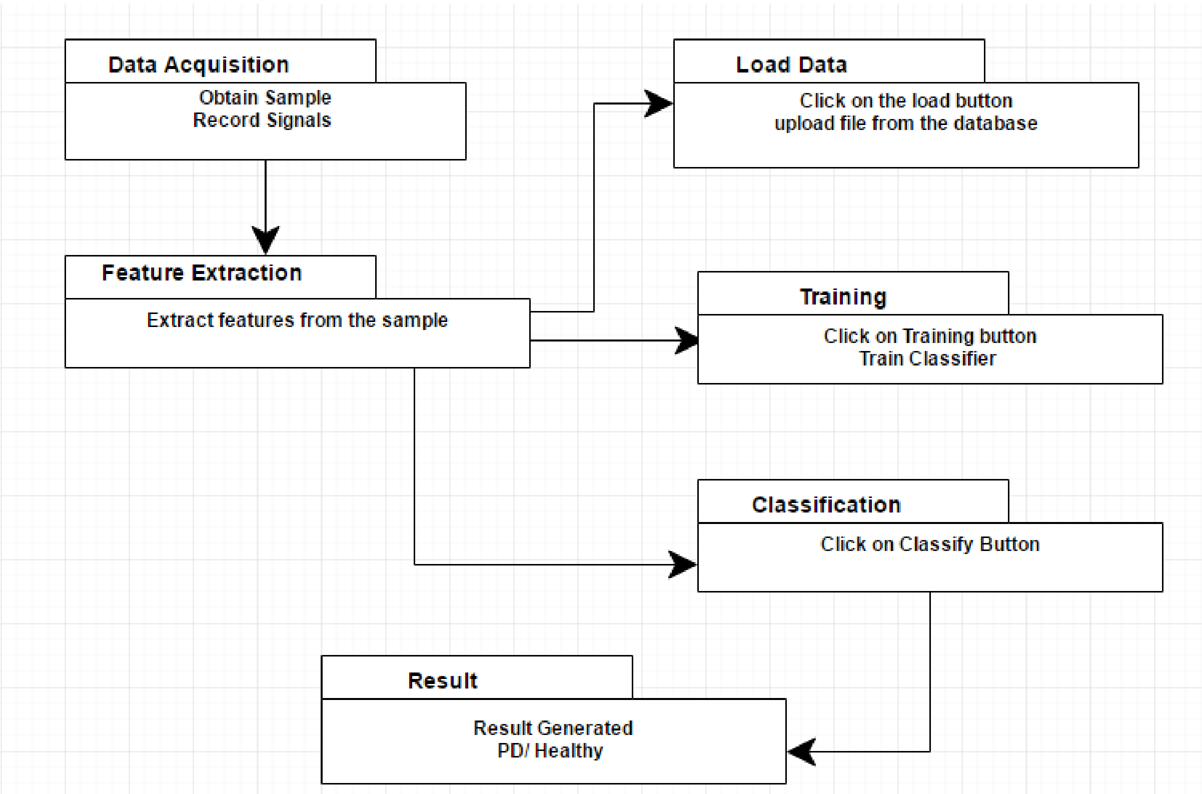

Figure 4.4: Package Diagram

4.4.2 Process Design

The process interaction in the system is summarized in Figure 4.5 while Figure 4.6 to Figure 4.10illustrate the sequence diagrams of the system.

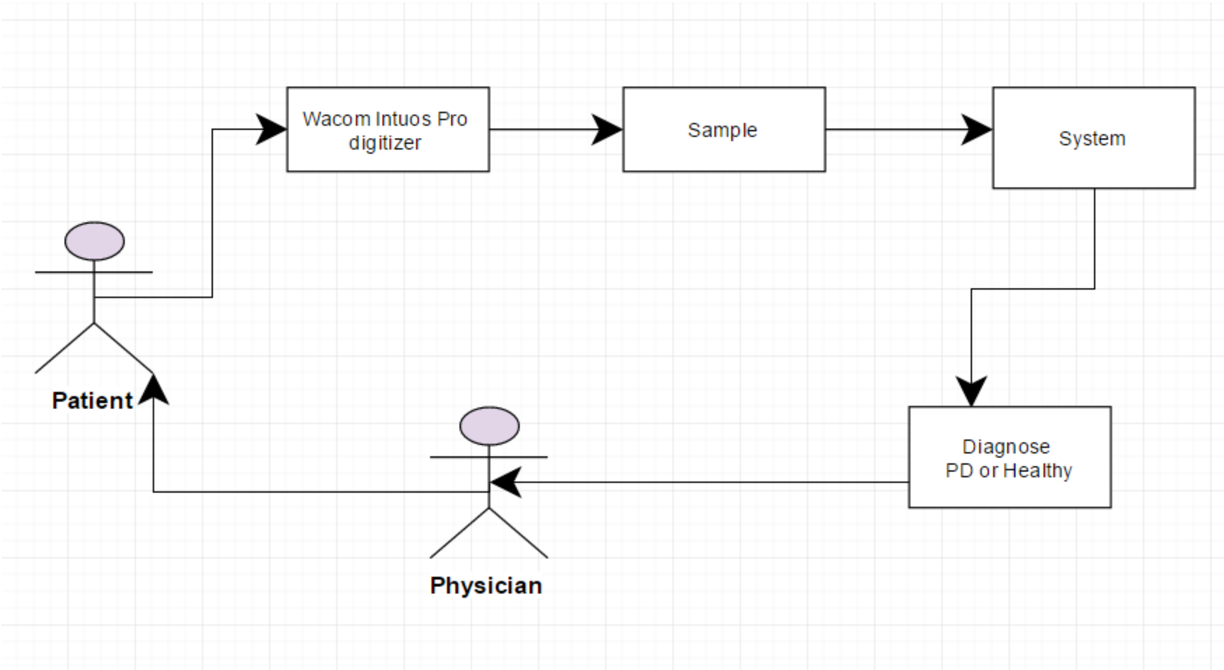

Figure 4.5: Process Interaction

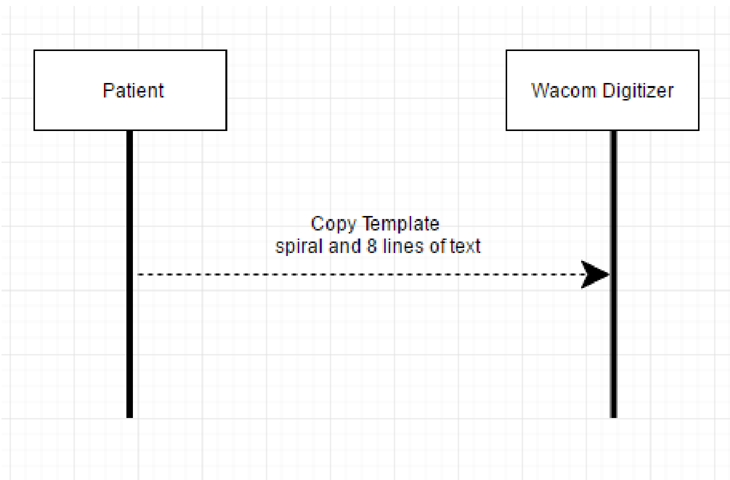

Figure 4.6: The user will be asked to copy template on digitizer

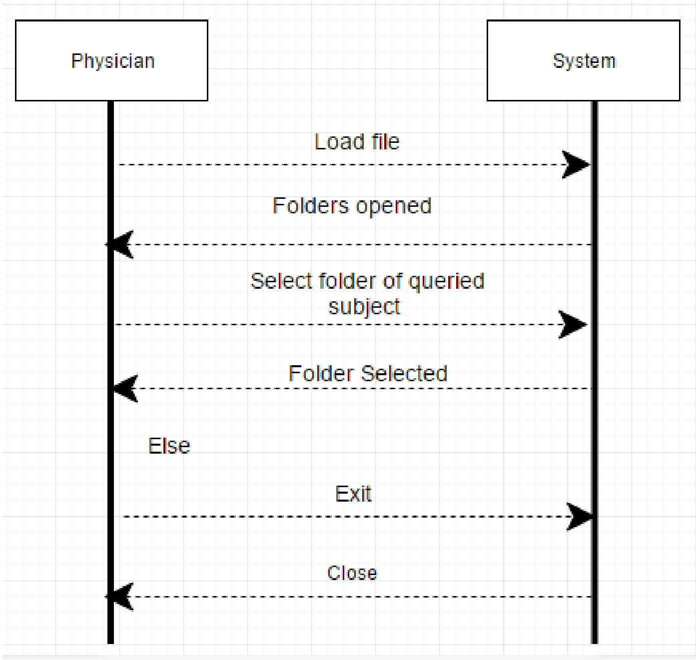

Figure 4.7: The user will load file

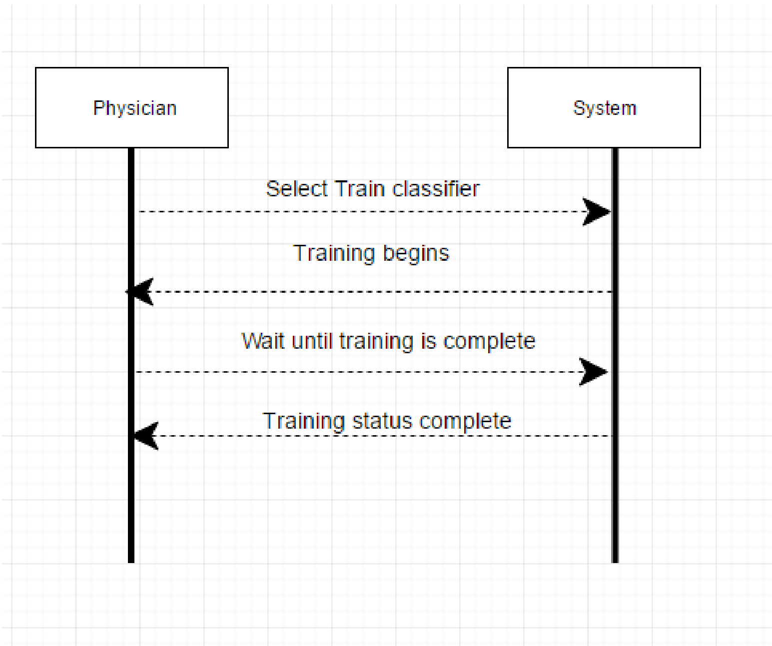

Figure 4.8: The user will train the classifier

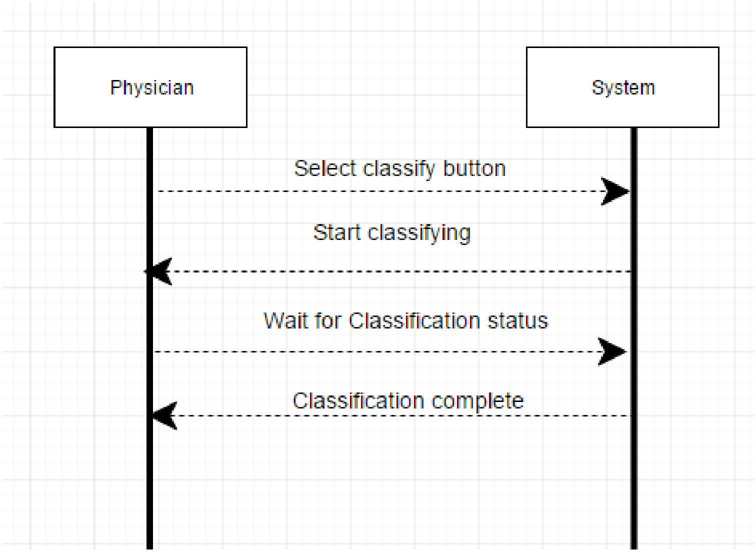

Figure 4.9: Classification

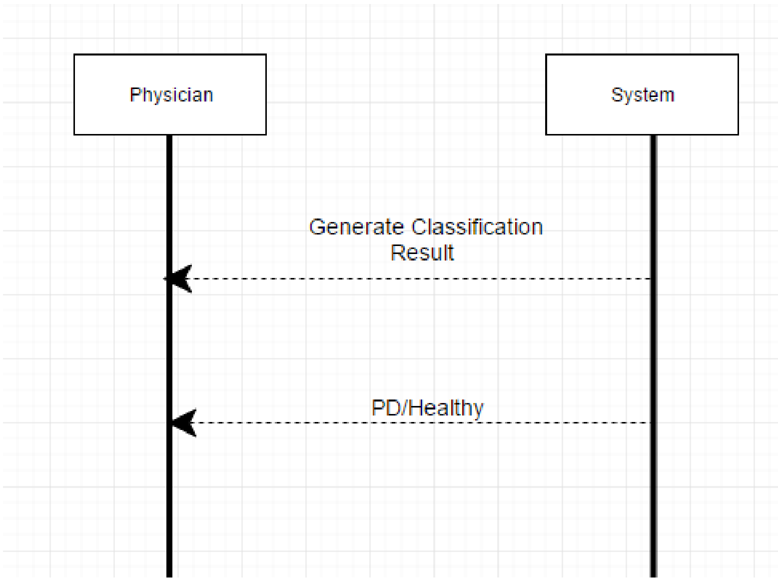

Figure 4.10: The System will declare as PD or Healthy

4.4.3 Security

The system is a desktop application and it should only be used by the practitioners. Patients should only be allowed to use the system under the observation of clinicians as he/she would have the detailed knowledge of the results which will be generated at the end of the test.

4.5 GUI Design

The GUI of a system requires to be simple and user friendly. The designed system will be used by physicians and patients instead of computer scientist, therefore the GUI had to be simple yet comprehensive. The system interface contains one form. The user will train the system first by pressing 'Train classifier' button. As soon training is complete the system displays the training status complete and show the 'Accuracy' of classifier as shown in Figure 6.1.User will click Test Classifier button to go to next form.

Figure 4.11: Training the Classifier!

Second form executes classification . As shown in Figure 6.2 user has choice of either give test sample to the classifier by clicking the browse button. The browse button opens folder dialog box where it can load test data sample from the PaHaW database.Once the test sample is selected user will click Classify button.The System will display Classification status Complete' and will display classification result on new forms as PD (Parkinson Patient) or Healthy.

Figure 4.12: Browse for test samples or Take Online Sample

Likwise user will click Take Online Sample button which opens up Wacom Input Test form for taking new sample from the Wacom Digitizer shown in Figure 6.3

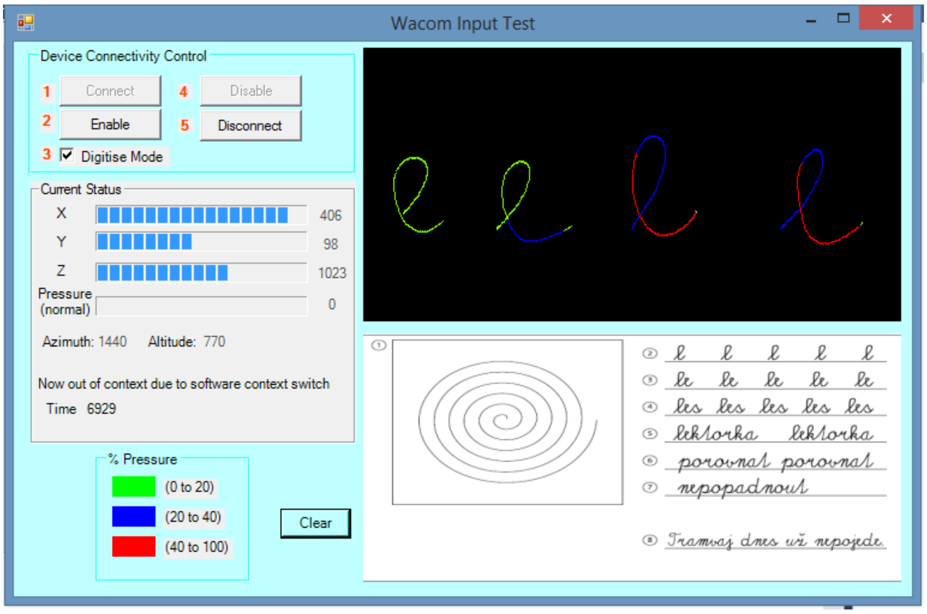

Figure 4.13: Wacom Input Test

User will click coonect button then enbale button .Once the user will check the digitize mode check box the device will be ready to input handwriting sample of user. The user will follow the handwriting template pasted on the form. As soon the user starts writing on the device using the pen , the handwriting will be displayed on the black screen . While the user continues drawing on the device cartesioan coordinates (x, y) , pressure , time , azimuth , altitude will change untill the user stops . The values of all these features will saved in a file . Eight sample will be recorded and a folder of subject will be created just like in PaHaW database. The test sample can then be fed to classifier . User will click classify button to process the test sample.

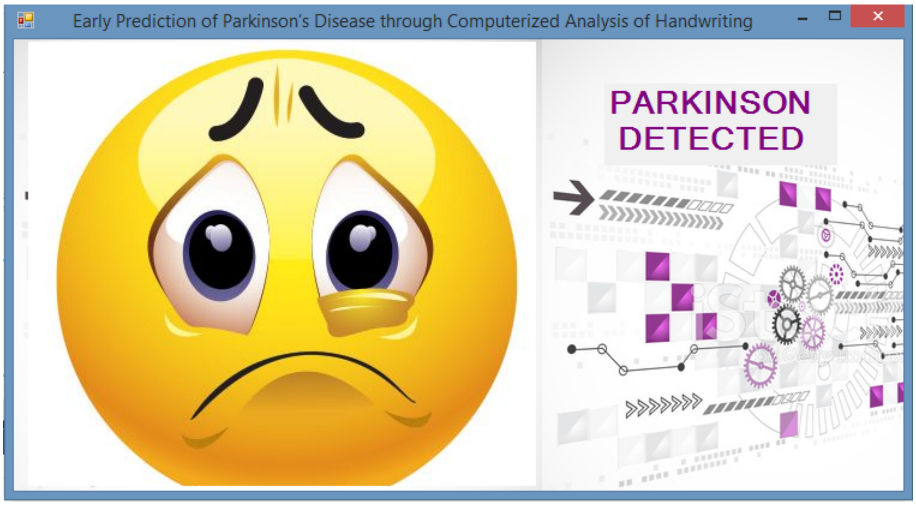

Figure 4.14: Subject is classified as PD patient

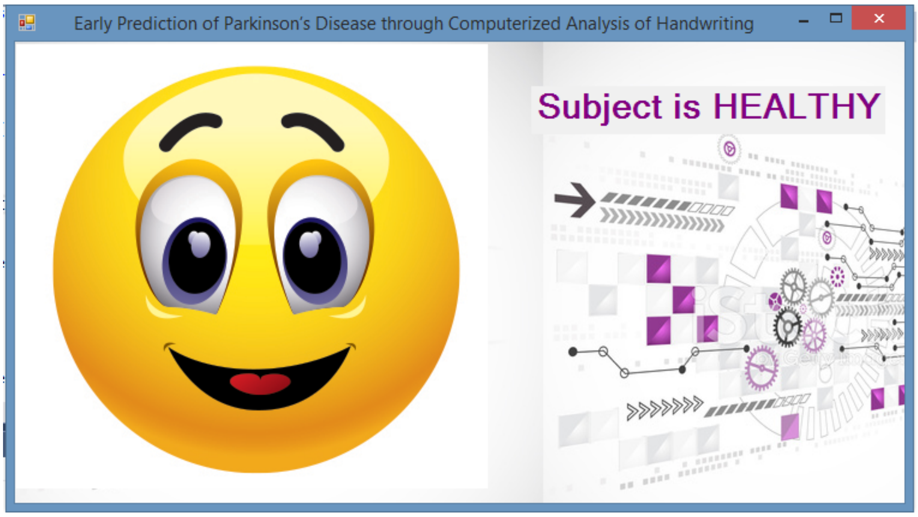

Figure 4.15: Subject is classfied as Healthy

Classification results are show in Figure 4.14 and Figure 4.15.

Chapter 5

System Implementation

Implementation details of our system are explained in this chapter comprising IDE, code structure, libraries employed, algorithmic details and methods implemented.

5.1 Development Enviroment

We employed Visual Studio Development tools for developing our system that works in Visual Studio Integrated Development Enviroment. These tools are used for creating the system from the planning phase, coding, testing, debugging, UI design, analyzing the code and performance and evaluating the system. Our system is implemented in C# language.

5.2 Libraries (Accord.NET)

The system mainly works on classification where online handwriting samples are input to the system to label given input sample as PD or Healthy. For classification, we have used Accord.Net Machine learning framework. Accord.NET framework contains multiple libraries for scientific computing, machine learning, pattern recogintion and statistical data processing etc. The libraries include the following.

- Accord.Math: Library with special Mathematical functions for scientific computations and matrix extension library with collection of numerical optimization algorithms. Accord.Math evaluates the formulas used for extracting features from the data samples and many other computations in classifying the data.

- Accord.Machine Learning: The most important library which operates the learning classifier. The library contains many models and algorithms like SVM, Decision Trees and Gaussian Mixtures etc. for machine learning. We have used Support Vector

Machine for training taking 60 handwriting samples with 30-30 distribution of PD and Healthy subjects from the PaHaW database.

- Accord.Statistic.Kernels: This library comprises of kernel functions/methods like Gausian kernel for machine learning. We implemented gausian kernel for transformation of data in the feature space.

- Accord.IO : Another very important library in the system implementation for loading and saving data from different file formats. The system picks files for training and testing from PaHaW handwriting Database. The files in the database are of SVC format and arranged in folders with each folder holding eight files of handwriting samples in digitized format.

5.3 Intuos Wacom Digitizer

The Parkinson prediction system will be capable of testing new online handwriting data samples as well. The PD detection system will take new live handwriting samples through Intuos Wacom digitizer. The Pen tablet will generate online data while the person is writing/ drawing on the tablet. It is a pressure sensitive multi touch wireless device with creative pen that delivers a wide range of pressure sensitivity and tilt recognition. The opacity and control line width depends on how lightly the user will touch the pen with the surface.

5.4 Algorithmic Development and Methodology

This section explains the implementation of proposed methodology and algorithmic details.

5.4.1 Training

Supervised learning is type of machine learning in which known dataset or labeled data is used to make prediction about unknown or test data. Training dataset consists of input data, a set of training examples and against each example there is response value called label which represents the class to which data belongs. The supervised learning algorithm creates a model which analyzes training data to map new examples. To validate the model, test dataset is fed to the learning algorithm or model which determines the class label for the test data. The learning algorithm makes generalizations from the known examples in training set and by doing statistical analysis of the data, employs knowledge deduced from the training examples makes generalizations about the unknown data and assign test data specified class to which it belongs.

In our system we used 60 percent of PaHaW Handwriting database samples to train the classifier which is Support Vector Machine. The PaHaW database contains labeled handwriting samples. We gave 30 Healthy handwriting samples and 30 Parkinson handwriting samples to train the classifier. Two classes PD (represented as 1 in code) and Healthy (represented as 0 in code) have been defined on the training samples and SVM classifier uses these two classes to analyze new examples and determine the classification based on some threshold. The samples are acquired by recording digitized signals from Wacom device while the subject draws a spiral and writes eight lines of text following a specific template. The SVM analyzes all the samples in the training set one by one.

5.4.2 Feature Extraction

Feature Extraction is technique where raw data is transformed into optimal features which are input for the learning algorithm or classifier. Multiple kernel learning is applied in supervised learning for feature extraction to improve classification accuracy. Kernels are mapping functions used to make labels for adjusting distance of a vector from the edge of class. As mentioned earlier, a significant difference is seen in handwriting of PD patients and healthy persons. A healthy person writing is stable and continuous where as person suffering from Parkinson will take more time to write than a healthy person and his/her writing will be jerky and not smooth. These facts can be used to explore new features. The PaHaW handwriting database contains handwriting samples with features in form of Cartesian coordinates of pen position, timestamp, pressure, azimuth and tilt. Features are computed separately for each of eight task and then combined together. We used these features to extract new features which are as follows.

- Stroke Speed: Stroke is continuous movement between two consecutive pen lifts. Spiral and letters can be drawn in one stroke continuously. A healthy person can easily draw spirals and loops in one go comfortably in less time. Parkinson affected person draws the spiral and loops in multiple strokes and takes more time. Therefore stroke speed is an effective feature and can be used as threshold to discriminate among two classes. Stoke speed is calculated by computing stroke length. Distance formula is implemented to calculate the stroke length. Once stroke length is calculated for each handwriting task it is divided by the stroke duration to compute the stroke speed.
  Stroke speed = Stroke Length/ Stroke duration [mm/s]

- Speed: Speed refers to the overall average speed of the complete task and can be computed by using the x and y coordinates and the timestamp information.Handwriting speed is calculated by formula

Speed=trajectory during handwriting / handwriting duration [mm/s]
Speed is basically distance coverd with rest to time.Trajectory is the total distance between the cartesian coordinated again calculated and implemented using distance formula.The trajectory during handwriting is divided by total time . The decsion criteria is that Pd patient handwriting speed will be slow than healthy person.

- Velocity: Change in position of pen in unit time is velocity. The rising and falling edges of the handwriting are computed by finding the difference between the maximum and minimum values. A visible difference is seen among the smoothness for different tasks. Velocity is computed by formula
Velocity:change in pen position/time [mm/s]

- Normalised on-Surface Time The digitized pen records signals when pen is lifted slightly above the surface and when pen touches the surface. Normalized surface time is the time while the pen is touching the screen surface by whole time duration. Whole time duration means the time during when pen lifted slightly above the surface plus the time pen is touching the surface. It is a common way of writing where we do lift the pen up and down. The formula is given by
Normalized on surface time = time in which pen touching surface/total writing duration

5.4.3 Classification

Classification is the process of taking a classifier which has been trained on some training data, running it on unknown data to determine the class for the test data. The Parkinson prediction Sytem is implemented with Support Vector Machine classifier (Figure 5.1). The classifier gives classification results by assigning PD lablel or Healthy label to the test sample.

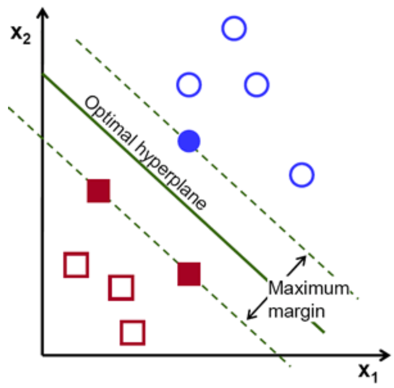

Figure 5.1: SVM Classifier

5.4.4 Live Test Samples - Intuos Wacom Digitizer

The overall process can be summarized as illustrated in Figure 5.2.

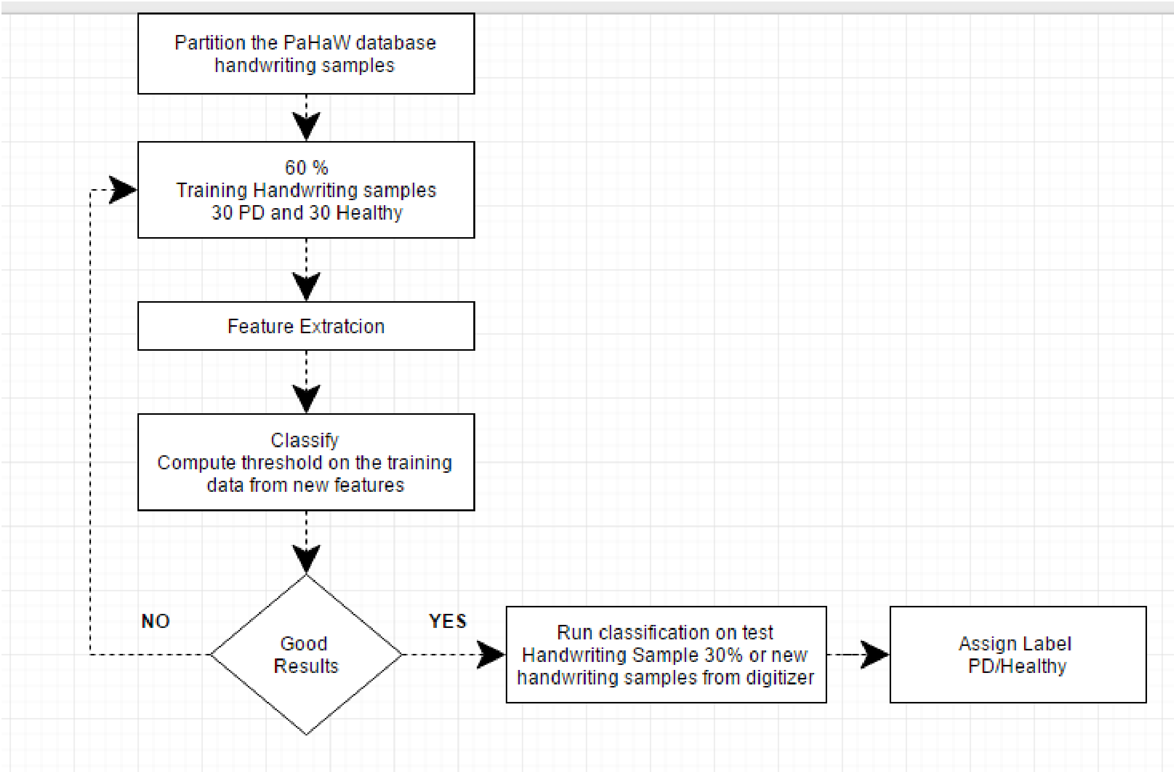

Figure 5.2: System Implementation

The PD detection system will work on fresh handwriting samples obtained from Wacom Digitizer. In brief the user will have to follow a template to draw a spiral and eight lines of text on the tablet screen. The handwriting sample of each task will be saved in the same format as that of the handwriting samples of PaHaW handwriting database. Eight seperate files for each task will be created and will be fed to the classifier for a decision.

Chapter 6

System Testing and Evaluation

System testing is the process of verification which confirms that the developed software is working in accordance with the requirements specified and is executing efficiently. Testing is the best way of ensuring that the developed software meets the quality standards or not. In this chapter we discuss the testing of our developed system.

6.1 Graphical User Interface Testing

Graphical User Interface(GUI) is one of the most important modules of any software as it is important to satisfy the user and make them at ease to use the software by making the interface simple, interactive and easily understandable. For our current developed system, our potential users are patients and medical experts which are not the experts of computer domain. There are two major forms requiring the user interaction and we have tried our best to maintain visual consistency for user ease. Eight subjects were requested to interact with the system and their feedback was recorded after the interaction sessions. The subjects were very clear and satisfied regarding the efficiency and easily understandable interfacing. Figure 6.1 , Figure 6.2 and Figure 6.3 illustrate the interactive forms of the system.

Figure 6.1: Training Interface

Figure 6.2: Classification Interface

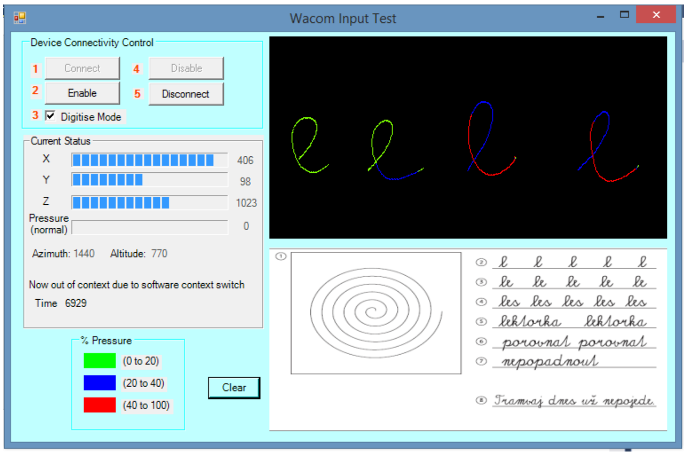

Figure 6.3: Acquire Online Data Sample

6.2 Usability Testing

The system is fairly simple to use and easy to understand and the usability guidelines have been followed in the design. As discussed previously, eight subjects were made to use the system and provided feedback on the different aspects of interface. There were no major issues reported in the system's usability and users completed the required tasks within or before the expected time.

6.3 Software Performance Testing

Software Performance testing is carried out to ensure that the developed software confines to the requirements specified and is efficient and reliable to use. In our developed system, reliability refers to the process of correct classification of a person to a diseased or healthy. There are total 75 samples out of which 60% of the data is used for training the classifier and rest 40% of data is used for testing. Summing up the classification accuracy comes out to be 83.3% as shown in Figure 6.4.

	Health	PD
Healthy	12	3
PD	2	13

Overall : 25 correct classifications out of 30, Rate: 83.3%

Figure 6.4: Confusion Matrix showing Classification Performance of System

6.4 Compatibility Testing

Compatibility Testing is the process of ensuring that the developed application is capable enough to run on different platforms. The present application is developed in Microsoft Visual Studio 2013 and is compatible with different versions of Windows Operating System.

6.5 Exception Handling

The system may generate an exception if the Accord.NET DLLs are not placed n the resource file of the project. The DLLs must be therefore added correctly in order to avoid the exception.

6.6 TestCases

6.6.1 Test Case 1: Train Classifier

Table 6.1: Test Case 1

Test Case ID:	TC1
Unit of Test:	Testing the Training Process
Steps to execute:	Click on Train Classifier button
Expected Result:	Training Process executed successfully
Actual Result:	Successful execution of Training Process
Status:	Pass

6.6.2 Test Case 2: Browse Folder

Table 6.2: Test Case 2

Test Case ID:	TC2
Unit of Test:	Testing the browse Process
Steps to execute:	Click on browse file button
Expected Result:	Browse folder option opened in windows explorer
Actual Result:	Browse Option worked successfully
Status:	Pass

6.6.3 Test Case 3: Classification of Subjects

Table 6.3: Test Case 3

Test Case ID:	TC3
Unit of Test:	Testing the Classification Process
Steps to execute:	Click on Classify button to begin classification
Expected Result:	Proper Classification of subjects
Actual Result:	Classified successfully
Status:	Pass

6.6.4 Test Case 4: Acquire Online Sample

Table 6.4: Test case 4

Test Case ID:	TC4
Unit of Test:	Testing the values of input data.
Steps to execute: required sample with help of Pen.	Connect the Wacom device and draw the
Expected Result:	Values successfully recorded in file.
Actual Result:	Values successfully recorded in file.
Status:	Pass

Chapter 7

Conclusions and Prespectives

7.1 Conclusion

Parkinson is one of the progressive disease mostly found in late ages and diagnosed only in later stages when the disease has progressed significantly. Many different ways have been proposed by researchers for early prediction of Parkinson disease, handwriting being one of the most effective of all these. This study was aimed at automating the process of using handwriting as a tool for Parkinson detection by using computational algorithms of image analysis and pattern recognition. The application is first trained by providing samples of PD patients and control subjects. Online features of handwriting include strokey speed, task speed, pen pressure and pen-up and down times etc. are employed to discriminate between the two classes using a support vector machine classifier. The system was evaluated on a benchmark database and realized a promising classification rate of 83%. The system can also be tested for new subjects by data acquisition using the Wacom digitizer.

7.2 Perspectives

The present system can be enhanced further by using other classifiers and combine the outputs of multiple classifiers to enhance the overall classification rates. Moreover, in addition to online featuers, offline features of handwriting can also be investigated and combined with the online features to have a richer representation of a sample. It is expected that this study will contribute towards the development of a complete diagnostic tool which will reduce the work load of psychologists by automating the process of disease diagnosis. The authors also hope that the system developed during this academic project will be helpful for researchers and developers aiming to automate other similar tests.

Chapter 8

Appendix-A

Samples of Tests

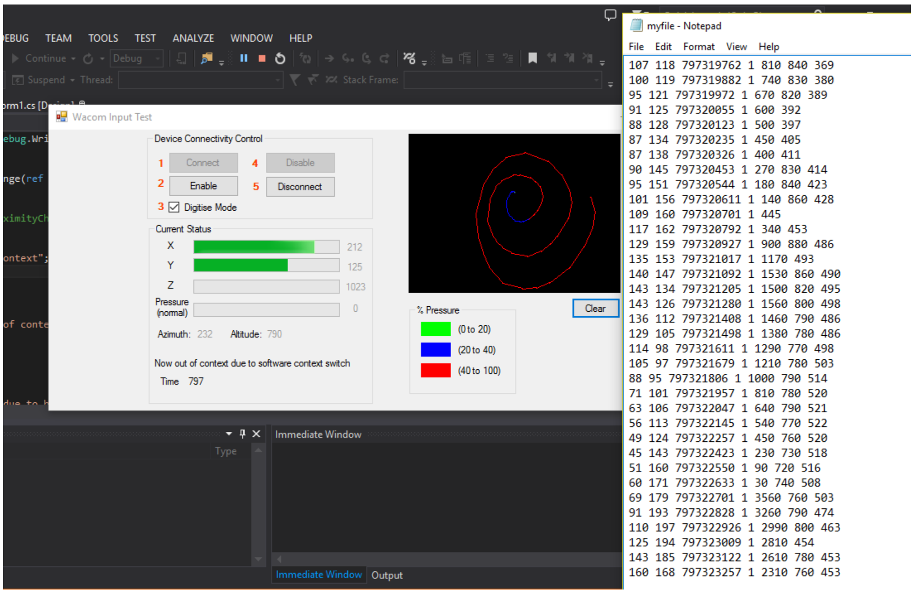

Figure 8.1: Sample Input of subject- Subject 1

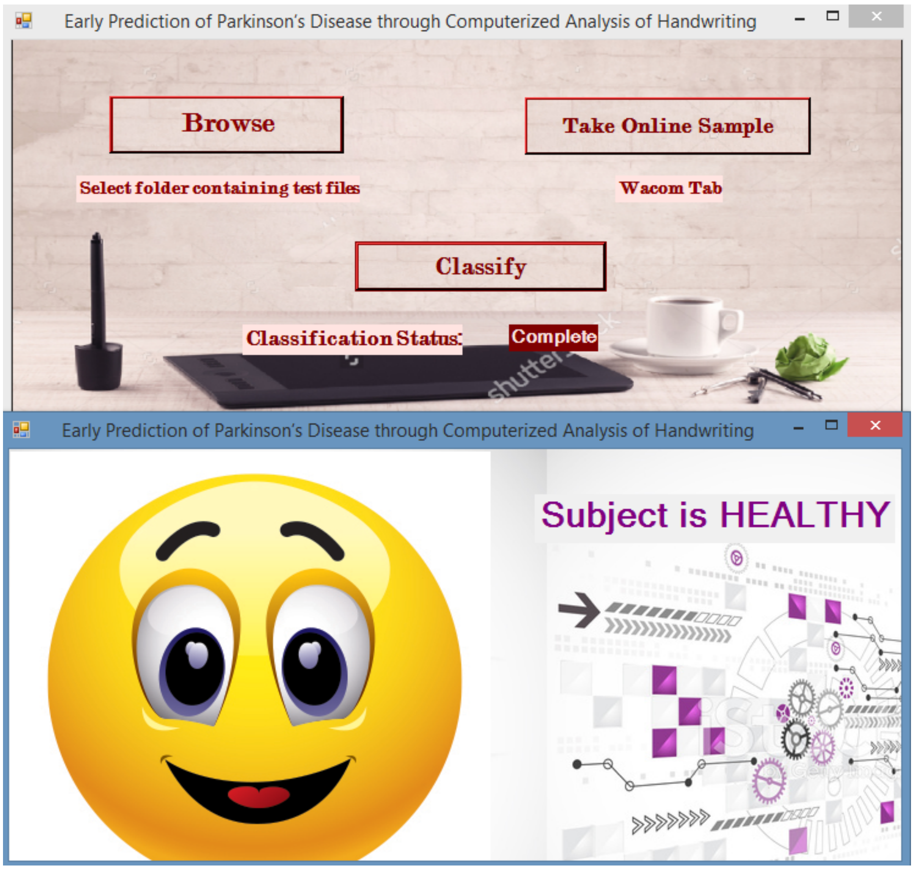

Figure 8.2: Output of of subject- Subject 1

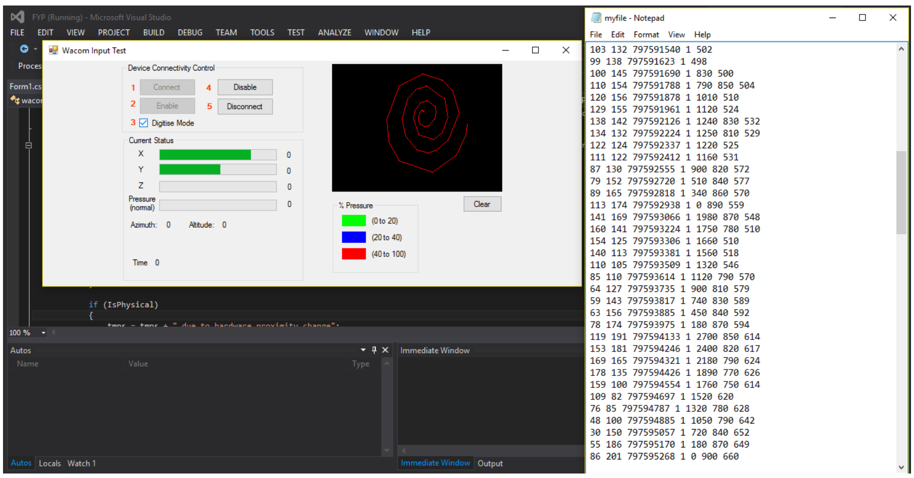

Figure 8.3: Sample Input of Subject- Subject 2

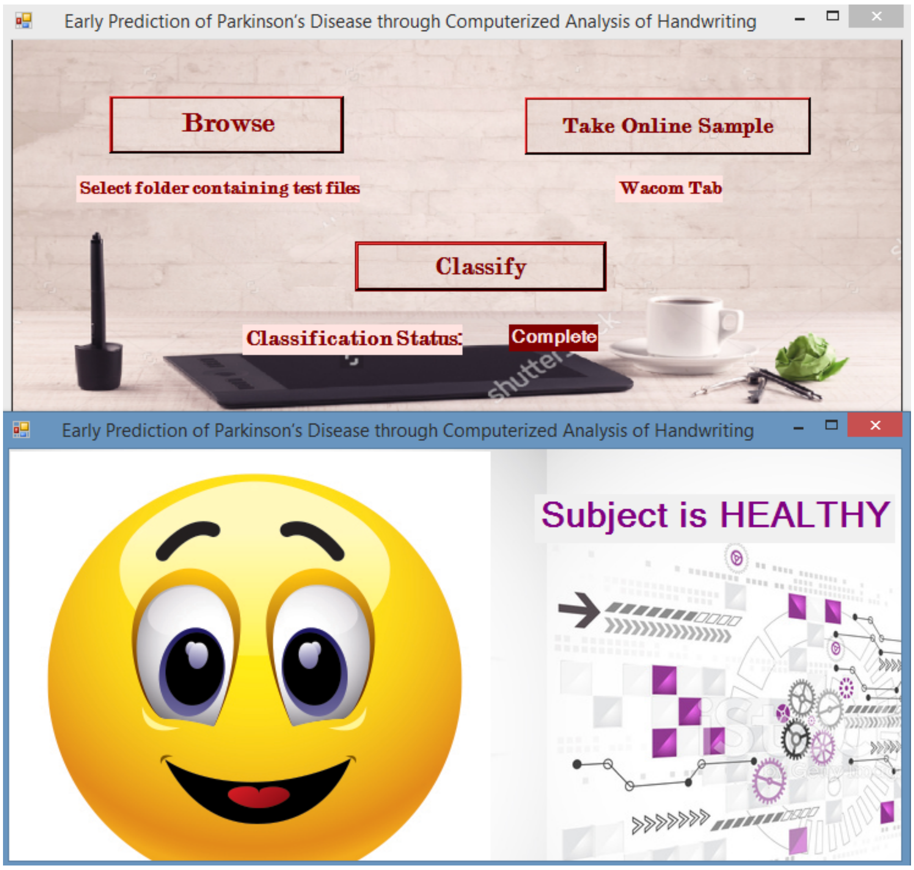

Figure 8.4: Output of subject- Subject 2

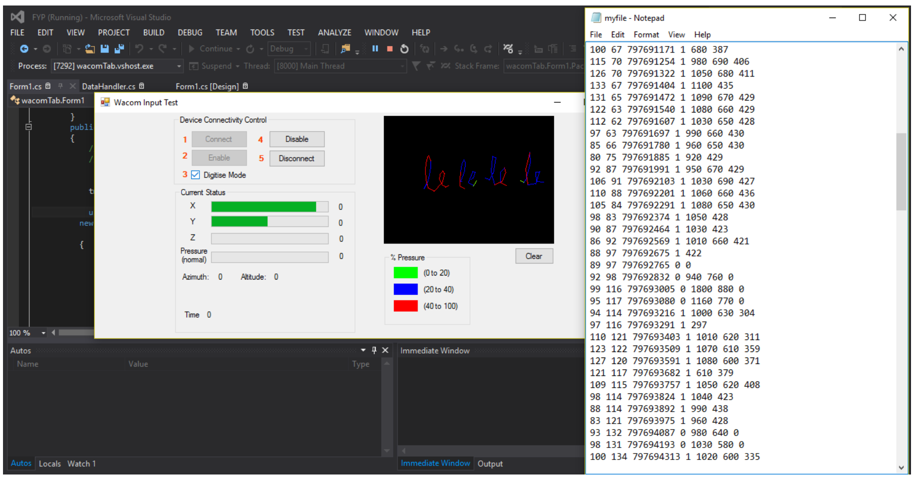

Figure 8.5: Input of subject- Subject 3

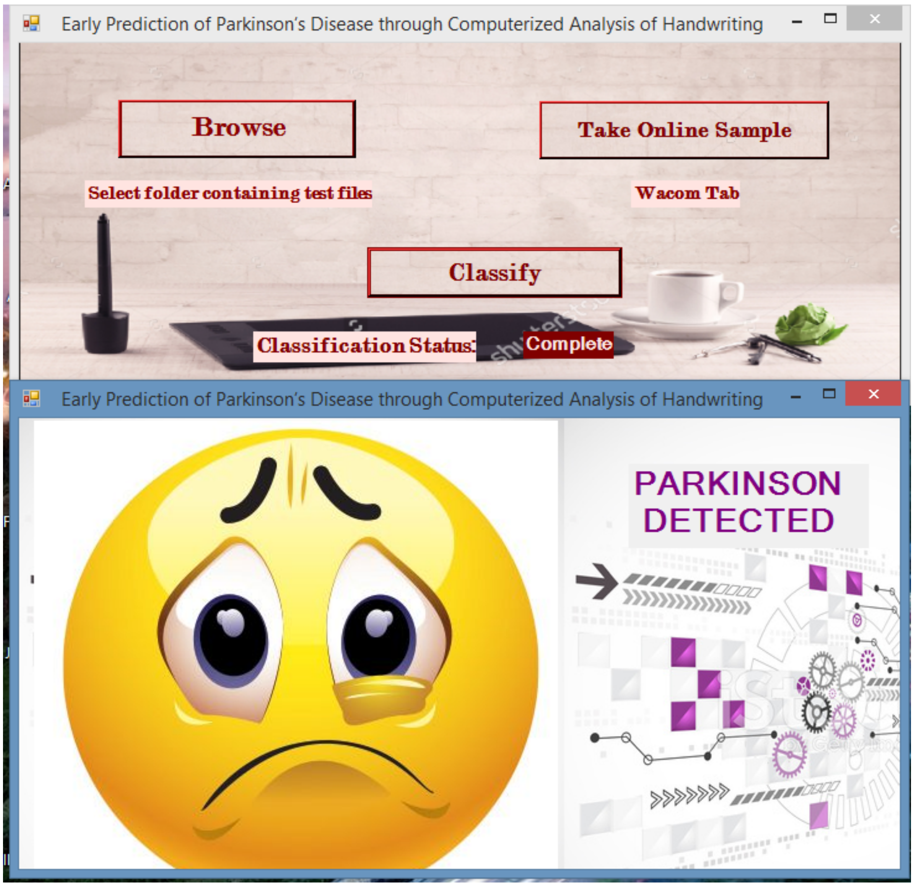

Figure 8.6: Output of subject- Subject 3

References

[1] HL Teulings, Jose L Contreras-Vidal, GE Stelmach, and Charles Howard Adler. Adaptation of handwriting size under distorted visual feedback in patients with parkinson's disease and elderly and young controls. *Journal of Neurology, Neurosurgery & Psychiatry*, 72(3):315–324, 2002. Cited on p. 1.

[2] Silke Anna Theresa Weber, Carlos Alberto dos Santos Filho, Arthur Oscar Shelp, Luiz Antonio Lima Resende, João Paulo Papa, Christian Hook, et al. Classification of handwriting patterns in patients with parkinson´ s disease, using a biometric sensor. *Global Advanced Research Journal of Medicine and Medical Sciences*, pages 362–366, 2014. Cited on p. 1.

[3] TE Eichhorn, T Gasser, N Mai, C Marquardt, G Arnold, J Schwarz, and WH Oertel. Computational analysis of open loop handwriting movements in parkinson's disease: a rapid method to detect dopamimetic effects. *Movement disorders*, 11(3):289–297, 1996. Cited on p. 1.

[4] Athanasios Tsanas, Max A Little, Patrick E McSharry, and Lorraine O Ramig. Accurate telemonitoring of parkinson's disease progression by noninvasive speech tests. *IEEE transactions on Biomedical Engineering*, 57(4):884–893, 2010. Cited on p. 1.

[5] Musa Peker, Baha Şen, and Dursun Delen. Computer-aided diagnosis of parkinson's disease using complex-valued neural networks and mrmr feature selection algorithm. *Journal of healthcare engineering*, 6(3):281–302, 2015. Cited on pp. 1 and 7.

[6] Peter Drotár, Jiří Mekyska, Zdeněk Smékal, Irena Rektorová, Lucia Masarová, and Marcos Faundez-Zanuy. Contribution of different handwriting modalities to differential diagnosis of parkinson's disease. In *Medical Measurements and Applications (MeMeA), 2015 IEEE International Symposium on*, pages 344–348. IEEE, 2015. Cited on pp. 3 and 8.

[7] Peter Drotár, Jiří Mekyska, Irena Rektorová, Lucia Masarová, Zdeněk Smékal, and Marcos Faundez-Zanuy. Evaluation of handwriting kinematics and pressure for differential diagnosis of parkinson's disease. *Artificial intelligence in medicine*, 67:39–46, 2016. Cited on pp. 3, 8, 9, 10, and 11.

[8] Athanasios Tsanas, Max A Little, Patrick E McSharry, Jennifer Spielman, and Lorraine O Ramig. Novel speech signal processing algorithms for high-accuracy classification of parkinson's disease. *IEEE Transactions on Biomedical Engineering*, 59(5):1264–1271, 2012. Cited on p. 7.

[9] Atilla Ünlü, Rüdiger Brause, and Karsten Krakow. Handwriting analysis for diagnosis and prognosis of parkinson's disease. In *International Symposium on Biological and Medical Data Analysis*, pages 441–450. Springer, 2006. Cited on p. 7.

[10] Sara Rosenblum, Margalit Samuel, Sharon Zlotnik, Ilana Erikh, and Ilana Schlesinger. Handwriting as an objective tool for parkinson's disease diagnosis. *Journal of neurology*, 260(9):2357–2361, 2013. Cited on p. 7.

[11] Peter Drotár, Jiří Mekyska, Irena Rektorová, Lucia Masarová, Zdeněk Smékal, and Marcos Faundez-Zanuy. A new modality for quantitative evaluation of parkinson's disease: in-air movement. In *Bioinformatics and Bioengineering (BIBE), 2013 IEEE 13th International Conference on*, pages 1–4. IEEE, 2013. Cited on pp. 8, 9, 10, and 11.