Abnormal Gait Detection using Smartphone

by

Satyam Satyajeet
201511055

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of

MASTER OF TECHNOLOGY
in
INFORMATION AND COMMUNICATION TECHNOLOGY
to

DHIRUBHAI AMBANI INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGY

April, 2017

Declaration

I hereby declare that

i) the thesis comprises of my original work towards the degree of Master of Technology in Information and Communication Technology at Dhirubhai Ambani Institute of Information and Communication Technology and has not been submitted elsewhere for a degree,

ii) due acknowledgment has been made in the text to all the reference material used.

Satyam Satyajeet

Certificate

This is to certify that the thesis work entitled ABNORMAL GAIT DETECTION USING SMARTPHONE has been carried out by SATYAM SATYAJEET for the degree of Master of Technology in Information and Communication Technology at *Dhirubhai Ambani Institute of Information and Communication Technology* under my/our supervision.

P S Kalyan Sasidhar
Thesis Supervisor

Acknowledgments

Foremost, I would like to express my sincere gratitude to my enthusiastic thesis guide Dr. *P S Kalyan Sasidhar* for the continuous support for my research, for his patience, motivation, and immense knowledge to the highest standard inspired me and motivate me. His guidance helped me in all the time of research and writing of this thesis. He also helped me for data collection.

I would also like to acknowledge respected faculties of my thesis evaluation committee Dr. *Sanjay Shrivastava* and Dr. *Gagan Garg*, for their very valuable comments on this thesis.

I take this opportunity to sincerely thanks the Dhirubhai Ambani Institute of Information and Communication Technology, Gandhinagar, Gujarat for providing me an infrastructural assistance in the form of research lab which buttressed me to perform my work comfortably.

I would also like to thank Dr *Ankit*, Dr *Gyan Ranjan* and Dr *Vivek* for helping through the Data Collection process.

I would also like to thank my family for the support they provided me throughout all my studies. Finally, there are my friends. We were not only able to support each other by deliberating over our problems and findings, but also happily by discussing about things other than just our papers.

Thank you everyone!

Satyam Satyajeet

Contents

Abstract

Gait cycle is repetitive walking pattern involving steps and strides. Difference between abnormal gait and normal gait lies between gait parameters and both are compared for prediction. We are proposing a method which is cheap and using only Smartphone embedded accelerometer to extract gait parameters. The advantages are low cost and low power supply requirements with everyone having Smartphone making it user friendly. We collected data for normal and abnormal patients having various kinds of diseases. Problems such as Rheumatoid Arthritis (RA), Osteoarthritis (OA), sciatica, calcaneal spur (or heel spur), Ankylosing spondylitis, Motor Injury, polio and Rotation of knee. The classifiers used were Naives Bayes (NB), Decision Tree (DT) and Random Forest (RF) out of which RF performed best giving 91.52% accuracy on 10-fold cross validation Set. DT and NB were giving accuracy of 86.38% and 89.69%.

List of Principal Symbols and Acronyms

g	Gravitational Acceleration
g_x	Gravity component along x-axis
g_y	Gravity component along y-axis
g_z	Gravity component along z-axis
C	Cadence
CM	Confusion Matrix
CV	Cross-Validation
DT	Decision Tree
N	Number of Steps
NB	Naive-Bayes
RF	Random Forest
SL	Step Length
SS	Stride Speed
ST	Step Time
STR	Stride Lentgth
STT	Stride Time

List of Tables

List of Figures

CHAPTER 1
Introduction

A walk comprises of many gait cycles. The span between successive initial contacts of the reference leg with the ground is termed as a *gait cycle*. Each gait cycle consists of two phases stance phase (the part in which the foot is in contact with the ground) and swing phase (the part in which the foot is in air). These two phases constitute 60% and 40% of a gait cycle respectively for a normal person [5]. A gait cycle starts with a heel strike and ends with a heel strike. A heel strike is the point of initial contact with the ground [5].

- **Stance Phase:**
  Stance phase starts with *heel strike* and ends with toe-off (Figure 1.1). Toe_off is the start of initial swing phase of the reference leg. In between there are events such as $foot\_flat$ in which the entire foot is in contact with the ground, midstance is the 50% of stance phase and $heel\_off$ is the starting point where heel leaves the contact with the ground.
- **Swing Phase:**
  Swing phase starts with toe_off and ends with *heel strike* (Figure 1.1). It has three subphases: early swing, mid swing and late swing [5]. In early swing phase, the acceleration of reference leg takes place upto mid swing when leg is in air and it contributes to 10-15% of gait cycle [5]. Mid swing phase contributes 10% of gait cycle and from mid swing to late swing deceleration takes place. Late swing contributes 15% of gait cycle.

Difference between abnormal and normal walking style lies between the difference in gait parameters. So we collected data from 35 people out of which 18 were having abnormal walking style and 17 were having normal walking style, for training purposes. We are using tri-axial accelerometer in smartphone and using AndroSensor application for collecting gravitational acceleration(g) on all the axes. Through investigation we found that g_x (gravity component along x-axis of the smartphone) is relevant parameter for exploring gait cycle. Abnormal walking style was observed in patients having diseases such Rheumatoid Arthritis, Ostrio Arthritis, Polio, Ankylosings Spondilitis, Heel Spur, Motor Injury, Lower Back Pain, Knee Pain, Sciatica. Out of total 18 patients 12 were female and 6 were male, showing that female has more abnormal walking style compared to men. For both abnormal and normal, gait parameters were extracted. We calculated gait parameters for both legs individually for a single subject. We made one dataset post-extraction for both left and right leg containing almost 185 observations. Lastly, we made three models one based on decision tree, one based on Random Forest and the other based on Naive-Bayes classifier. We concluded that *stride_length*, *swing_ratio* can combinedly classify the subject′s abnormality using the simplest decision tree, while using random forest *stride_speed* was the most important parameter for classification with other three parameters *stride_length*, *number_of_steps* and *swing or stance_ratio*.

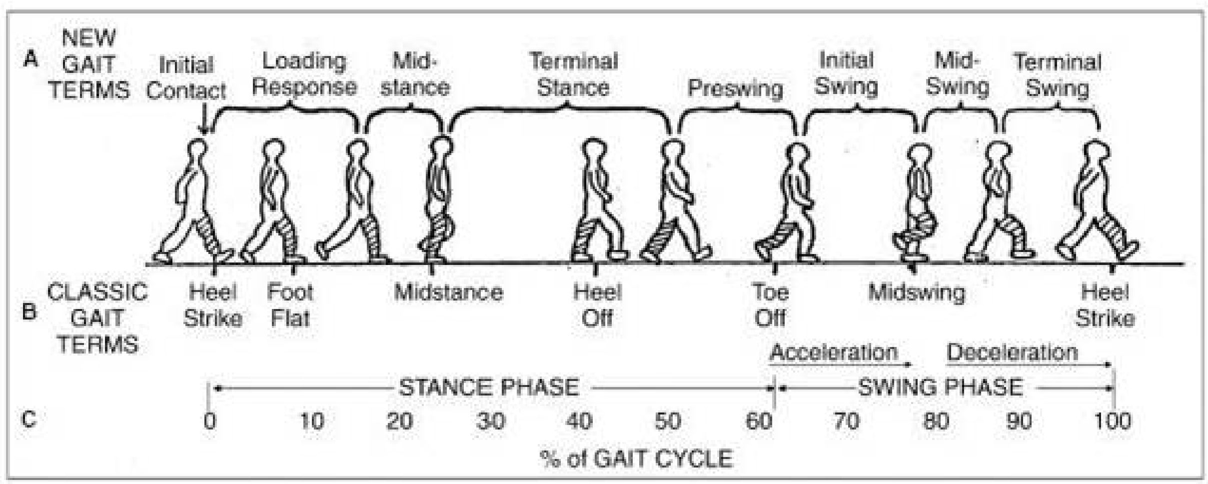

Figure 1.1: Various phases in Gait Cycle [2]

1.1 Problem Statement

1.2 Motivation

Non-Technical Motivation: Preventing the user by alerting them at an earlier stage helps them improvise their walking style. Also, one lacks resources for checking gait symptoms, so they might have to visit physiotherapist very often which will cost more. As most people have Smartphone with them nowadays, it is feasible to develop a solution using Smartphone which can detect and alert the users. Physiotherapist needs to check the improvement of patients at regular intervals. So that we can propose a method to reflect improvements in pateint′s health graduallly.

Technical Motivation: Smartphone has inbuilt sensors, processors, huge computational power and it is portable. With so much of abundant resources in a Smartphone, it has become a platform where problems can be solved. Solutions proposed till now are only differentiating between normal and PD or Lower back pain.

As usually doctors track the patients how he walks depending on certain number of parameters such as step length, limping etc while walking and then decide whether they walk abnormally or normally. Hence the problem can be classified as a pattern recognition problem.

1.3 Literature Survey

Smartphone nowadays a lot of embedded sensors such as gyroscope, accelerometer and compass. A huge amount of memory to act as data collector and powered by Android or ios Operating Systems for computation and processing purposes. Research on extraction of gait parameters and walking style detection using Smartphone is on rise from 2-3 years. Most of the papers are using External Sensors like gyroscope; accelerometer or pressure sensors to extract gait parameters such as $Step\ Length$, $Stride\_Speed$, $Cadence$ and $Number\_of\_Steps$ which requires battery to provide power. They are also using Bluetooth or Wi-Fi connectivity between Smartphone and External Sensors which also consumes power of Smart phone. Some of them are using embedded sensors on Smartphone.

The real time estimation of gait parameters which has Shoe-worn inertial sensor (Accelerometer and Gyroscope) and also requires a Li-ion battery (with a lifetime of 6h)[3]. The subjects they took for consideration were normal and Parkinson′s disease (PD) [3]. Shoe with 1200mAh Li-battery (3.7 - 4.2 V) supply and One 3-axis accelerometer, one 3-axis gyroscope, one 3-axis magnetometer sensors, pressure sensor and Bluetooth for wireless connectivity were used to extract gait parameters [6]. A Smart Shoe comprising of 4 pressure sensors and arduino to send data to Smartphone at different region of leg was used with Wi-Fi connectivity. They used only two patients data for classification between abnormal and normal [7]. The fall detection system real-time used only 10 PD patient′s data for building supervised learning model and they used external wearable sensors

[8]. All these methods are using external wearable sensors which requires a lot of power, if we have smart phones with all relevant sensors then we can save the wireless communication power for transmitter-receiver and external power supply. The patient can have different diseases which can affect gait parameters differently using a single kind of patient for classification purposes can lead to misclassification when a new kind of patient arrives. The extraction of gait parameters using signal processing is also emerging in recent years. The Short Time Fourier Transform ($STFT$) and Wavelet Transform (WT) were used to measure Step and Stride frequency but it was also using external shoe worn sensors and power supply in conjugation with Smartphone [4]. $Autocorrelation$ and DTW (Dynamic Time Wrapping) are used to find similarity between two signals. Patients with lower back pain there asymmetry between both legs or unilateral problems, autocorrelation and $DTW4$ in conjugation can be used to measure similarity between steps of both the legs in real-time [10] but it fails to extract all the relevant parameters of gait and is useful for only unilateral patients.

Using Smartphone embedded sensors is to compute gait parameters in real-time has no power related issues but similarity between patients is as issue for real-time as patient can be at any part of globe with any kind of disease affecting any region. So, method should be robust in real time but 20 patients with lower back injury for classification [1] in real time need to be robust in detecting for any other kind of abnormality. These all data can be highly correlated for patients with similar disease.

Mean step length and velocity are relevant parameters for abnormality detection [1]. If we can measure step length and associated speed of subject then they can used for our classifier to detect abnormality. Step Length (SL) is a relevant feature to detect the gait of patients affected by motor disorders such as Freezing of Gait [9]. Stance phase i.e. the amount of time person spends in contact with ground in one gait cycle is 60% of gait cycle and Swing phase i.e. the amount of time person spends in air in one gait cycle is 40% of gait cycle for normal people [5]. The phone placement if we use Smartphone′s inbuilt sensors can be placed at knee, ankle or lower back of subject, we can place the sensors at any convenient position of patient [8].

1.4 Terminology

Distance (spatial), Time (temporal) variables are generally the parameters associated with gait cylce [5]. These parameters are defined below:

1.4.1 Distance Variable

1. Step Length [SL] - Distance between corresponding successive parts of heel contact of the opposite feet. For people having normal gait cycle right SL equals left SL [5].

2. Stride Length [STR] - Distance between successive parts of heel contact of the same foot. For people having normal gait cycle it is double the SL for any foot [5].

1.4.2 Time Variable

1. Step Time [ST] - The amount of time spent during single step. It is time elapsed between heel strike of one foot and heel strike of opposite foot [5].

2. Stride Time [STT] - The amount of time it takes to complete one stride [5].

1.4.3 Derived Variable

1. For a particular dataset. Number Of Steps [N] - The no of steps is the total number of steps to complete a fixed distance.

2. Cadence [C] (N/min) - Number steps per unit time.

3. Stride Speed [SS] ($Meter/min$)-

$$SS = \frac{C * STR}{120}$$

1.4.4 Stance and Swing phase :

1. Stance Phase($SP\%$)-It is the period of gait cycle during which the reference feet remains contact with ground [5].

2. Swing Phase($SWP\%$)-It is period of gait cycle during which reference feet remains in air [5].

1.5 Related Work

In Section 1.3 we discussed limitations of using smart-shoe having hardware sensors embedded in it, as it requires constant power supply and uses Bluetooth or Wi-Fi for connectivity with SmartPhone. We are using inbuilt sensors of smartphone using AndroSensor application for data collection purposes. We are using only accelerometer in Smartphone to extract gait parameters. We have collected data for wide diversity of patients which has three classes mild, moderate and severe for supervised learning model preparation. The gravitational acceleration (g) which can be computed from accelerometer data was used to build our method of extraction of gait parameters. AndroSensor Mobile Application was used to directly take the values of gravity in all the axes(x, y, z) of smartphone while walking. These gravity dataset with time was analyzed in MATLAB to extract all the gait parameters. We made a dataset of 185 observations containing all these gait parameters with the class of patient under the supervision of physiotherapist whether they are having mild, moderate, severe abnormality or they are normal. At last we used classifiers NB, RF, DT for classification purposes we used 10-$fold$ CV for adding randomness to the testing set. Our model is trained with not specific kind of disease but with diversity of diseases and only accelerometer is used for building our System.

We are going to use smartphone as platform to collect data and sense data using its inbuit acclerometer making it cheaper in terms of cost and power requirements.

CHAPTER 2

System Overview

Our system has six steps which we used to classify the patient data. Initially, we collected gravity components along all axes using embedded accelerometer in smartphone with the help of AndroSensor application. AndroSensor gives GUI for raw accelerometer, gravity and linear accelerarion using filters. With support of AndroSensor we started recording the data when patient started walking and when patient stopped we turned off the recording.

Variation along gravity y and gravity z was less in comparison with gravity x component for each and every user. So we extracted gait parameters in MATLAB using gravity along x-axis of the smartphone.

For each subject all the essential gait parameters such as *stride_length*, *step_length*, *swing_ratio*, *stance_ratio*, *stride_time*, *step_time*, *stride_speed* and *cadence* were extracted in MATLAB. All subject data after extraction of parameters were recorded in *.csv* file for classification purposes.

We used 3 different classifiers (Naive-Bayes, Decision Tree and Random Forest) for detecting the abnormality using relevant gait parameters. *RF* performed best among the three classifiers with four features and 500 decision trees. We used correlation matrix to remove the intercorreelation associated with the gait parameters and then applied *DT* and *NB* on these parameters for classification.

We marked four classes of patients:

- Normal
- Mild
- Moderate
- Severe

Lastly our 3 classifiers detected the class of patients on testing set prepared with randomly.

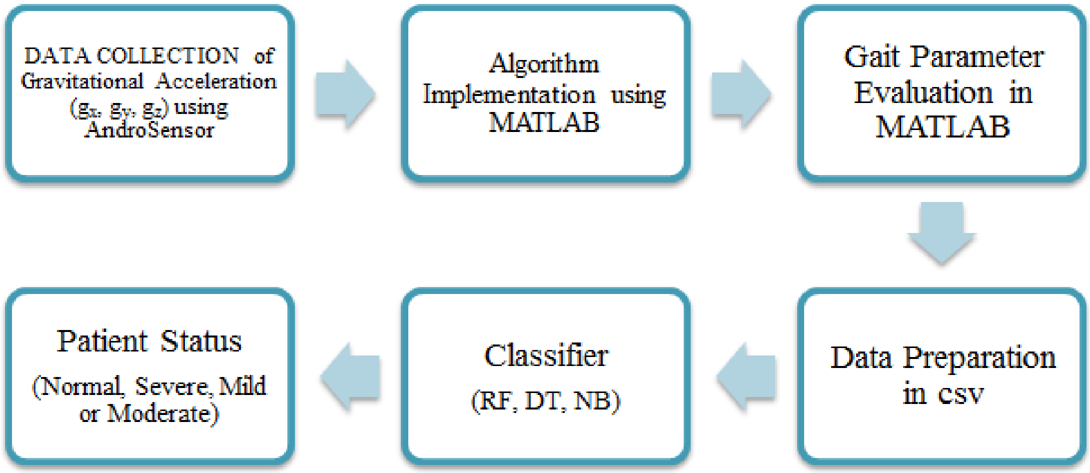

Figure 2.1: System Architecture

2.1 Experimental Methodology

2.1.1 Softwares and Hardware Used

Inbuilt Accelerometer Sensor of Smartphone MOTO G4 Plus and Yu Yuphoria, AndroSensor application for *data collection*, *MATLAB* for extracting *spatio − temporal* gait parameters and *R* for classification purposes. Accelerometer Sensor is generally used to measure the acceleration which is the rate of change of velocity and force produced in it. It works on *MEMS* (Micro Electro Mechanical Systems) technology.

The small size of *MEMS* chip is a boon for embedding it in a smartphone. Accelerometer is a raw sensor available in a smartphone whereas from Android 2.3 onwards synthetic sensors such as gravity are also available which modifies the raw sensor data to factor out the force due to gravity. When the phone is at rest it displays $9.8m/s^2$ in positive z direction. Figure 2.2 shows screenshots of AndroSensor application GUI.

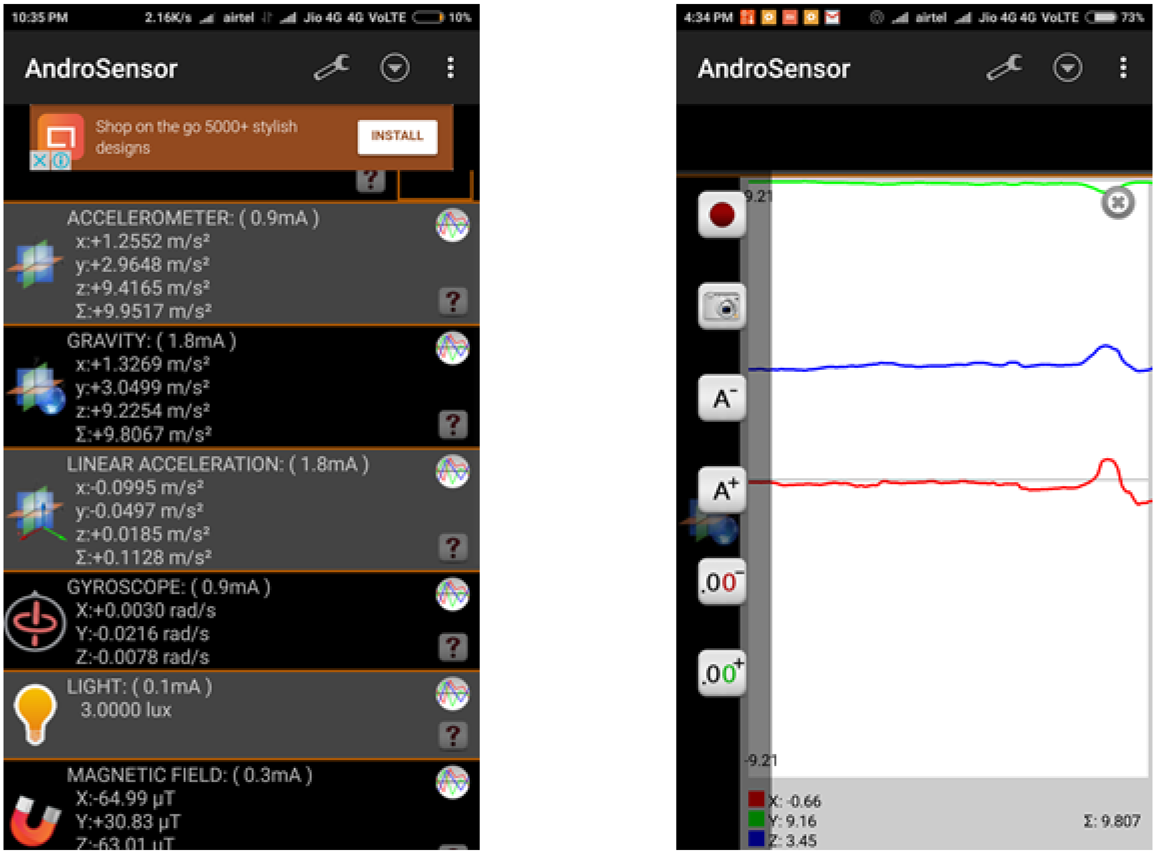

Figure 2.2: AndroSensor App Screenshots showing gravity variations

2.1.2 Data Collection

Smartphone was placed at the ankle of the subject using band for supporting purpose. All the subjects were instructed to walk for a distance of $6m$ normally in straight path. The recording interval was $5ms$ and the sampling frequency is $200Hz$. We started the recording and instructed the patient to walk and when they finish $6m$ walk test then we stop the recording. Depending on patients severity the pateint was allowed to walk 2, 3 and 4 rounds from both legs i.e. left leg and right leg respectively. Total we collected data which has 185 observations from 17 normal and 18 patients. All the patients and normal people were between the age of 20-60 except 1 who was 60. Out of 18 patients, 5 were male and 13 were female.

We collected data of 18 patients suffering from *RA*, *Ostrio Arthritis*, *Heel Pain*, *Heel Spur*, *Sciatica*, *Rotation of knee*, *Ankylosings Spondilitis*, *Polio*, *Knee* and *Motor injury* and *Lower back pain*. These data were collected at physiotherapist centres and clinics under supervision of physiotherapists and doctors. The class associated with patient abnormality in both the legs was marked mild, severe and moderate by consulting the physiotherapists and doctors which was helpful for training the supervised models.

Figure 2.3: Smartphone Placement at ankle

Name	Age	Sex	Disease	Left Leg	Right Leg
Arvind Patel	43	Male	Rheumatoid Arthritis	Moderate	Moderate
Geeta	49	Female	Ostrio Arthritis	Severe	Severe
Deepika	47	Female	Ostrio Arthritis	Severe	Severe
Monika	23	Female	Heel Pain in Right leg	Svere	Severe
Parul	33	Female	Left Leg sciatica	Mild	Mild
Kanta Patel	48	Female	Left knee problem	Severe	Severe
Raj Kumar Sharma	31	Male	Rotation of knee right	Moderate	Severe
Ritu	33	Female	Polio	Severe	Severe
Arvind Singh	41	Male	Rheumatoid Arthritis	Moderate	Moderate
Harshit Pratik	27	Male	Ankylosings Spondilitis	Mild	Normal
Sushilaben Patel	47	Female	Abnormal walking	Moderate	Moderate
Shakuntla Behen	40	Female	Heel Spur	Normal	Mild
Harshna	49	Female	Rheumatoid Arthritis	Mild	Normal
Pooja	40	Female	Rheumatoid Arthritis	Moderate	Mild
Jv Soni	60	Female	Rheumatoid Arthritis	Severe	Severe
Sonali	24	Female	Knee injury	Mild	Normal
Harsh	23	Male	Motor injury	Mild	Normal
Ankit	40	Male	Knee and Back pain	Mild	Mild

2.2 Phone orientation and gravity(g) variation

As we know that, one gait cycle starts at heel strike and ends at heel strike. When subject is instructed to walk then in one gait cycle the change in orientation and axes of phone is as depicted in below figure.

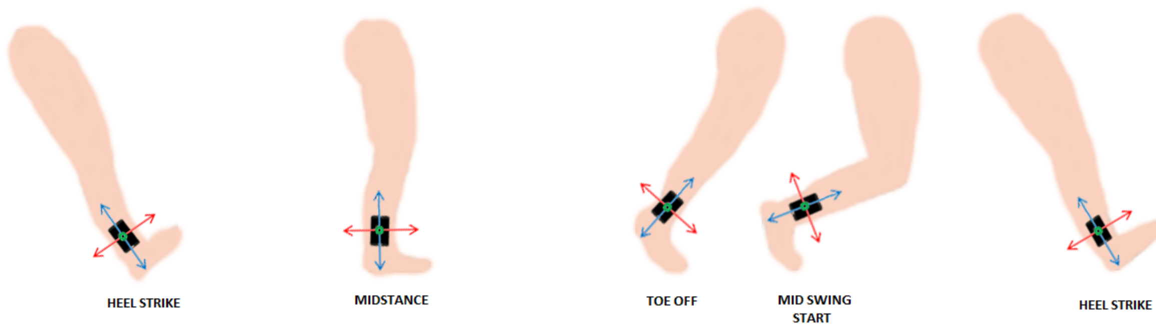

Figure 2.4: Phone orientation changes in Gait cycle

Figure 2.5 depicts the phone′s variations in gravity during one gait cycle.

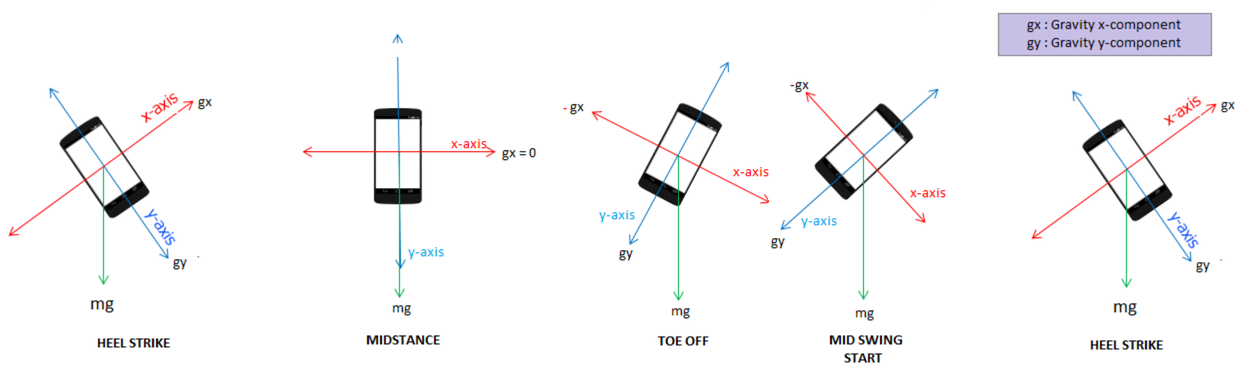

Figure 2.5: Gravity components and weight(mg) acting on phone during gait cycle for subject′ Right Leg

$$Weight = mass * gravity$$

Weight is a force due to gravity which acts on centre of gravity towards ground vertically. So in figure 2.5 weight acts on phone during walking in downward direction towards ground. The gravity sensor shows the forces measured by accelerometer that are caused only by gravity. During 1 gait cycle g is downwards but as the phone orientation changes the axes changes and so does the components of gravity along each axis i.e. x and y. Along z there is a little variation as subject is constrained to walk in straight path not in zig-zag way.

$$\vec{g} = \vec{g_x} + \vec{g_y} + \vec{g_z}$$
$$g = \sqrt{g_x^2 + g_y^2 + g_z^2} = 9.8m/s^2$$

At heel strike the phone gains maximum gravity component in positive x direction as shown in figure 2.5 as the tilt of phone is maximum in positive direction with reference to ground, while the g_y will obtain second minimum as since g remains constant.

At start of mid-swing phase the phone will have maximum gravity component along negative x direction 2.5 as the tilt of phone is maximum in negative direction with reference to ground, while the g_y will obtain its minimum value as g remains constant. At mid-stance the g_x will have zero crossing and as the force mg and y axis of phone are in same direction as shown in Figure 2.5. So, g_y will be maximum 9.8 m/s^2 and g_x will be nearly equal to zero. At toe-off the value of g_x will be negative as shown in Figure 2.5. All these terms are shown in Figure 2.6

Note: These all are valid when we are using right leg as reference, whereas while considering left leg positive peaks are replaced by negative peaks and vice versa. This is due to change in axes i.e. positive x-axis changes to negative x and vice versa.

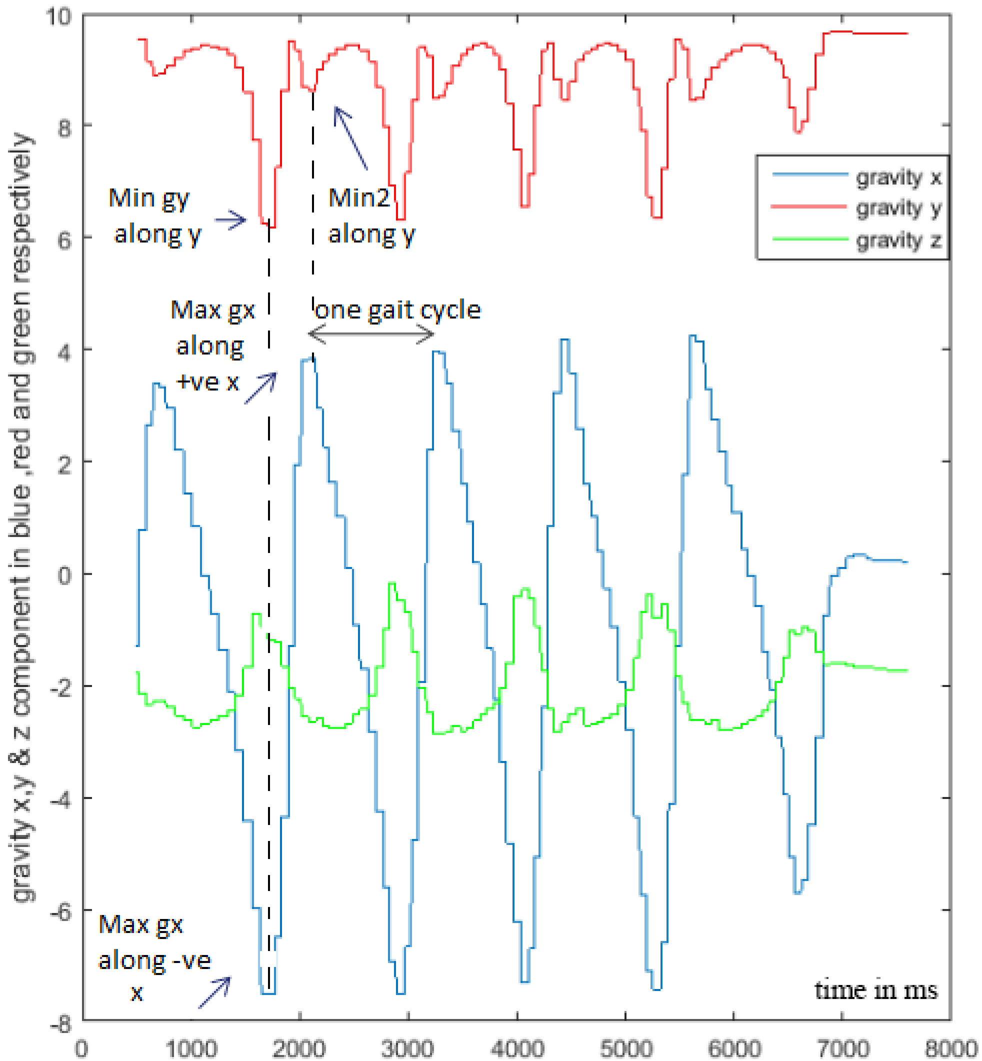

Figure 2.6: g_x and g_y maxima and minima with time

2.3 Hypothesis

Heel Strike is identified through maximum positive peak with right leg as a reference and maximum negative peak with left leg as a reference in g_x plot with time. Mid stance is identified through zero crossing of g_x plot with time. Start of mid-swing is identified maximum negative peak with right leg as a reference and maximum positive peak with left leg as a reference. The detailed explanation of different stages of gait cycle and variation in g_x, g_y and g_z is provided in Figure 2.7.

	Heel Strike	Foot Flat	Mid Stance	Heel Off	Toe Off	Mid Swing-1	Mid Swing-2	Heel Strike
g_x	Max-positive (Peak)	Positive but decreases.	Ideally Zero.	Negative and increases in negative direction.	Increases in Negative direction.	Max-peak in negative direction.	Nearly zero.	Max-positive (Peak).
g_y	Second Minimum. .	Increases	Max(9.8)	Decreases	Decreases	First Minim um	Max(9.8)	Secon d Minim um
g_z	Almost constant	Almost constant	Almost constant	Almost constant	Almost constant	Almost constant	Almost constant	Almost constant
	MAX (+ve) Peak to MAX (-ve) Peak					MAX(-ve)Peak to Max(+ve)		

Figure 2.7: Method to extract Gait Parameters

So we can identify heel strike and start of mid-swing through plot of g_x vs time for both left leg and right leg. As user is constrained to move in straight path the variations along z is insignificant and almost constant. The identification of stages for left leg and right leg are shown in plot below:

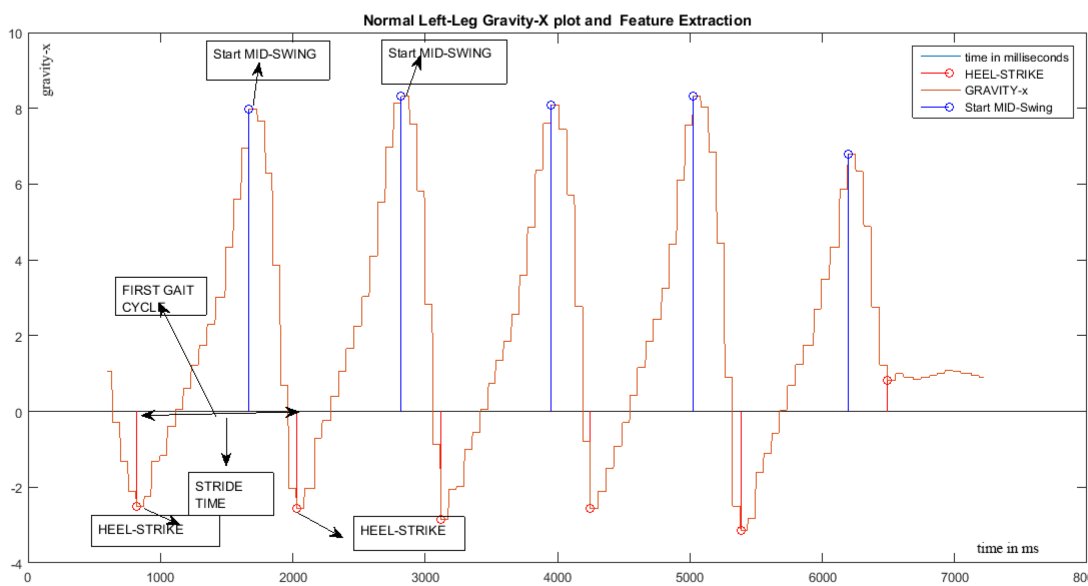

Figure 2.8: Heel Strike and Mid-swing start identification through g_x plot for left leg

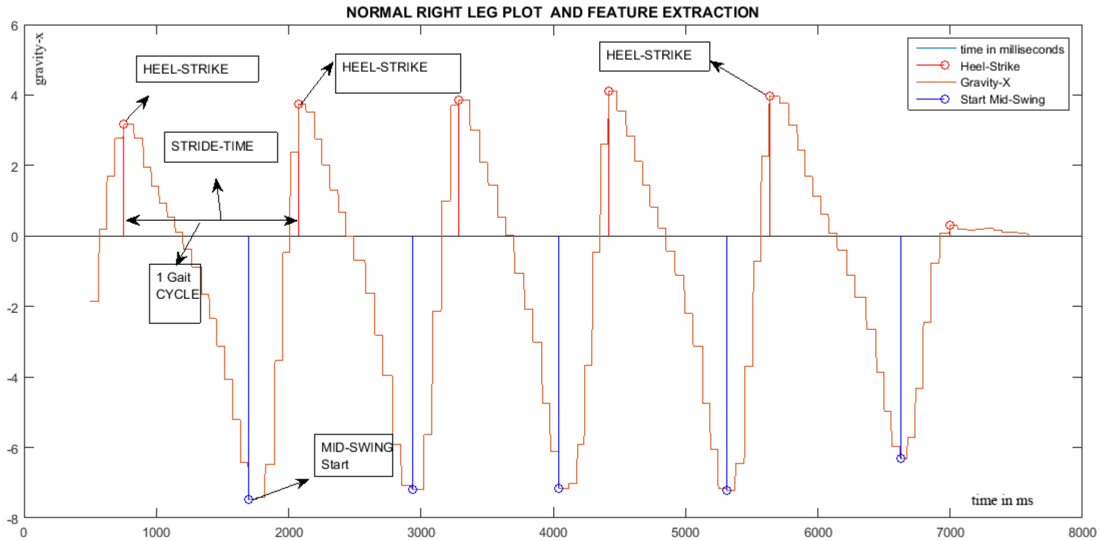

Figure 2.9: Heel Strike and Mid-swing start identification through g_x plot for right leg

2.4 Proposed Algorithm

For both right and left leg as it is assumed that subject starts with heel strike and ends with heel strike. So total number of positive peaks(heel strike) will be one more than total number of negative peaks (start mid-swing). So we are iterating over total number of negative peaks to get differences between heel strikes.

Algorithm for analysis of right leg

Input: g_x values for a fixed distance d of $6m$ and time is in milliseconds(ms)

Variables used: d = $6m$
tt = total time taken for $6m$ walk (in ms) = *finish_time - start_time*

Output: Values for various Gait Parameters.

Solution.

for $i = 1 : \text{length}(save1)$ //save1 array stores +ve peaks from input g_x
$stride\_time[i] = save[i+1] - save[i];$ //array to store time between two +ve peaks
end

$stride\_length = (d/tt) * mean(stride\_time);$
$step\_length = stride\_length/2;$
$step\_time = stride\_time/2;$
for $i = 1$: length($save2$) // save2 array stores negative peaks from input gx
 $d1[i]$ = abs($save2[i]$ - $save1[i]$); //array to store time interval b/w +ve to -ve peak
 $d2[i]$ = abs($save2[i]$-$save1[i+1]$); //array to store time interval b/w -ve to +ve peak
 $d3[i]$ = $save1[i+1]$ - $save1[i]$; // gait cycle time i.e. time interval between two heel strikes
 $d4[i]$ = 0.13 * $d3[i]$; //13% of gait cycle
 $d5[i]$ = $d2[i]$ + $d4[i]$; // swing phase time
 $d6[i]$ = $d3[i]$ - $d5[i]$; // stance phase time for identifying toe-off
end
$stance\_ratio$ = mean($d6$) / (mean($d5$) + mean($d6$));
$swing\_ratio$ = mean($d5$) / (mean($d5$) + mean($d6$));
$number\_of\_steps$ = length($save1$) + length($save2$); //sum of total number of +ve and -ve peaks
$cadence$ = ($number\_of\_steps$ / tt) * 60000;
$stride\_speed$ = $cadence$ * mean($step\_length$);

Note: For left leg we are iterating over total number of positive peaks and here negative peaks (heel strike) are one more than total number of positive peaks(start mid-swing). So $save1$ in left leg is storing all the values of negative peaks time instance and $save2$ is storing value of positive peaks time instance. All the steps of the algorithm for analysis of left leg remains the same, except these two parameters (save1 and save2).

The above algorithm for detecting swing phase requires detection of toe-off time instant but we can detect only start of mid-swing. So we set 13% of total gait cycle time which is added to the time between start of mid-swing (-ve peak) to heel strike (+ve peak) to detect toe-off point. As in [5] toe-off is 10-15% before start of mid-swing. So we took average of 10 and 15% to detect toe-off, which is approximately 13%. Stride length can form a good decision boundary in comparing normal with moderate and severe class where as Swing Ratio can also form good decision boundary between normal and mild abnormal subject.

CHAPTER 3

Statistical Analysis

In this chapter we discuss the plots associated with different classes of patients and their comparison. Also, we discuss the average values of gait parameters for every class. Then we mention the frequency of occurrence of parameters for each class i.e. mild, normal, moderate and severe. We have used correlation based feature selection for applying to NB classifier as all the features under NB are assumed to be independent. Also we have mentioned the performance associated with DT, NB and RF classifiers.

3.1 Plots

In this section we describe the various associated plots with different classes and comparison is made between the plots. Here for each class of patients we present plots for both legs. The plots are given below:

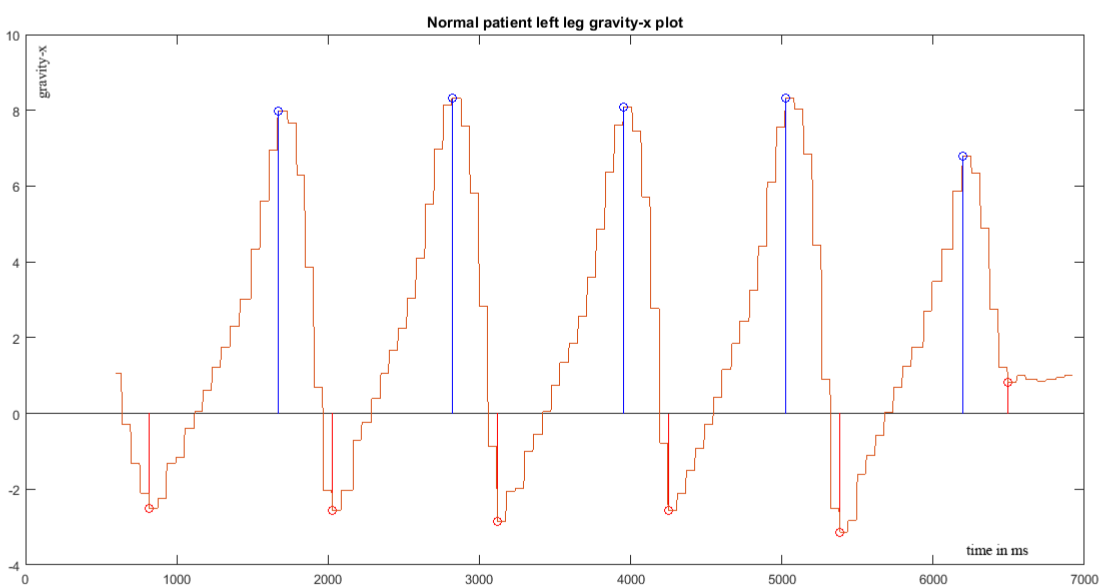

Figure 3.1: g_x plot for normal pateint left leg

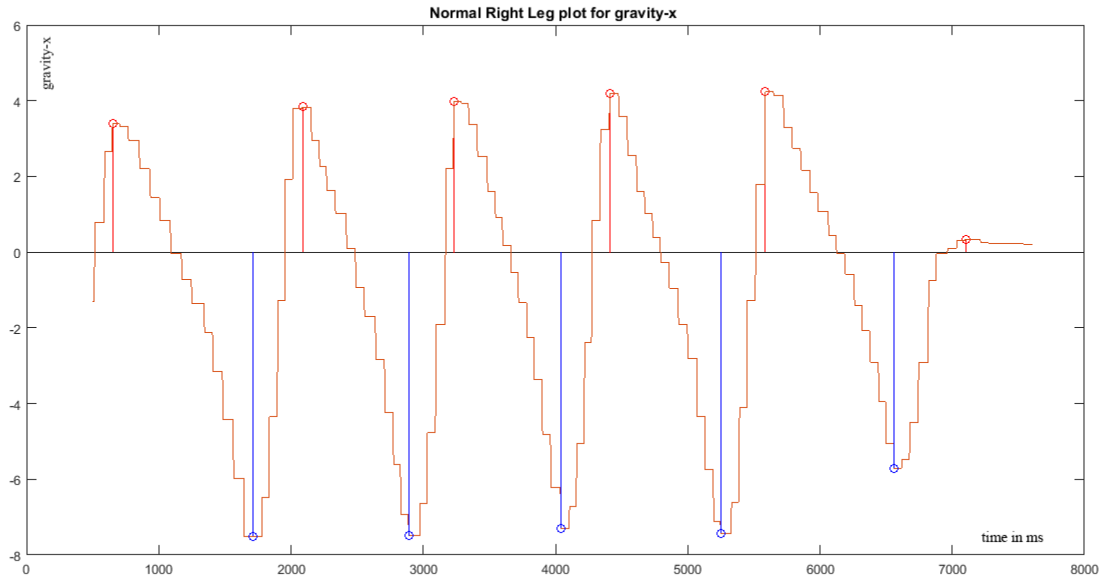

Figure 3.2: g_x plot for normal pateint right leg

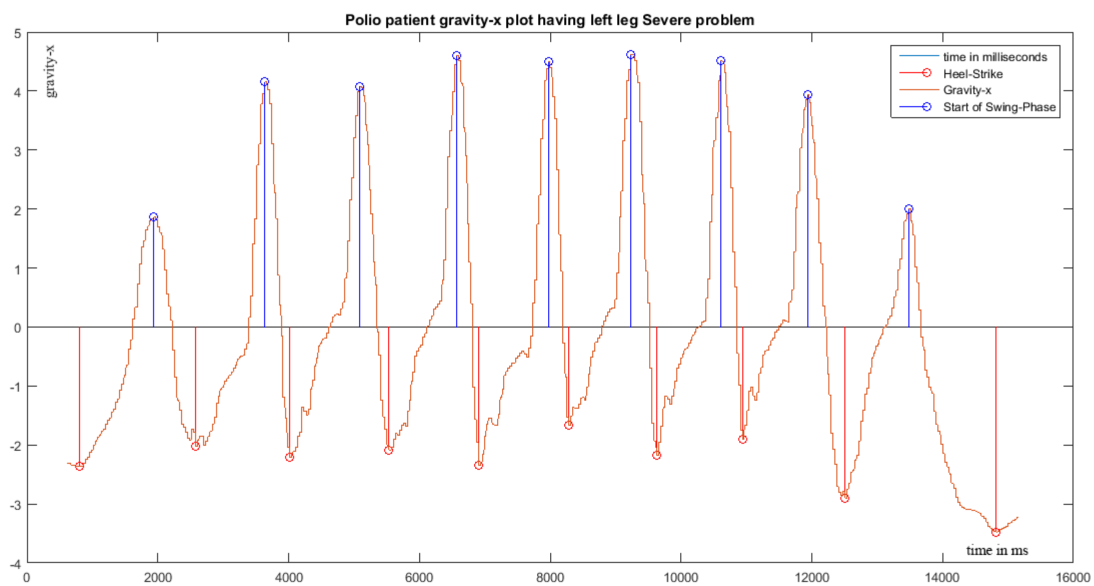

Figure 3.3: g_x plot for severe pateint left leg

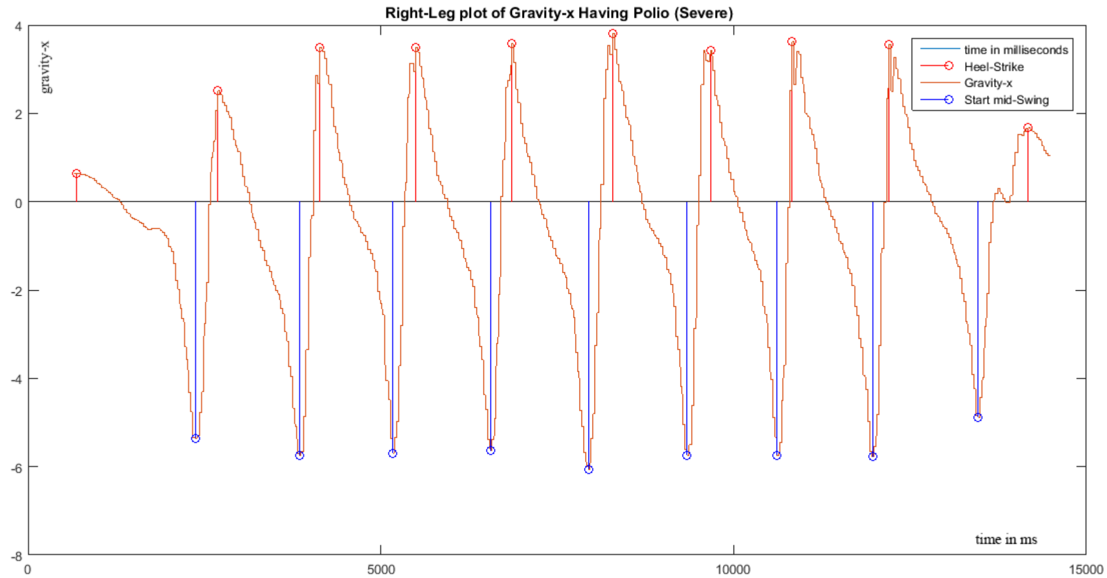

Figure 3.4: g_x plot for severe pateint right leg

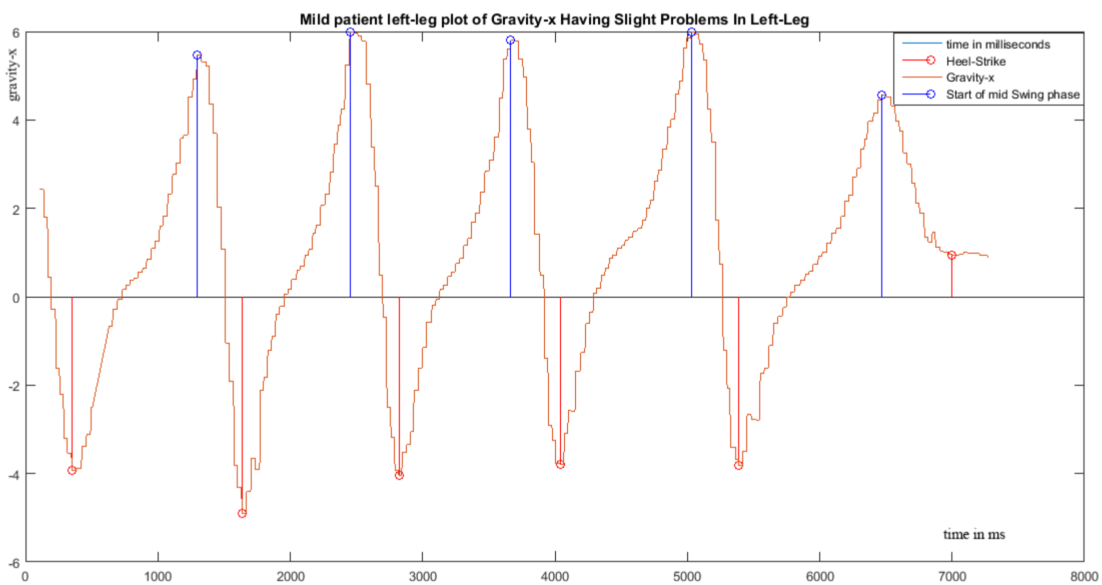

Figure 3.5: g_x plot for mild pateint left leg

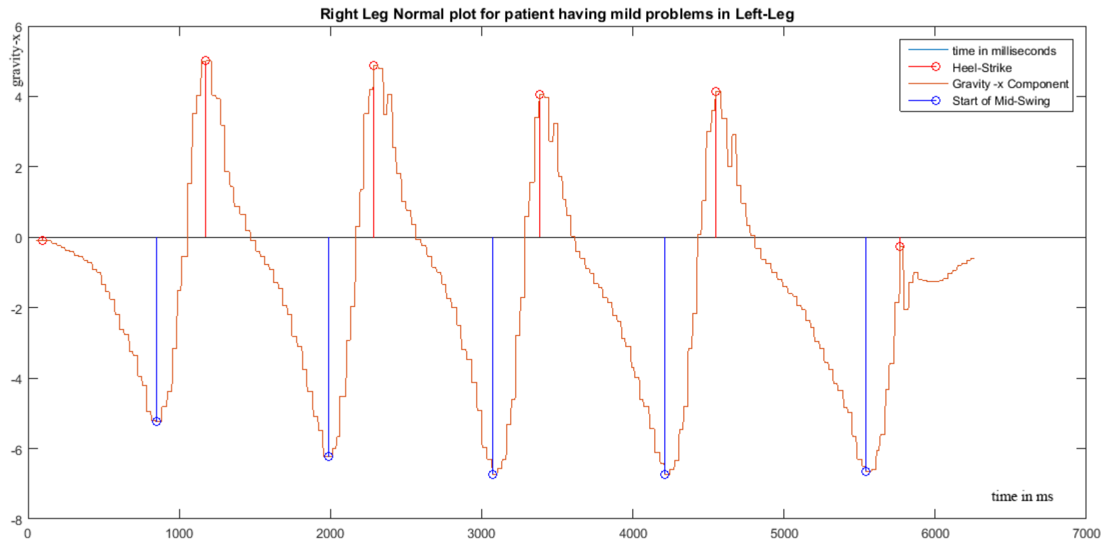

Figure 3.6: g_x plot for mild pateint right leg

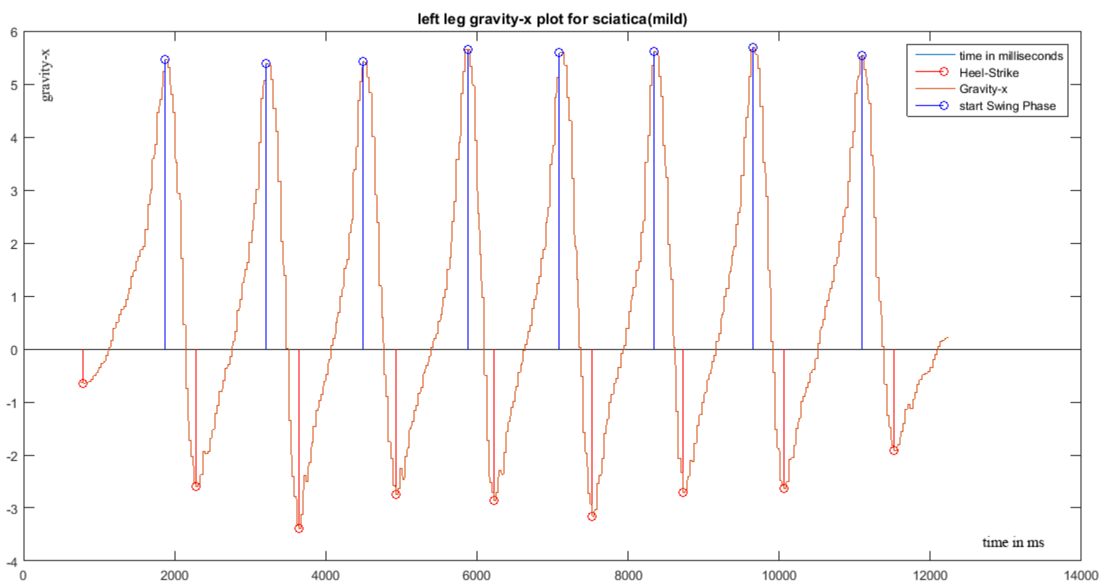

Figure 3.7: g_x plot for mild pateint left leg

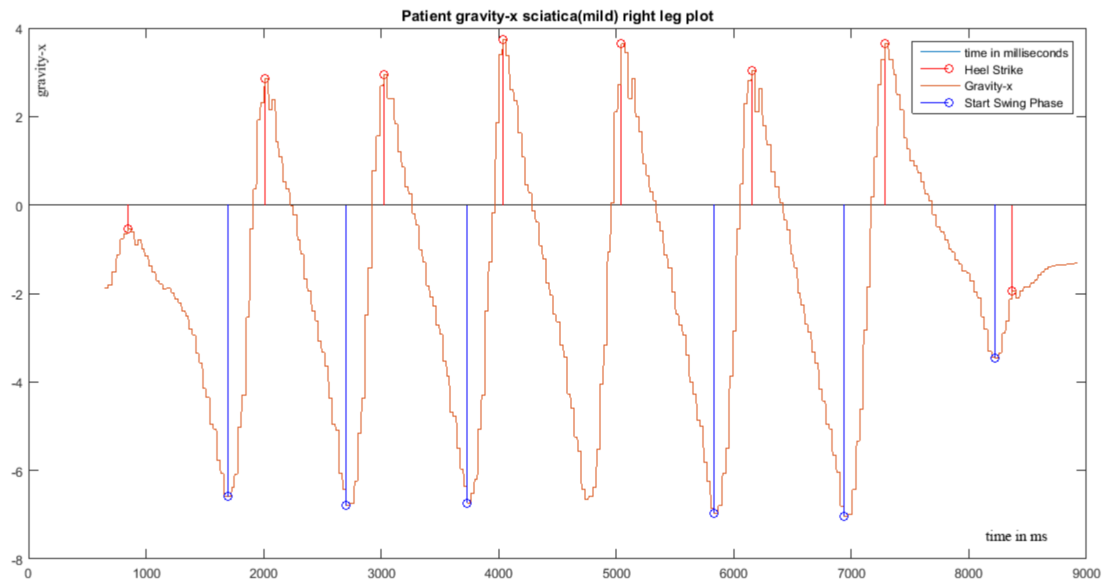

Figure 3.8: g_x plot for mild pateint right leg

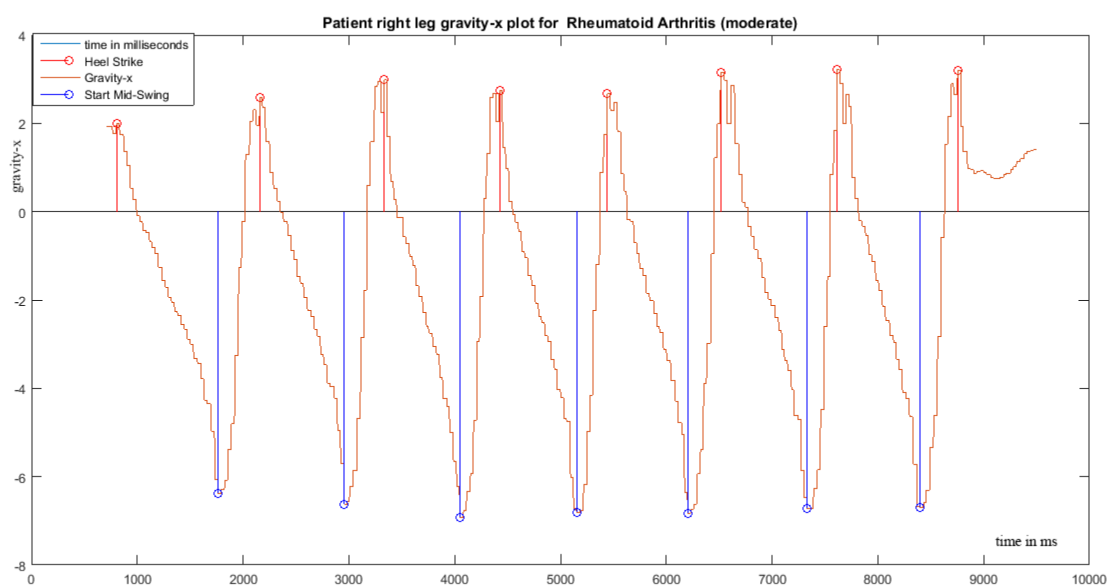

Figure 3.9: g_x plot for moderate pateint right leg

Plots Comparision

Number of steps which is sum of positive and negative peaks is more for severe patients in comparison to normal, moderate and mild patients. This proves that Severe patients takes more no of steps to complete $6m$ walk test which means their SL reduces and STR also reduces i.e the distance covered between one heel Strike to another heel Strike. The total time for finishing 6 meter walk test is also increased (Figure 3.5). Moderate patients usually take less number of steps to complete 6*m* walk, compared to severe patients but it is more than normal patients (Figure 3.9). Mild patients usually are kind of similar to normal class (Figure 3.7) as they take same number of steps to complete the 6m walk test, but some mild patients take few more steps than normal person (Figure 3.5). In moderate class and in mild class number of steps are same (Figure 3.9 and Figure 3.7). So we conclude that number of steps is not a good feature to differentiate between normal, mild and moderate patients. (Figure 3.17) Number of steps to complete 6m walk may depend on height of subject also.

3.2 Average of Gait Parameters and Histograms

We made a file containing all the parameters extracted from MATLAB for both legs for every patient with his class marked. Then, we computed median, mean and mode for each and every parameter and class.

Average stride speed(in $meter/min$) of normal class is highest whereas it is least for Severe class which is obvious. The stride speed of mild class and moderate class are overlapping and mild and normal class are also overlapping. It gives clear indication that stride speed alone cannot differentiate between our classes of patients. (Figure 3.14)

Cadence (in N/min) is directly proportional to number of steps and inversely proportional to total time for completion of walk. As normal and mild subject takes less no of steps to complete 6 meter walk in shorter time of walk whereas patients(moderate and Severe) takes more no steps in longer total time for walk, cadence values overlap for all the classes of patients.

Average Stride length(in cm) is maximum for normal patients and its mean differs from all other classes, giving indication that it can form a good feature to distinguish among the classes. (Figure 3.15)

Average Stride time(in *milliseconds*) also ovelaps for all the classes.

In severe class, the maximum swing ratio (48) was observed in left leg of a patient suffering from polio because she could not put more weight on left leg and she had to put more weight on her right leg. So, during a gait cycle her swing ratio was more for left leg and Stance ratio of the other (right) leg was increased which was observed to be 65.37.

The mean Swing Ratio for normal and mild class was forming a decision boundary which indicates that mild patients while walking were putting less weight on the leg which had problem, which means their feet were more in air as compared to normal people. (Figure 3.16) So, Swing ratio can be used to distinguish between normal and mild patients.

Similarly, Stance phase which is a phase of gait cycle in which reference leg is in contact with ground. It is decreased for mild class compared to normal because swing phase increases ($SP\%$ = 100 - $SWP\%$). For normal class in one gait cycle, 60% is swing phase and 40% is stance phase [5].

Patient Class	Mean ± SD	Maximum	Minimum	Median
Stride Speed	56.487 ± 7.695	39.43	73.97	56.77
Cadence	100.66 ± 7.3	84.9	119.97	101.14
No of Steps	10.66 ± 0.97	9	15	10
Stance Ratio	58.22 ± 1.69	51.94	63	58.08
Stride Length	111.81 ± 9.99	77	130	113
Stride Time	1177.758 ± 87.06	1008	1412	1162
Swing Ratio	41.76 ± 1.69	37	48.06	41.92

Figure 3.10: mean and sd of gait parameters of normal class

Patient Class	Mean ± SD	Maximum	Minimum	Median
Stride Speed	28.37 ± 4.02	37.45	22.05	27.39
Cadence	88.89 ± 8.71	103.77	77.95	87.09
No of Steps	18.5 ± 1.58	21	17	19
Stance Ratio	60.82 ± 2.98	65.37	55.89	60.85
Stride Length	62.72 ± 6.36	72	52	64
Stride Time	1332.1 ± 131.72	1524	1118	1342
Swing Ratio	40.42 ± 3.45	48	34.63	40.28

Figure 3.11: mean and sd of gait parameters of severe class

Patient Class	Mean ± SD	Maximum	Minimum	Median
Stride Speed	44.02 ± 7.7	59.9	34.18	40.95
Cadence	96.26 ± 7.22	110.16	83.88	96.80
No of Steps	13.03 ± 1.55	15	11	13
Stance Ratio	55.28 ± 5	62.63	33.55	55.54
Stride Length	91.44 ± 12.69	114	77	88
Stride Time	1228.18 ± 108.08	1453	1028	1211
Swing Ratio	44.71 ± 5	66.45	37.37	44.46

Figure 3.12: mean and sd of gait parameters of mild class

Patient Class	Mean ± SD	Maximum	Minimum	Median
Stride Speed	36.205 ± 2.56	39	29.77	36.7
Cadence	90.5 ± 7.17	101.7	71.66	90.94
No of Steps	14.56 ± 0.81	15	13	15
Stance Ratio	60.77 ± 1.89	63.79	57.1	60.88
Stride Length	80.31 ± 5.52	89	72	79.5
Stride Time	1316.68 ± 129.86	1706	1181	1302
Swing Ratio	39.22 ± 1.89	42.9	39.21	39.12

Figure 3.13: mean and sd of gait parameters of moderate class

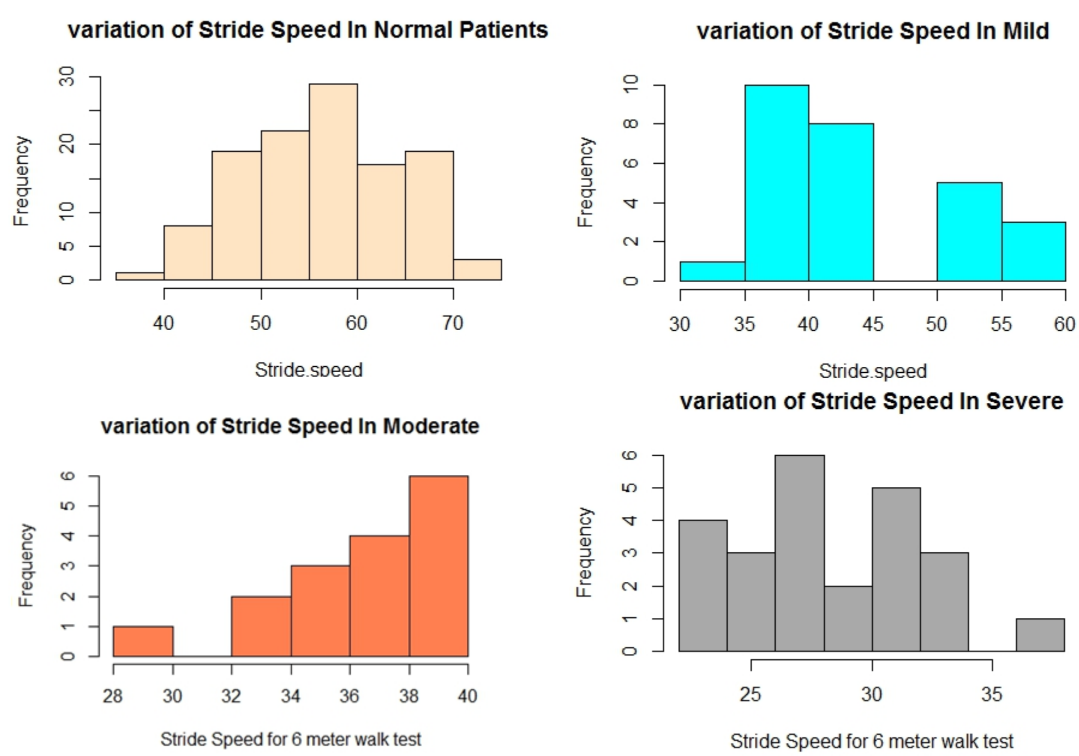

Figure 3.14: Variation of stride speed in all classes

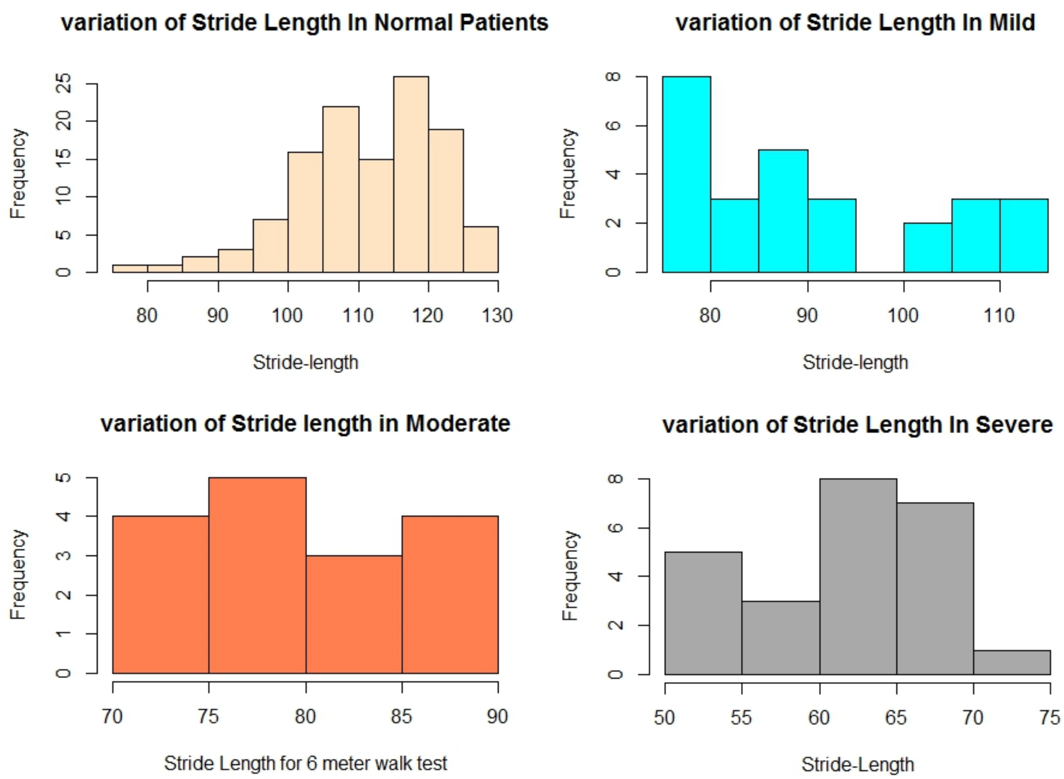

Figure 3.15: Variation of stride length in all classes

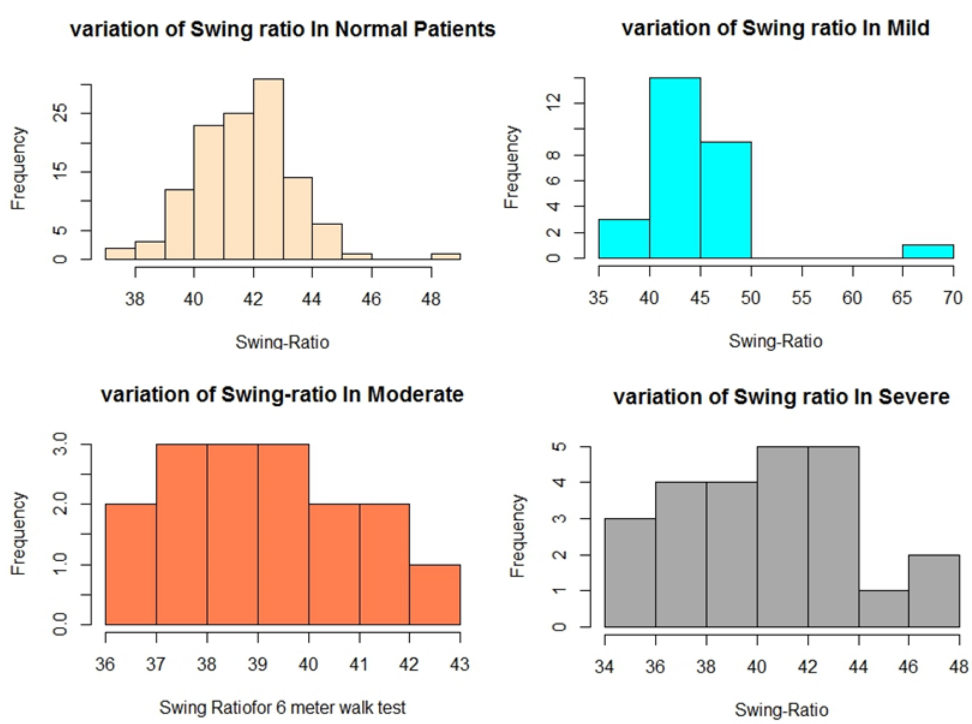

Figure 3.16: Variation of swing ratio in all classes

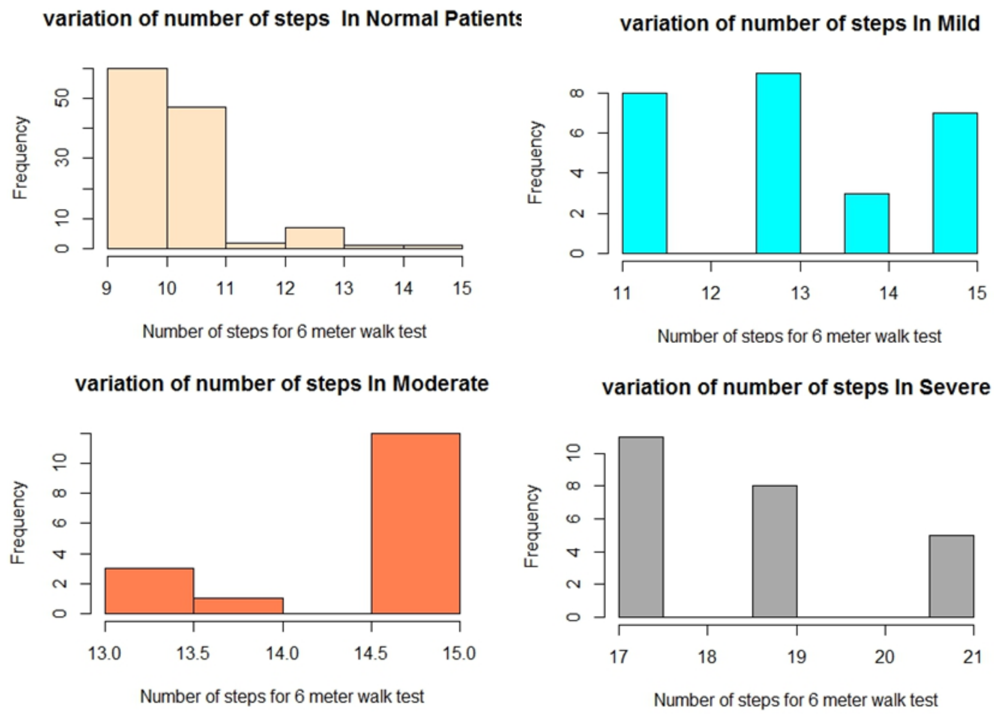

Figure 3.17: Variation of number of steps in all classes

3.3 Correlation based feature selection

The need of feature selection is listed below :

- Reducing the number of features makes supervised model more interpretable.
- It avoids overfitting of the model
- For better perception of feature relation with response variable
- Some of the classifiers like NB, LDA and QDA uses correlation based feature selection as they assume that features are independent with each other i.e. no inter-correlated features are used to train the model.

For a given sample of data with two features x_i and y_i for n observations, r which is the correlation coefficient is given by the following formula:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

where: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Now coming on to our dataset we have total 8 features i.e. Stride Speed, Stride Length, Cadence, Number of Steps, Swing Ratio, Stance Ratio, Stride Time and Status of patient which is the output as it has four levels (Moderate, Normal, Severe and Mild). To get a correlation Matrix first parameter is compared with all the remaining parameters. Similarly, for other parameters also iteratively. Finally, in R we get the correlation matrix which describes the amount of correlation for one feature with others. So we performed correlation test for our dataset. Results are as shown below Figure:

	Stride Speed	**Cadence**	**Number of Steps**	**Stance Ratio**	**Stride Length**	**Stride Time**	**Swing Raio**	**Status**
Stride Speed	1	0.76	-0.89	-0.3	0.96	-0.64	4.2	-0.3
Cadence	0.76	1	-0.51	-0.34	0.58	-0.86	0.24	-0.13
Number of Steps	-0.89	-0.51	1	0.3	-0.96	0.52	-0.19	0.32
Stance Ratio	-0.30	-0.34	0.30	1	-0.28	0.35	-0.9	-0.31
Stride Length	0.96	0.58	-0.96	-0.28	1	-0.5	0.16	-0.33
Stride Time	-0.64	-0.86	0.52	0.35	-0.50	1	-0.26	0.10
Swing Ratio	0.20	0.24	-0.19	-0.90	0.16	-0.26	1	0.36
Status	-0.30	-0.13	-0.32	-0.31	-0.33	0.10	0.36	1

Table 3.1: Correlation Matrix

The value of r can be 0, 1 or -1. 1 means the output variable is perfectly positive correlation whereas the value 0 indicates no correlation with output and -1 indicates perfectly negative correlation with the output. The value greater than $|0.7|$ indicates high correlation whereas value between $|0.5| to |0.7|$ indicates moderate correlation with the output. However if the value of r is less than $|0.5|$ and greater than $|0.3|$ then it shows that it is weak positive/negative correlation. As status of patient is 0 for "normal", 1 for "mild", 2 for "moderate" and 3 for "severe" negative correlation indicates that as the value of variable increases the tendency is of being normal and vice versa.

From Table 4.4 features which show weak correlation with the output are Stride Speed, Swing Ratio, Stance Ratio, Number of Steps and Stride length. This gives clear interpretation that no one attribute can strongly predict the output. We have to take the variables that are not highly intercorrelated. So in our case we cannot take stride speed, Stride length and number of steps together as parameters as they are highly intercorrelated.

There are lot of interdependencies among the the features as one parameters is derived from other we want the model should contain less number of uncorrelated features to avoid overfitting during training to make model more interpretable. Stride Length and Swing Ratio are the two features highly correlated with status of pateint with less intercorrelation.

CHAPTER 4

Classifiers

We are using pruned Decision Tree, Naive-Bayes and Random Tree classifiers for detecting the status of patients. Total 185 observations out of which 118 were normal, 16 were having moderate abnormality, 24 were having severe abnormality and lastly 27 were having mild abnormality. We splitted our dataset into 80% for training purpose and 20% for testing purpose.

We used 5-fold CV for decision tree and Naive-Bayes classifier, as using 10-fold CV also the learning tree changed two times and confusion matrix was also the same. Also the error rate i.e. missclassification rate was the same for both 5-fold CV and 10-fold CV for Decision Tree and Naive-Bayes. We used 10-fold CV for Random Forest classifiers as it showed more accuracy.

Cross-validation was used to avoid overfitting, deciding size of decision tree and size random forest. No samples in testing sets are repeated using cross-validation providing randomness in testing dataset.

4.1 Decision Tree

Decision tree are very interpretable and its perception is easy. It also signifies the important features. Any Kind of relationship like logarithmic or polynomial between features does not affect decision trees. Pruning in Decision tree is done to avoid overfitting. If pruning is not done then the tree will grow until number nodes in leaf in DT equals to number of observations. This may be good for training and perform excellent on training set but as size of DT grows its performance on new data or validation set decreases.

4.2 Decision Tree Model

We made decision tree model on ($Status \sim Stride\_Length + Cadence + Swing\_Ratio$ + $Stride\_Time$) and pruned the tree on the size of tree having minimum relative error. While iterating 5-fold CV we saw that four times it has least relative error with size 4 (Figure 4.1) and only one time it has least relative error with size 5 (Figure 4.3). We used modified 5-fold CV in which 80% was used for training and 20% was used for testing purposes. We trained our model on four parameters but it took only two parameters Stride Length and Swing Ratio. The first DT model (Figure 4.2) indicates that if $stride\_length$ is greater than 90cm then it should be Normal Class and if it is less than 71cm then it is severe class. While if stride length is greater than 71cm but less than 90cm and if swing ratio is greater than 43% then it belongs to mild class otherwise moderate class.

The second DT model (Figure 4.4) indicates that if the stride_length is greater than 90cm and swing_ratio is less than 44cm then it should be normal class and if the swing_ratio is greater than 44cm then it is mild class. If Stride_length is less than 71cm and swing_ratio is greater than 43cm then it belongs to moderate class but if swing ratio is less than 43cm then it belongs to mild class. If stride length is greater than 73cm and less than 90cm then it belongs to severe class.

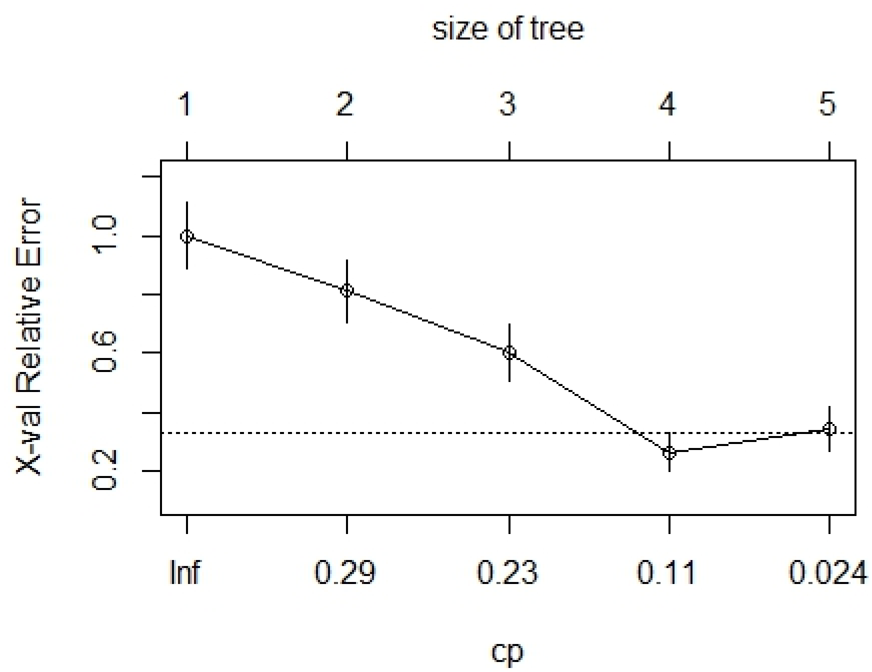

Figure 4.1: Decision Tree model on size 4 with least relative error

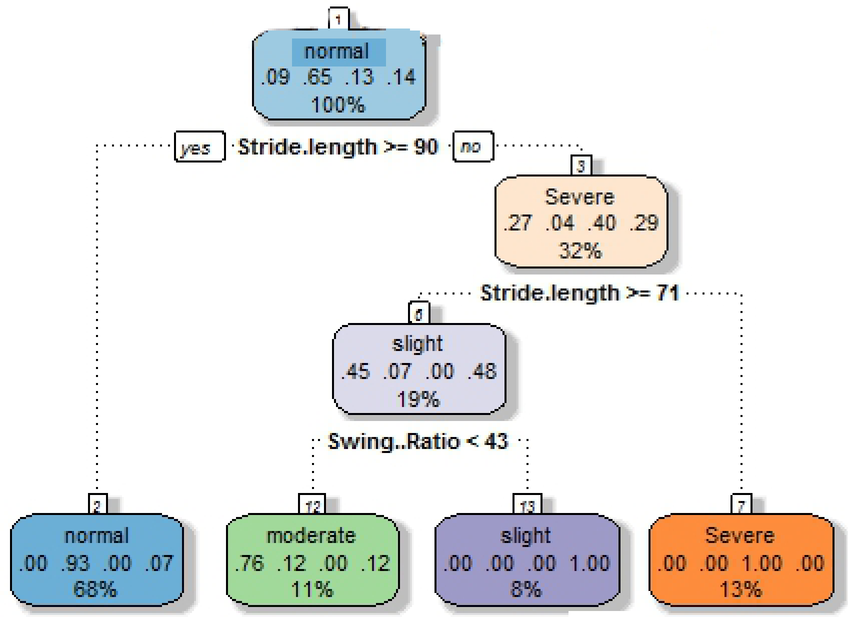

Figure 4.2: Pruned Decision Tree model on size 4

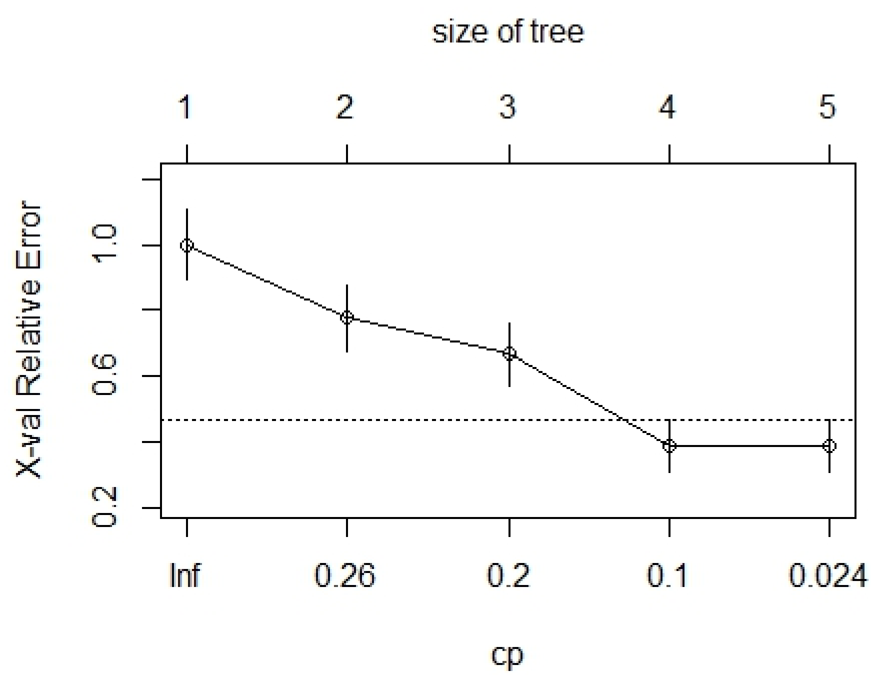

Figure 4.3: Decision Tree model on size 4 with least relative error

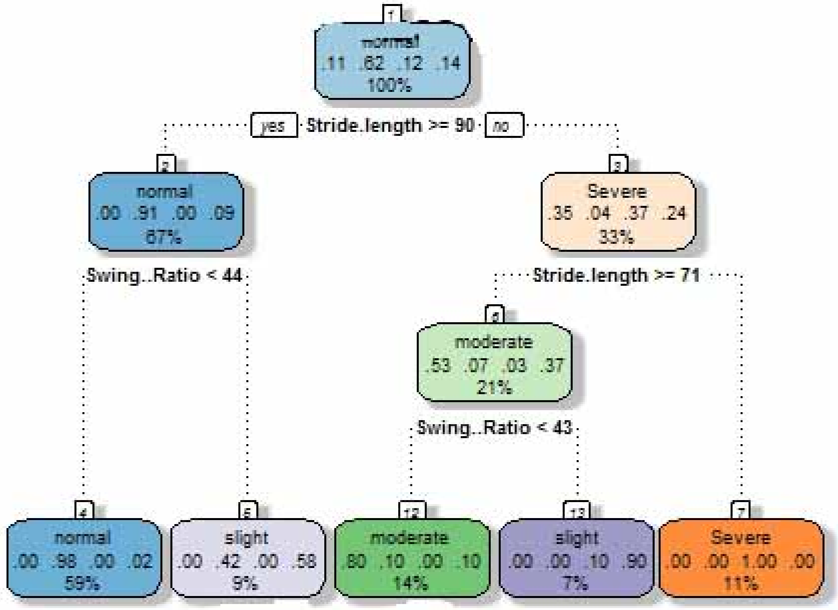

Figure 4.4: Decision Tree model on size 5

4.2.1 Results

Confusion Matrix (CM) shows the amount of misclassification in the testing set. The diagonal elements in the confusion matrix indicate the number of observations were correctly predicted throughout the pruned decision tree models. The non-diagonal elements indicate that their actual class and predicted class are different. The first column shows the actual class whereas the first row shows the predicted class. So each cell (Actual, Predicted) in the CM represents the number of classifications or misclassifcations. The As, we iterate five times we have 5 models of which 4 are same and 1 is different. The first CM (Table 4.1) misclassified 2 of moderate patients as 1 normal and 1 mild and 2 normal patients as mild patients. The second CM (Table 4.2) misclassified 1 moderate patient to normal class and 2 of each normal and severe patient as mild class. The third CM (Table 4.3) missclassified 5 normal patients to mild class. The average accuracy of pruned decision tree model was 89.07% considering misclassification in testing set.

	i=1	**i=2**	**i=3**	**i=4**	**i=5**
Accuracy	89	93.75	88.1	90.25	84.28

Table 4.1: Accuracy of DT for 5-fold CV

Prediction	Moderate	Normal	Severe	Mild
Moderate	2	1	0	1
Normal	0	22	0	2
Severe	0	0	4	0
Mild	0	0	0	2

Table 4.2: CM-1 for DT

Prediction	Moderate	Normal	Severe	Mild
Moderate	5	1	0	0
Normal	0	30	0	2
Severe	0	0	8	2
Mild	0	0	0	2

Table 4.3: CM-2 for DT

Prediction	Moderate	Normal	Severe	Mild
Moderate	4	0	0	0
Normal	0	25	0	5
Severe	0	0	6	0
Mild	0	0	0	2

Table 4.4: CM-3 for DT

4.3 Naive-Bayes

NB is a probabilistic classifier and it is based on Bayes Theorem. It can be used when we have small number of training observations but it assumes that the features in the training set are independent which is not the case here as some are derived parameters. So, we used correlation based backwards elimination to get rid of dependencies between features. We trained our NB classifier on this model ($Status \sim Stride\_Length + Cadence + Swing\_Ratio + Stride\_Time$) and used 80% dataset for training and rest 20% for testing. We used 5-fold CV for adding randomness in training and testing sets.

The accuracy is defined as the ratio of number of perfect classification to the total no of observations in testing set i.e ratio of sum of diagonal elements in confusion matrix to the sum of all elements in confusion matrix.

The Sensitivity(Se) is the probablity that subject is in actual class and detected in actual class.

The Specificity(Sp) is the probablity that subject is not in actual class and not detected in actual class.

The Kappa(k) compares accuracy with expected accuracy. The value of kappa in the range of 0.80 to 1 is an excellent agreement, in range of 0.6 to 0.8 is good agreement, in range of 0.4 to 0.6 is an average agreement and below this is poor or fair agreement.

4.3.1 Results

Results are as shown in Table 4.5. The average value of accuracy for NB classifier was 86.38% for iterating five times. The average value for Kappa was 0.72 which is good agreement. Average Se and Sp obtained for Moderate class after iterating five times was 0.76 and 0.96. Average Se and Sp obtained for Normal class after iterating five times was 0.91 and 0.92. Average Se and Sp obtained for Severe class after iterating five times was 0.9 and 0.97. Average Se and Sp obtained for Mild class was 0.575 and 0.942.

		i=1	**i=2**	**i=3**	**i=4**	**i=5**
	Accuracy	88.37	88.37	79.07	90.7	86.05
	Kappa	0.6434	0.7964	0.6526	0.7863	0.7448
Moderate	Sensitivity	0.5	1	0.5	1	0.8
	Specificity	0.97561	0.9474	0.94872	0.97619	0.97368
Normal	Sensitivity	0.9444	0.9615	0.9167	0.9091	0.8621
	Specificity	1	0.8824	0.7895	1	0.9286
Severe	Sensitivity	0.6667	1	1	0.8571	1
	Specificity	0.95	1	0.9444	1	1
Mild	Sensitivity	0.5	0.5	0.375	1	0.5
	Specificity	0.95122	0.97143	0.97143	0.92683	0.90244

Table 4.5: Results on 5-fold CV for NB classifier

Confusion matrix: We are have total 5 confusion matrix but here we are showing only three. The first CM (Table 4.6) misclassified 1 moderate patient as severe and 3 mild patients as normal class. The second CM (Table 4.7) misclassified 4 mild patients 1 as normal and 3 as mild, 2 mild patients 1 as normal and 1 as mild, 2 severe patients as moderate and 1 mild patient as normal. The third CM (Table 4.8) misclassified 2 of each moderate and normal patients as mild and 1 of the mild patients as normal.

Prediction	Moderate	Normal	Severe	Mild
Moderate	1	0	1	0
Normal	0	30	0	0
Severe	0	0	6	0
Mild	0	3	0	2

Table 4.6: CM-4 NB Classifier

Prediction	Moderate	Normal	Severe	Mild
Moderate	2	1	0	1
Normal	0	22	0	4
Severe	2	0	7	0
Mild	0	1	0	3

Table 4.7: CM-5 NB classifier

Prediction	Moderate	Normal	Severe	Mild
Moderate	5	0	0	2
Normal	0	25	0	2
Severe	0	0	4	0
Mild	0	1	0	4

Table 4.8: CM-6 NB classifier

4.4 Random Forest

Random Forest Classifier is classifier that is made of lot of decision trees and the classification is based on the output class of Decision Trees by calculating probability. It removes the amount of correlation in different decision trees made with different bootstrapped samples by selecting different i features out of total n features. It usually is accurate and provides the list of important values and it reduces variance by using bootstrapped samples but it sometimes overfits. We used cross-validation to check for how many features should the model be trained to give maximum accuracy. From the Figure 4.5 we came to know that it gives highest accuracy by using 4 features in the training set. So, we executed random forest with 10-fold Cross-validation(CV), 500 decision trees and with 4 features in R.

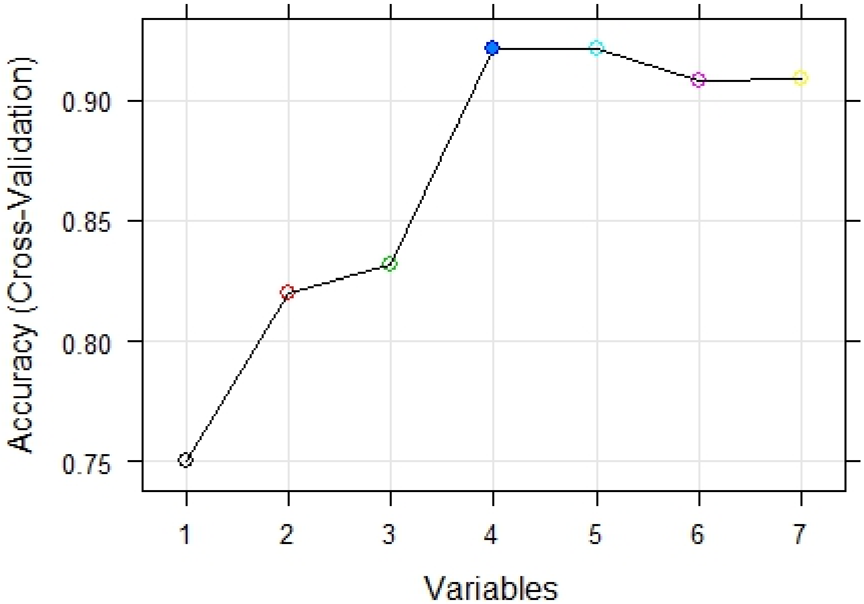

Figure 4.5: Cross-validation for determining number of variables in RF model

4.4.1 Results

The results are summarized in table 4.9. The average value of accuracy for RF classifier was 90.48% for 10-fold CV which was maximum among all the classifiers. The average value for Kappa was 0.832 which is very good agreement. Average Se and Sp obtained for Moderate class for 10-fold CV was 0.821 and 0.975. Average Se and Sp obtained for 10-fold CV was 0.965 and 0.92. Average Se and Sp obtained for Severe class obtained for 10-fold CV was 1 and 0.993. Average Se and Sp obtained for 10-fold CV Mild class was 0.62 and 0.967.

		i=1	**i=2**	**i=3**	**i=4**	**i=5**	**i=6**	**i=7**	**i=8**	**i=9**	**i=10**
	Accuracy	1	0.944	0.9474	0.8889	0.8235	0.913	0.9524	0.86	0.85	0.9167
	Kappa	1	0.8831	0.914	0.7844	0.6531	0.8321	0.9009	0.79	0.7297	0.85
Moderate	Sensitivity	1	1	1	0	0.5	1	1	1	0.66	1
	Specificity	1	1	1	0.9411	0.933	1	1	1	0.88	1
Normal	Sensitivity	1	1	1	1	0.9167	0.933	1	1	0.9231	0.875
	Specificity	1	0.833	0.8889	1	1	0.875	0.857	0.75	1	1
Severe	Sensitivity	1	1	1	1	1	1	1	1	1	1
	Specificity	1	1	1	0.9375	1	1	1	1	1	1
Mild	Sensitivity	1	0.75	0.8	0.667	0	0.8	0.5	0.333	0.6667	NA
	Specificity	1	1	1	1	0.87	0.944	1	1	0.9412	0.9166

Table 4.9: Performance of Random Forest on 10-fold CV

Confusion Matrix:

Confusion Matrix obtained are shown below (Table 4.10, 4.12 and 4.11).

Prediction	Moderate	Normal	Severe	Mild
Moderate	2	0	0	0
Normal	0	15	0	0
Severe	0	0	4	0
Mild	0	0	0	1

Table 4.10: Confuison Matrix for RF best performance in 10-fold CV for i=1

Prediction	Moderate	Normal	Severe	Mild
Moderate	1	0	0	1
Normal	0	11	0	0
Severe	0	0	2	0
Mild	1	1	0	0

Table 4.11: Confuison Matrix for RF worst performance in 10-fold CV for i=5

Prediction	Moderate	Normal	Severe	Mild
Moderate	1	0	0	0
Normal	0	14	0	1
Severe	0	0	2	0
Mild	0	1	0	4

Table 4.12: Confuison Matrix for RF average performance in 10-fold CV for i=6

4.4.2 Important Variable in Construction of RF

The Gini-index is measure of impurity level while building the decision tree same as MSE(Mean Square Error) in Regression. The variables with a large mean decrease Gini are more important predictors for RF. The mean decrease gini is measure of node purity meaning the variable having high mean decrease gini can be used to split node into childrens by this variable. Stride Speed has highest MeanDecreaseGini whereas Stride length also is competing with Stride Speed. Number of Steps is less important than Stride Speed and Stride Length whereas Swing Ratio and Stance Ratio are competing with each other. Cadence and Stride Time are have lowest mean decrease Gini. The importance was derived from RF model. So, Stride Speed, Stride Length, Number of Steps and Stance or Swing ratio may be the most important 4 features used by RF (Figure 4.6).

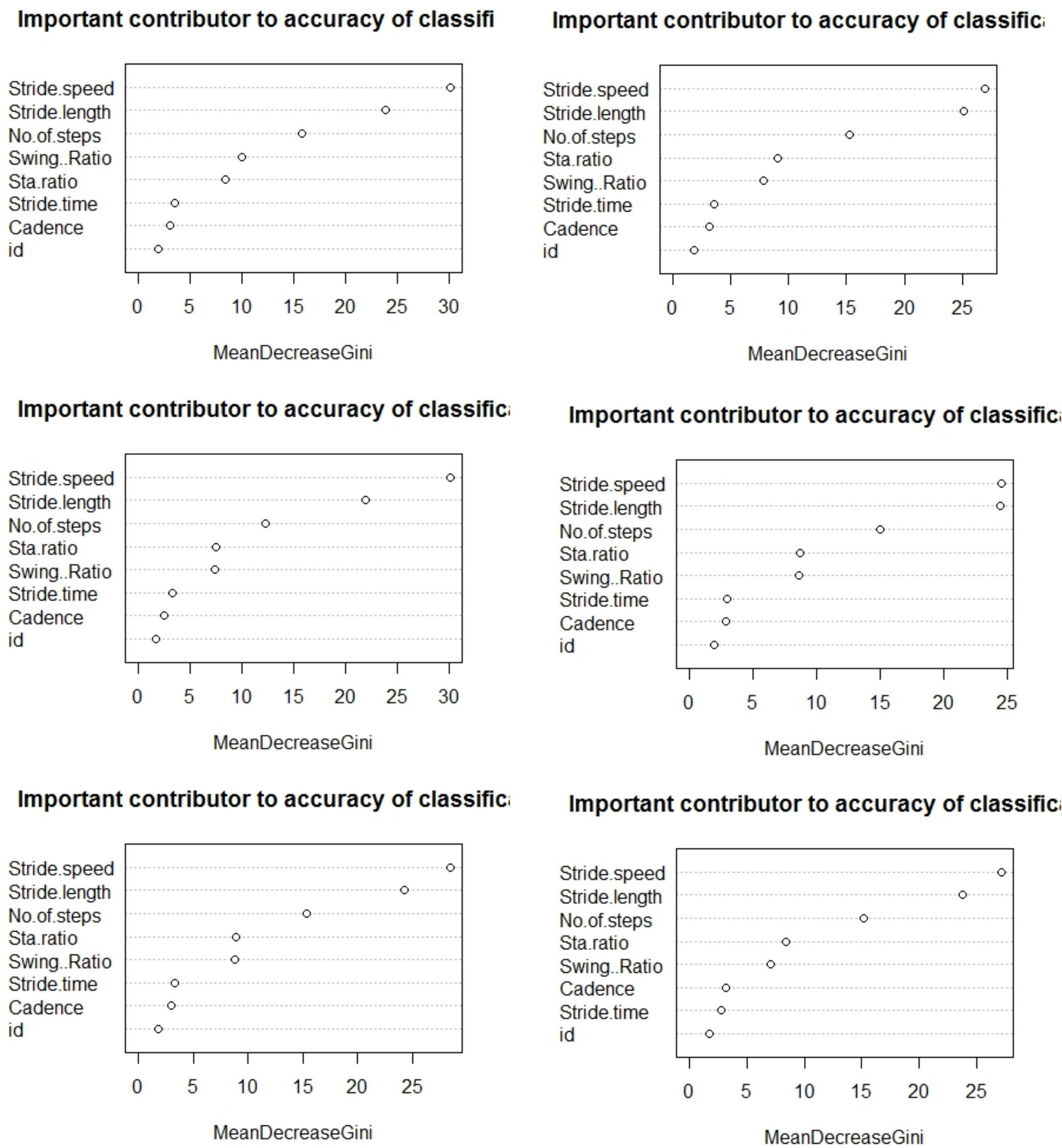

Figure 4.6: MeanDecreaseGini Plots Indicating Importance of Features

RF is performing better over other classifiers, improving sensitivity of mild class and kappa.

CHAPTER 5

Conclusion

In this chapter we finally present the conclusions,drawbacks and further improvement in the methods which can be used to make real time andriod application.

5.1 Conclusions

In this thesis we have presented a solution for abnormal gait detection using smartphone as data collection platform and post extraction of relevant gait parameters followed by classification process. We concluded that *Stride length* and *Swing Ratio* are useful parameters for distinguishing mild and normal patients who usually have similar features. The *Stride Length* forms clear decision boundary between Normal and Severe class. Moderate class and Mild class were also distinguised through these two parameters. However, *RF* model indicates that *Stride Speed* is most impotant parameter that it has used to distinguish the classes, followed by *Stride Length*, *Number of Steps* to complete 6*m* walk test and *Swing* or *Stance Ratio*. Out of Three classifiers used *RF*, *NB*, *DT*; *RF* performs best with 90.48% accuracy, followed by *DT* with 89.07% accuracy. The *NB* classifier gave 86.38% as it treats all the parameters as independent or having Normal Distribution.

5.2 Problems with System

In 6*m* walk test we have to start the recording manually by turning on the *AndroSensor* before the patient starts walking and turn off the *AndroSensor* after patient completes the 6*m* walk. The tie gap between the time of turning on the recording button and time at which the user actually started walking was approximately 0.5 seconds. Similarly, some time was also spent when user comleted walk and recording was stopped. It was also around 0.5 seconds. These offsets before the start of walk and after completion of the walk, was creating noise in the system. This noise initially and finally was removed manually by starting at *Heel_Strike* and ending the walk at *Heel_Strike* for both the legs and analysing between the *Heel_Strikes*.

The Subject was instructed to take a step with the refrence leg initially i.e. the leg in which SmartPhone is placed, to start at a *Heel_Strike*.

The total number of data collected was 185 but for patient class it was only 67 which is only 36.21%. So, if more data for patient (*Mild*, *Moderate* and *Severe*) can be collected then our supervised algorithms can perform more better.

5.3 Future Work

For real time estimation of gait parameters, the problems discussed in previous Section 5.2 has to be removed. This can be done by removing the manual recording of subject, which can be done by extending the walk path to around 12 meter and mark the start and end of 6 meter walk path. A second Smartphone can be used to control the the Smartphone placed at the ankle of subject. Using the second smartphone we could start the recording when Subject reaches Start of $6m$ remotely and alsol terminate the recording when subject reaches end of $6m$ remotely. Using this technique one can remove the noise in the System.

If around 50% of the total dataset can be obtained for the patient class then accuracy for Supervised Algorithm can be improved. Then the robust Decision Tree model can easily be implemented after extraction of gait parameters to build an andriod application post analysis.

References

[1] H. Chan, H. Zheng, H. Wang, R. Sterritt, and D. Newell. Smart mobile phone based gait assessment of patients with low back pain. In *Natural Computation (ICNC), 2013 Ninth International Conference on*, pages 1062–1066. IEEE, 2013.

[2] S. J. Cuccurullo. *Physical medicine and rehabilitation board review*. Demos Medical Publishing, 2014.

[3] A. Ferrari, P. Ginis, M. Hardegger, F. Casamassima, L. Rocchi, and L. Chiari. A mobile kalman-filter based solution for the real-time estimation of spatio-temporal<? pub _newline?> gait parameters. *IEEE transactions on neural systems and rehabilitation engineering*, 24(7):764–773, 2016.

[4] D. Han, V. Renaudin, and M. Ortiz. Smartphone based gait analysis using stft and wavelet transform for indoor navigation. In *Indoor Positioning and Indoor Navigation (IPIN), 2014 International Conference on*, pages 157–166. IEEE, 2014.

[5] P. K. Levangie and C. C. Norkin. *Joint structure and function: a comprehensive analysis*. FA Davis, 2011.

[6] F. Lin, A. Wang, Y. Zhuang, M. R. Tomita, and W. Xu. Smart insole: A wearable sensor device for unobtrusive gait monitoring in daily life. *IEEE Transactions on Industrial Informatics*, 12(6):2281–2291, 2016.

[7] A. J. A. Majumder, S. I. Ahamed, R. J. Povinelli, C. P. Tamma, and R. O. Smith. A novel wireless system to monitor gait using smartshoe-worn sensors. In *Computer Software and Applications Conference (COMPSAC), 2015 IEEE 39th Annual*, volume 2, pages 733–741. IEEE, 2015.

[8] S. Mazilu, M. Hardegger, Z. Zhu, D. Roggen, G. Troster, M. Plotnik, and J. M. Hausdorff. Online detection of freezing of gait with smartphones and machine learning techniques. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2012 6th International Conference on*, pages 123–130. IEEE, 2012.

[9] L. Pepa, F. Verdini, and L. Spalazzi. Experimental evaluation of a smartphone based step length estimation. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 7647–7650. IEEE, 2015.

[10] A. A. Perez and M. A. Labrador. A smartphone-based system for clinical gait assessment. In *Smart Computing (SMARTCOMP), 2016 IEEE International Conference on*, pages 1–8. IEEE, 2016.