GAIT ANALYSIS FOR SEVERITY RATING OF PARKINSON'S DISEASE USING MACHINE LEARNING AND DEEP LEARNING ALGORITHMS

A THESIS

Submitted by

BALAJI E

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

FACULTY OF INFORMATION AND COMMUNICATION ENGINEERING

ANNA UNIVERSITY

CHENNAI 600 025

JULY 2021

ANNA UNIVERSITY CHENNAI

CHENNAI - 600 025

CERTIFICATE

The research work embodied in the present Thesis entitled **"Gait Analysis for Severity Rating of Parkinson's Disease using Machine Learning and Deep Learning Algorithms"** has been carried out in the Department of Biomedical Engineering, PSG College of Technology, Coimbatore. The work reported herein is original and does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion or to any other scholar.

I understand the University's policy on plagiarism and declare that the thesis and publications are my own work, except where specifically acknowledged and has not been copied from other sources or been previously submitted for award or assessment.

Balaji E
RESEARCH SCHOLAR

Brindha D
SUPERVISOR
Assistant Professor (Sl.Gr)
Department of BME
PSG College of Technology
Coimbatore

ABSTRACT

Parkinson's disease (PD) is a chronic neurodegenerative brain disorder, which affects the ability of the person to perform the regular activities. While diagnosing PD, neurologists often use several clinical manifestations like motor and non-motor symptoms and rate the severity based on the Unified Parkinson Disease Rating Scale (UPDRS). This kind of rating largely depends on the expertise and experience of the clinicians. and it is not only subjective but also inefficient. Gait pattern, which plays a vital role in assessing the human mobility, is a significant biomarker to classify whether the subject is healthy or affected with PD. Hence, in this work, we aim to investigate the gait pattern of healthy and PD subjects using machine learning (ML) and deep learning (DL) algorithms to design an automatic and non-invasive PD diagnosis and severity rating system that can assist the neurologists in their daily PD diagnosis. In this regard, the vertical ground reaction force (VGRF) gait dataset is utilized and the following three ML and DL classification algorithms are explored: 1. Supervised machine learning based classifier 2. Convolutional neural network (CNN) classifier and 3. Long-short term memory(LSTM) network classifier. Firstly, To obtain the optimal feature set for the classification model, a correlation based feature selection technique is employed. Secondly, the statistical analysis of the VGRF sensor data is performed to differentiate between healthy control and PD patients. Four supervised machine learning algorithms namely decision tree(DT), support vector machine (SVM), ensemble classifier(EC) and Bayes classifier(BC) are used to classify the stages of PD based on Hoehn and Yahr (H&Y) scale. Moreover, to avoid data overfitting problem and enhance the classification accuracy, the 10 fold cross validation technique is utilized. Even though the supervised machine learning classifiers yield high accuracy in detecting the PD, one of the major limitations is the need for hand-crafted features.

Hence, to overcome the hand-crafted feature approach, secondly, we explore the CNN classifier for multi-class classification for various frame sizes. To avoid the data over-fitting, L2 regularization technique, which penalizes the weight parameters of the nodes, is used in combination with the dropout layer. For optimizing the loss function, a stochastic gradient descent (SGD) algorithm is employed because it reduces the computational burden for large dataset. Experimental results substantiate that the proposed DCNN architecture outperforms state-of-the-art artificial neural network (ANN) classifiers and achieves the highest classification accuracy of 98.45%. Thirdly, to evaluate the potential of LSTM classifier, which is highly suitable for learning the long-term temporal dependencies in the gait cycle, we study the PD rating based on modified H&Y scale and UPDRS scale for three different walking scenarios. For solving the cost function, Adam, a stochastic gradient-based optimizer, is employed and the severity of PD is categorized based on UPDRS and H&Y scale. The experimental results reveal that Adam optimized LSTM network can effectively learn the gait kinematic features and offer an average accuracy of 98.6% for binary classification and 96.6% for multi-class classification

ACKNOWLEDGEMENT

I wish to record my deep sense of gratitude and profound thanks to my research supervisor **Dr. Brindha D**, Assistant Professor (Sl.Gr), Department of Biomedical Engineering, PSG College of Technology, Coimbatore, for her keen interest, inspiring guidance, constant encouragement with my work during all stages, to bring this thesis into fruition.

I am extremely indebted to **Dr. Vidhyapriya R**, Professor & Head, Department of Biomedical Engineering, PSG College of Technology, Coimbatore, for her valuable suggestions and support during the course of my research work.

With a great sense of pleasure and privilege, I extend my gratitude to the principal **Dr. Prakasan K** and the management of PSG College of Technology for providing the necessary facilities and support.

I also thank the faculty and non-teaching staff members of the Department of Biomedical Engineering, PSG College of Technology, Coimbatore, for their valuable support throughout the course of my research work.

Finally, I express my sincere gratitude to my parents for cheering me up at every stage during my doctoral study. Without their love, endless support and encouragement, this thesis might not have been possible.

BALAJI E

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

PD	Parkinson's disease
UPDRS	Unified Parkinson Disease Rating Scale
H & Y	Hoehn & Yahr
ML	Machine learning
DL	Deep learning
CNN	Convolutional neural network
LSTM	Long short term network
DT	Decision tree
SVM	Support vector machine
EC	Ensemble classifier
BC	Bayes classifier
NB	Naïve Bayes
ROC	Region of convergence
AUC	Area under curve
VGRF	Vertical ground reaction force
LDA	Linear discriminant analysis
PPV	Positive predictive value
NPV	Negative predictive value
MCC	Mathew's correlation coefficient

WOIB	Without impairment of balance
WIB	With impairment of balance
WIPR	With impaired postural reflexes
SGD	Stochastic gradient descent
ANN	Artificial neural network
ReLU	Rectified linear unit
FC	Fully connected
RNN	Recurrent neural network
DNN	Deep neural network
CM	Confusion matrix

CHAPTER 1

INTRODUCTION TO PARKINSON'S DISEASE

PD is a movement disorder which manifests a multitude of motor and non-motor symptoms and consequently impacts the quality of life. Even though PD is still incurable, early prognosis of this disease can help to plan for appropriate medication/surgery so that the progression of the disease can be minimized. Some of the most common non-motor symptoms are cognitive impairment, sleep disorders and depression. The motor symptoms of PD include tremor, rigidity, slow movement, also called bradykinesia, and postural instability (Tropea & Chen-Plotkin 2018; Mirelman et al. 2019). The average age of onset for PD is about 60 years. According to the Parkinson's foundation(PF), USA, it is estimated that about 8-10 million people worldwide are affected by the PD. Arvid Carlsson, a Swedish neuropharmacologist who was awarded the Nobel prize in Physiology in 2000, found the major reason for PD being the lack of a neurotransmitter in the brain called "dopamine". The dopamine deficiency can lead to several mental and neurological issues, including executive dysfunction and dementia. Figure 1.1 illustrates the manifestations of both motor and non-motor symptoms of PD. Even though PD is a kind of extrapyramidal disorder without any apparent etiology, the recent investigations reveal that the major causes for the PD could be attributed to gene mutation, excessive exposure to toxins and aging. Numerous non-motor symptoms of PD start ahead of the motor symptoms (Alam et al. 2017; Joshi et al. 2017). Three of the most common motor symptoms of PD are: tremor, limb rigidity, and slowness of movement (bradykinesia). Similarly, the common non-motor symptoms of PD are sleep disorder, depression and fatigue. Person affected with PD may also have problem with balance, posture and coordination. Even though the symptoms of PD

are common, progression of PD may vary largely from person to person and change over time (Zeng & Wang 2015; Das 2010).

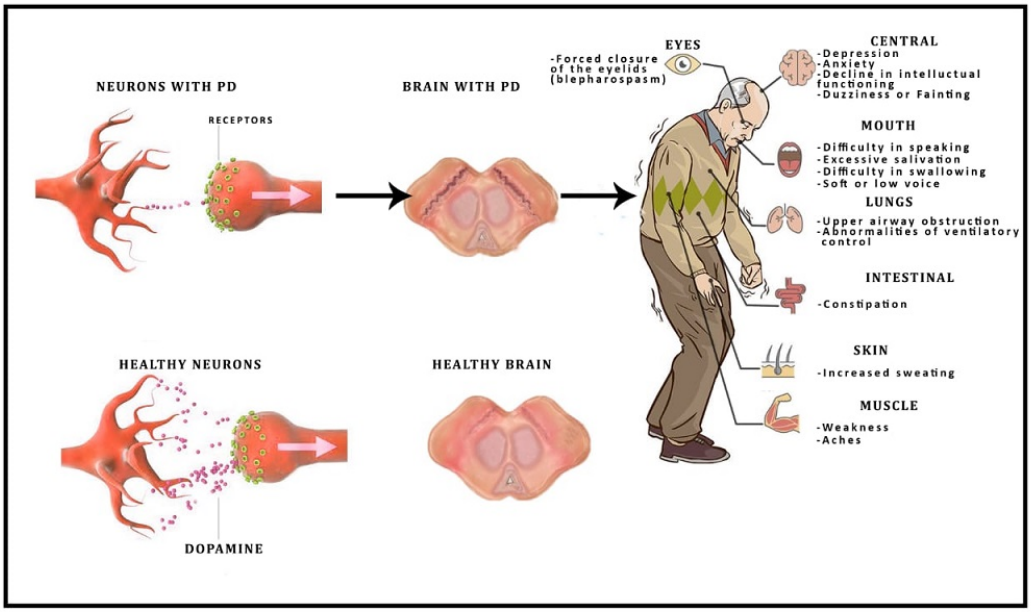


Figure 1.1: Causes for PD and its symptoms

1.1 PD-rating scales

The two commonly used clinical rating scales to monitor PD progression are UPDRS and H&Y scale (Goetz et al. 2004). While the former is largely a comprehensive tool to assess the non-motor symptoms including mental behaviour, mood and social interaction, the latter is entirely based on the mobility, which takes only the motor symptoms into consideration to classify the stages of PD (Joshi et al. 2017).

1.1.1 UPDRS

MDS-UPDRS is a conventional rating scale to assess the stages of PD based on the response to the questionnaire and score prepared by clinicians.

UPDRS, which typically contains 42 items, are categorized into four parts: Part I-non-motor activity of daily living, Part II-motor activity of daily living, Part III-motor examination and Part IV -motor complications. According to UPDRS, the symptoms of PD are rated on a 5 point scale, ranging from 0 to 4. The higher the score, the more the severity of the PD. The highest UPDRS score is 199, representing the worst possible stage of PD.

1.1.2 H&Y scale

H&Y scale, proposed by Melvin D. Yahr and Margaret M. Hoehn in 1967, contains five stages, as illustrated in Table 1.1, and it is still considered to be a universally accepted rating scale for its effectiveness to provide an overall validation of severity according to functional disability. Moreover, H & Y scale is easy to apply and quick to finish the rating.

Table 1.1: PD severity rating based on H & Y scale

Scale	Functionality	Stage
1	Unilateral-Minimal	No functional disability
2	Bilateral-Midline	Without Impairment of Balance
3	Bilateral-Moderate	With Impaired Postural Reflexes
4	Need assistance to stand	Severely Disabled
5	Confined to bed or wheelchair bound	Completely Disabled

1.2 Existing methods for PD diagnosis

Some of the physiological signals used for PD diagnosis are speech, handwriting, tremor and gait.

1. Speech based methods

Speech signals manifest the non-motor symptoms of PD. Particularly, the phonotation of sustained vowels, words can help to discriminate between healthy and PD patients. In this direction, for predictive telediagnosis of Parkinsonism, University of Oxford collaborated with the National Centre for Voice, Denver to collect a variety of voice samples from people afflicted with PD and made the dataset publicly available in UCI repository. Several researchers have used the UCI speech signals and investigated the performance of ML algorithms to classify the stages of PD based on UPDRS (Sakar et al. 2013). For optimal feature selection, the time-frequency based features such as jitter, shimmer etc from voice samples have been utilized.

2. Handwriting based methods

The handwriting, which entails cognitive, perceptual-motor and kinesthetic components, can act as a potential biomarker when there is any significant change in it. As it is a non-invasive method, researchers have investigated PD prognosis using handwriting dynamics. Commonly, using the graphics tablet and LCD monitor, the handwriting database is acquired for three different tests namely 1. Static Spiral Test, 2. Dynamic Spiral Test and 3. Stability Test. In this direction, some of the notable results using ML are reported in (Pereira et al. 2018; Gupta et al. 2020).

3. MRI based methods

In another direction, the advent of brain imaging modalities provides great impetus to the computer based PD diagnosis. Particularly, the structural magnetic resonance imaging has been widely used for its ability to render high-resolution structure of the brain tissues and non-invasiveness char-

acteristics. Similarly, an another tool called DatScan, a type of single-photon emission computed tomography (SPECT) scanning technique, is also used for diagnosis of PD. Recently, ML algorithms are used to detect the presence of dopamine in the substantia nigra using the MRI images (Amoroso et al. 2018; Battineni et al. 2019). However, the cost associated with the MRI based PD diagnosis is considerably high and as an alternative approach, gait based PD analysis has been largely used.

4. Gait based methods

Gait based classification which can be used for diagnosing motor control deficits is not only cost effective but also easy to acquire the gait pattern (Shrivastava et al. 2017). Interestingly, gait cycle has some prominent characteristics like periodicity, deterministic behavior and spatio-temporal features. Moreover, unlike speech and handwriting based methods that give only the non-motor symptoms, the gait analysis renders clinicians to assess motor symptoms and visualize how severely the subjects are affected (Turner & Hayes 2019). Hence, based on the assessment, the clinicians can suggest the therapeutic treatment to delay the progression of the disease.

1.3 Motivation

Currently, in the clinical setting, the diagnosis and severity rating of PD is based on the visual observation and UPDRS score prepared by the clinicians. Since the visual observation is subjective, it may result in biased assessment. Hence, recently, there has been a paradigm shift towards utilizing machine learning (ML) algorithms to uncover the hidden patterns in the physiological signals, which can help the clinicians in their daily diagnosis of PD. One of the significant reasons for using ML algorithms in bioinformatics is their ability to handle large volume of data and discover biomarkers, resulting in not

only accurate prediction but also reduced time for diagnosis.

1.4 Objectives

1. To explore the statistical features of gait pattern through VGRF sensor data and identify the significant biomarkers to differentiate between healthy and PD patients.
2. To investigate the kinematic features of gait pattern and extract the primary kinematic parameters using correlation based feature selection approach so as to address the limitation of statistical analysis, which is influenced by the weight factor of the subject under study.
3. To implement the supervised machine learning based classifiers, including SVM, DT, EC, and BC, to diagnose the stages of PD based on UPDRS and H & Y scale.
4. To overcome the hand-crafted feature requirement of machine learning classifiers using deep learning approach and evaluate the performance of CNN models to quantify the stages of PD.
5. To exploit the long-term temporal dependencies in the gait cycle through LSTM classifier and implement both binary class and multiclass classifications.

1.5 Contributions

The major contributions of this thesis are as follows.

- An automatic and non-invasive gait classification framework using supervised machine learning algorithms is presented to assist

the clinicians to diagnose the severity of PD. Utilizing the vertical ground reaction force (VGRF) dataset from the publicly available Physionet database, we have performed the statistical and kinematic analyses of gait pattern and extracted the salient biomarkers from the gait cycle using the Pearson's rank correlation test to reduce the redundant/insignificant features.

- Four supervised machine learning algorithms namely decision tree (DT), support vector machine (SVM), ensemble classifier (EC) and Bayes classifier (BC) are utilized to classify the stages of PD based on the optimal feature vector obtained from the statistical and kinematic analyses of gait pattern. Moreover, to avoid data overfitting issue and enhance the classification accuracy, the 10 fold cross validation technique is utilized.

- To minimize the dispersion in gait samples and decorrelate them from physical parameters like height and weight, a multiple regression normalization approach is used. Experimental validation substantiates that compared to several state-of-the-art methods, the proposed approach, which uses optimal spatiotemporal features, can offer better prediction of PD severity rating based on the H&Y scale.

- In the second approach, we present a DCNN based PD classification method mainly to avoid the need for hand-crafted features in the machine learning algorithms. We have analyzed the 1D and 2D DCNN architectures for different frame sizes to assess the performance of one dimensional input signal method and two dimensional image input method in CNN framework. Then, the performance of the proposed DCNN architecture to classify the severity of PD is assessed based on modified H & Y scale using gait pattern obtained for three different walking scenarios such as 1. walking on a level ground 2. walking on a treadmill and 3. walking with rhythmic auditory stim-

ulation (RAS).

- Dividing the pre-processed input pattern into frames, we test the performance of the classifier using the confusion matrix. Utilizing the stochastic gradient descent algorithm for CNN loss function optimization and L2 regularization technique to avoid data overfitting, we achieve an average stage classification accuracy of 98.45 % for PD stage prediction.
- In the third approach, a PD severity rating technique using an Adam optimized LSTM classifier is put forward. Firstly, the binary classification of PD is addressed by exploiting the long-term temporal gait sequence. Subsequently, the severity rating of PD is identified based on UPDRS and H&Y scale.
- The performance of LSTM classifier is tested on the VGRF gait dataset, which includes three different walking scenarios. Moreover, to avoid data overfitting, the proposed approach utilizes dropout and L2 regularization techniques.
- Adam optimizer, which has less memory requirement and few hyper parameters tuning, is employed to solve the cost function. The LSTM classifier has achieved an average accuracy of 96.6% for multi-class classification and provides 3.9% improvement in overall accuracy compared to the existing techniques

1.6 Organization of the thesis

Chapter 1 presents the PD overview, causes and the methods available for diagnosis. The clinically accepted PD rating scales such as UPDRS and H&Y scales are briefed along with the motivation and contributions of this thesis.

Chapter 2 briefs the statistical analysis of gait pattern and explains the dataset considered for the study. The supervised machine learning algorithms and the performance metrics considered to assess the classifiers are also presented. Then, the performance of classifiers evaluated using the confusion matrix and the region of curve (ROC) is discussed.

Chapter 3 provides a brief introduction to the spatial and temporal features of gait and presents the proposed kinematic analysis of gait pattern for severity rating of PD. Subsequently, the performance of the ML classifiers are assessed using the six key performance metrics and the comparative study of the proposed approach with the other related approaches are presented.

Chapter 4 describes the deep learning based PD severity rating approach using gait cycle analysis. The architecture and different layers of DCNN are introduced. Then, the performance of 1D and 2D DCNN for binary class and multi-class classification problems are discussed. The experimental results obtained using the Keras and Tensorflow platform are presented.

Chapter 5 explains the LSTM network classifier for binary and multi-class classification of PD using the gait analysis. It also presents the Adam optimizer for solving the objective function and the L2 and dropout layer to avoid the data overfitting issue. Finally, the performance of the LSTM classifier compared with those of the related approach is presented for validation.

Chapter 6 gives the concluding remarks of the thesis along with some of the open challenges, which require further investigation.

CHAPTER 2

STATISTICAL ANALYSIS OF GAIT FOR PD SEVERITY RATING

The tremor and gait abnormality are the initial manifestations of PD. Gait pattern, which involves the sequence of periodic and rhythmical pattern of foot movements, has some interesting and distinct properties like periodicity, irregular cycle and deterministic behavior (Chen et al. 2013). Particularly, the kinematic features of gait pattern have significant biomarkers which can be critical for not only identifying the presence of PD but also quantifying the progression of the disease (Creaby & Cole 2018). Moreover, the spatiotemporal variables are less prone to physiological parameter like age, height and weight; hence, they can serve as the basis to extract the significant features for PD prognosis (Zhang & Ma 2019). Even though the PD diagnosis has been studied extensively based on gait cycle analysis, the kinematic features of gait abnormality for prognosis of PD have not been investigated to its full potential. Hence, the major objective of this paper is to explore the significant spatiotemporal features of the gait pattern so that the stages of PD, which need finer investigation of walking pattern, can be precisely identified.

In this study, an automatic and non-invasive gait classification system for PD diagnosis and severity rating is presented. From the gait pattern acquired using 16 foot worn sensors, the prominent temporal and spatial features are extracted. To obtain the optimal feature set for the classification model, a correlation based feature selection technique is employed. Firstly, the statistical analysis of the VGRF sensor data is performed to differentiate between healthy control and PD patients. In statistical analysis, the weight of the sub-

ject influences the foot plantar pressure and classifier output. Hence, we also perform kinematic analysis to avoid biased classification. Unlike several other ML based approaches which perform the binary classification, in this work, we aim to solve the multi-class classification problem of PD using discriminative feature set obtained from the spatiotemporal features. Four supervised machine learning algorithms namely decision tree(DT), support vector machine (SVM), ensemble classifier(EC) and Bayes classifier are used to classify the stages of PD based on H & Y scale. Moreover, to avoid data overfitting problem and enhance the classification accuracy, the 10 fold cross validation technique is utilized. The efficacy of the classifier model is validated using the region of convergence (ROC) curve and confusion matrix.

2.1 Related Work

Several previous studies have assessed the gait features of subjects affected with Parkinsonism. For instance, based on gait and tremor features extracted using statistical analysis, Perumal & Sankar 2016a presented a linear discriminant analysis (LDA) based pattern classification algorithm for early detection and monitoring of PD. Utilizing the kinetic features to identify a PD tremor due to Parkinsonism, they analysed the frequency domain characteristics of tremor and obtained an average accuracy of 86.9%. However, their approach is reduced to a binary classification problem in which only the presence of PD is detected and stages of PD have not been identified.

Aşuroğlu et al. 2018 proposed a locally weighted random forest regression model to eliminate the effects of interpatient variability in gait features and utilized the ground reaction force sensor data to model the relationship between gait patterns and PD symptoms. Based on the 16 time-domain and 7 frequency-domain features, they provided a quantitative assessment of PD motor symptoms. Nevertheless, only the statistical analysis of GRF, which is influ-

enced by the weight of the subject, is used for identifying the PD symptoms and the kinematic analysis and the severity level of PD have not been reported.

Using wavelet based feature extraction technique for gait characteristics and extracting 40 features based on statistical analysis and frequency distributions, Lee & Lim 2012 proposed a PD classification approach which made use of neural network with weighted fuzzy membership functions and obtained the maximum classification accuracy of 81.63%. To differentiate between idiopathic PD patients and healthy subjects, they used the bounded sum of weighted fuzzy membership functions for selecting minimum features based on a non-overlapping area distribution method. Nonetheless, the large number of gait features obtained from the wavelet approximation and detailed coefficients will significantly increase the computational complexity.

For computer assisted PD severity detection, Zhao et al. 2018b put forward a two channel model, which fuses long short-term memory and convolutional neural network to learn the spatiotemporal gait pattern from VGRF data. Even though significant improvement in the classifier accuracy is achieved using the deep learning architecture, it has been reported that higher prediction accuracy can be achieved by fusing the multiple data sources like MRI image and biochemical data. To understand the periodicity and randomness in gait signals, Prabhu et al. 2018 performed the recurrence quantification analysis (RQA) along with statistical analysis and reported that RQA works well even for short length gait time series data. Moreover, for optimal selection of feature vectors, they used Hill-climbing feature selection method and compared the classification performance of probabilistic neural network with that of the SVM. Figueiredo et al. 2018 showed that using kernel principal component analysis and genetic algorithm, which have the capability to process the nonlinear data and converge to global optimum, could significantly reduce the dimensionality features of gait parameters for efficient classification. They presented a com-

parative analysis of machine learning approaches for gait pattern recognition, and used a cross validation method to assess the performance of classification algorithms for human walking recognition in clinical applications.

Oung et al. 2018 proposed a multi-class PD classification approach with the combination of speech signals and gait pattern, which are analysed based on empirical wavelet transform (EWT) and empirical wavelet packet transform (EMPT), respectively. Applying the Hilbert transform to extract the wavelet energy and entrophy based features, they achieved a classification accuracy of 95% using the extreme learning machine classifier. However, the feature dimensionality selection/reduction to obtain the optimal feature set, which can reduce the computational burden, has not been implemented. To obtain the significant differences in the spatial-temporal features between PD patients and healthy controls, Wahid et al. 2015 used a multiple regression (MR) normalization approach and evaluated the efficacy of machine learning algorithms in classifying PD gait after normalization. Notwithstanding, the MR model was developed and tested for a relatively small dataset with body mass index of majority of the subjects being greater than 25 that may produce increased dispersion of data.

Several of the aforementioned studies have used time-domain and frequency-domain features to differentiate the healthy control and PD subjects. Nevertheless, the major limitation is that such features can not be directly related to clinical indicators which could assist the neurologists to rate the severity level (Khoury et al. 2019a; Ricciardi et al. 2019). Therefore, the prime focus of this paper is to develop a functional tool for diagnosing and severity rating of PD based on the VGRF data and seek to identify the discriminative kinematic features which can be linked to the clinical indicators. Furthermore, using the correlation based optimal feature set extracted from spatiotemporal domain, the proposed PD diagnosis approach aims to improve the classifier accuracy.

2.2 Dataset Description

We used the publicly available gait dataset in Physionet for the proposed study (web 2019). The gait dataset, which was collected at the Laboratory for Gait & Neurodynamics, Movement disorders Unit of the Tel-Aviv Sourasaky Medical centre, Israel, was contributed by three group of researchers namely Yogev et al. 2005, Hausdorff et al. 2007, and Toledo et al. 2005b. The gait dataset contains three gait pattern acquired through 1. walking on a level ground 2. walking with rhythmic auditory simulation (RAS) and 3. walking on a treadmill, is utilized. The prime novelty of this work lies in using the correlation based spatiotemporal features from gait data to enhance the early diagnosis of PD.The dataset contains the gait pattern from 93 patients affected with PD and 73 healthy subjects. To study the stride-to-stride dynamics of the PD patients, the walking pattern was obtained for three different conditions. The dataset collected by (Yogev et al. 2005) consists of gait pattern for normal walking on a level ground, and the dataset contributed by Hausdorff et al. 2007 contains the gait cycle for walking at a comfortable pace with RAS, and the contribution from Toledo et al. 2005b comprises a gait time series data for walking on a treadmill.

Since the gait stability is impaired in PD subjects, investigating the gait dynamics through the force sensor data can reveal the aspect of locomotor dyscontrol and assist to quantify the stride-to-stride variability. Hence, in this work, for quantifying the gait disorder of the subjects with and without the external rhythms, the three datasets are collectively considered for identifying the severity of PD (Toledo et al. 2005a). Henceforth, for easy representation, the three datasets are referred to as Ga (Yogev et al. 2005), Ju (Hausdorff et al. 2007), and Si (Toledo et al. 2005b) . Table 2.1 gives the demographics of the subjects who volunteered to take part in the experimentation for the collection of gait dataset. Table 2.2 presents the total number of healthy and PD subjects with their level of severity in each dataset determined according to H& Y scale.

Table 2.1: Demographics of healthy subjects and PD patients in three datasets

Dataset	Group	Subjects	Male	Female	Age (Yrs)		Height	Weight
					Mean±SD	Range	(metre)	(Kg)
Ga (Yogev et al. 2005)	Healthy	18	10	8	57.9±6.7	37-70	1.68±.08	74.2±12.7
	PD Patient	29	20	9	61.6±8.8	36-77	1.67±.07	73.1±11.2
Ju (Hausdorff et al. 2007)	Healthy	26	12	14	39.31±18.51	20-74	1.83±.08	66.8±11.07
	PD Patient	29	16	13	66.80±10.85	44-80	1.87±.15	75.1±16.89
Si (Toledo et al. 2005b)	Healthy	29	18	11	64.5±6.8	53-77	1.69±.08	71.5±11.0
	PD Patient	35	22	13	67.2±9.1	61-84	1.66±.07	70.3±8.4

Table 2.2: Number of subjects in the three dataset based on severity rating

Dataset	Healthy	Severity 2	Severity 2.5	Severity 3
Ga (Yogev et al. 2005)	18	15	8	6
Ju (Hausdorff et al. 2007)	26	12	13	4
Si (Toledo et al. 2005b)	29	29	6	0

For measuring the forces underneath a foot as a function of time, the gait capturing system contained a pair of shoes and a recording unit attached to the waist of the subjects. Each shoe had eight VGRF sensors, as illustrated in Figure 2.1. Table 2.3 gives the relative positioning of the sensors in $x-y$ coordinate system. The gait data was collected at a sampling rate of 100 Hz from 16 VGRF sensors (8 per foot) for 2 mins. To neglect the startup effects, the first 20s data from each of the subject's time series data was removed.

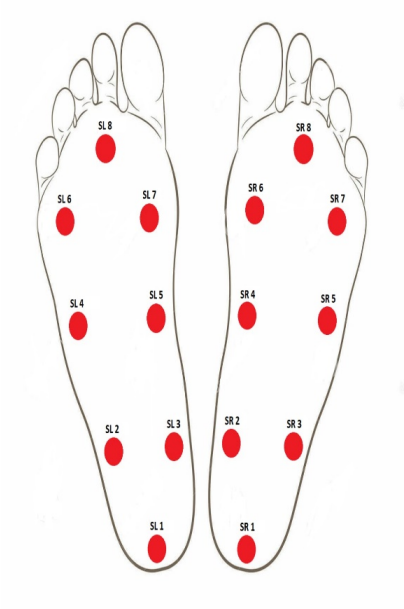
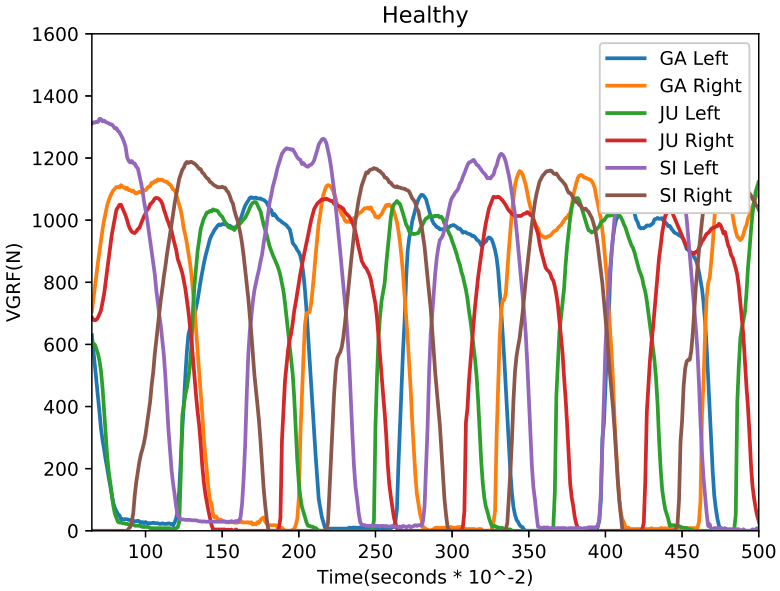


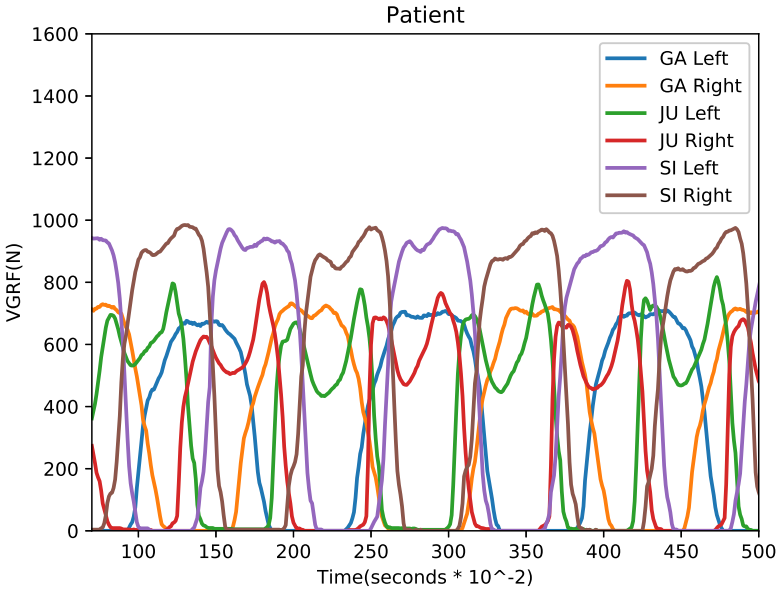
Figure 2.1: VGRF sensor positioning underneath each feet

Table 2.3: VGRF sensor relative positioning in the left and right feet (SL-sensor left, SR-sensor right)

Sensor	X axis (mm)	Y axis (mm)
SL1	-500	-800
SL2	-700	-400
SL3	-300	-400
SL4	-700	0
SL5	-300	0
SL6	-700	400
SL7	-300	400
SL8	-500	800
SR1	500	-800
SR2	700	-400
SR3	300	-400
SR4	700	0
SR5	300	0
SR6	700	400
SR7	300	400
SR8	500	800



(a)



(b)

Figure 2.2: VGRF signals of (a) HOA (b) PPD

Figure 2.2 , which shows the exemplar walking pattern for both healthy older adult (HOA) and person with Parkinson's disease (PPD) from three datasets, highlights that the plantar pressure exerted by the PD patient while walking is

almost half of the pressure exerted by the healthy controls. Moreover, Figure 2.2 b highlights that in the dataset provided by Si (Toledo et al. 2005b) when the PPD walked with external cueing through RAS the mobility increased with substantial increase in the walking speed of the subjects. Specifically, we can notice the improvement in swing and stride time variabilities in the PPD compared to HOA. It is important to note that as gait pattern is subjective to the weight and height of the patients, the classification based on the statistical parameters is largely biased and may result in poor stage classification. Hence, we aim to extract the significant spatiotemporal biomarkers from gait pattern for efficient stage classification.

2.3 Methodology

Figure 2.3 shows the proposed gait classification framework for severity prediction of PD. In data analysis, to reduce the start-up and end-up effects of gait, first and last two gait cycles of each trial were discarded. The statistical and kinematic analyses of the gait pattern are individually performed to extract the prominent biomarkers. Subsequently, the optimal feature set obtained using the rank correlation method is given to the classifier model to predict the stages of PD (Mu et al. 2018). To assess the performance of the classifier model 6 performance metrics are reported. In the following section, we present the statistical and kinematic feature extraction techniques and discuss the ML algorithms used in the classifier model.

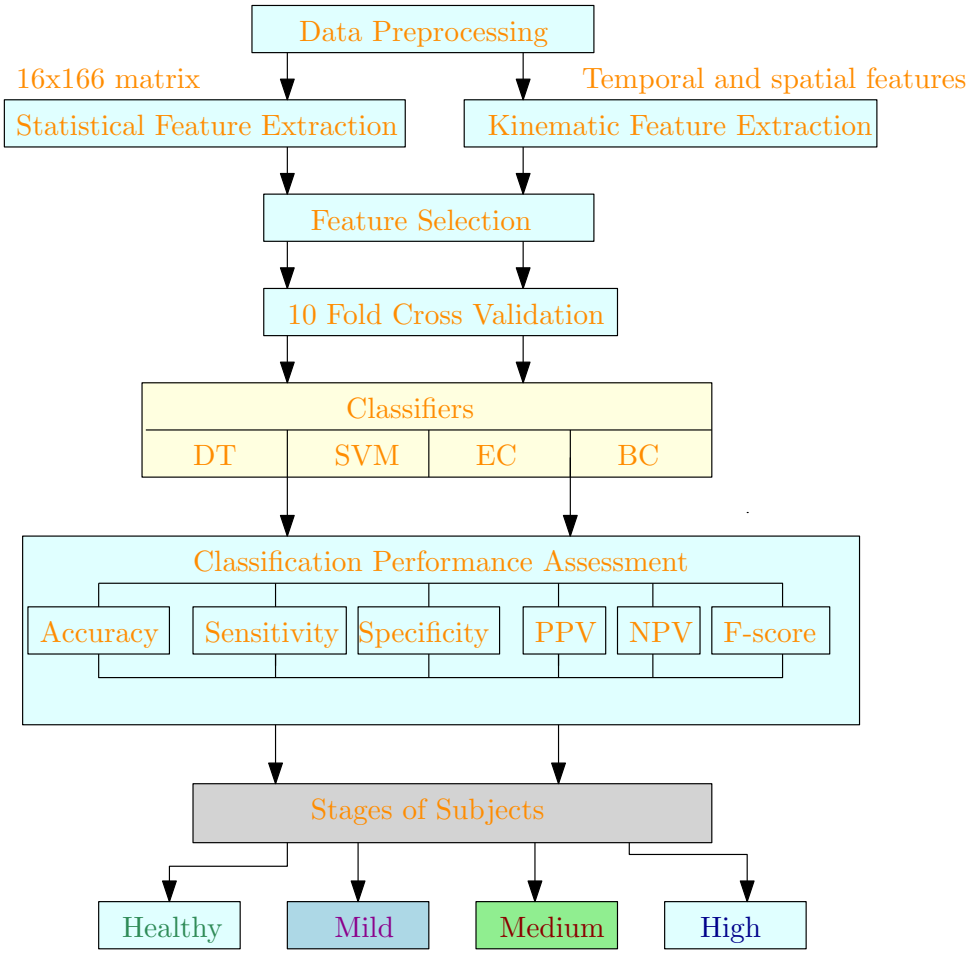


Figure 2.3: Proposed gait classification framework for stage classification of PD

2.3.1 Feature Extraction using statistical analysis

We compute the fluctuation magnitudes of 16 sensors to statistically analyse the gait pattern of both healthy subjects and PD patients. Table 2.4, which gives the mean, median and standard deviation of both the subjects, reveals that the fluctuation of the mean value of the respective sensors in 2 feet of the PD patient is relatively higher than that of the healthy subjects, particularly in SL1, SL4, SL5 and SL6 for left feet sensors and SR1, and SR7 for right feet sensors. Due to higher fluctuations of mean value, we utilize the mean of the VGRF sensor data as a significant biomarker for PD classification. More-

over, for assessing the efficacy of the classifiers, we use 10 fold cross validation approach, which divides the dataset into 10 subsets.

Table 2.4: Statistical parameters of VGRF signals from 16 sensors

Sensors	Healthy subjects			PD patients		
	Mean	Median	SD	Mean	Median	SD
SL1	77.71	85.80	29.98	56.54	50.29	32.45
SL2	63.66	63.35	20.43	55.15	51.56	20.42
SL3	51.52	48.43	20.47	51.17	48.97	21.17
SL4	57.56	51.94	23.42	66.28	63.34	25.89
SL5	25.54	18.29	21.08	38.89	33.76	26.67
SL6	66.61	64.60	20.72	67.34	67.42	21.29
SL7	71.71	68.58	20.16	78.89	78.73	22.50
SL8	36.67	32.35	19.91	38.99	33.98	21.35
SR1	75.48	79.62	28.22	55.50	56.73	30.91
SR2	62.74	59.37	19.69	55.45	53.02	19.26
SR3	54.14	52.65	23.02	52.95	49.79	22.94
SR4	61.93	59.98	22.12	67.77	67.87	21.88
SR5	30.40	23.12	23.76	36.62	31.02	23.15
SR6	65.31	64.35	22.08	65.66	66.16	19.01
SR7	71.39	66.94	22.12	80.66	78.41	22.24
SR8	39.29	34.94	17.44	39.09	37.31	14.24

Each subset is cross-validated with other subsets such that 90% of data is used for training and 10% of data is utilized for testing. As a result, it avoids data overfitting and guarantees that the already trained and tested data will not be included for the next subset. Furthermore, by applying the Shapiro-Wilk test with the significance level of 5%, the normal distribution of the gait attributes is assessed. Figure 2.4 illustrates the foot pressure output of VGRF sensor for both healthy and PD subjects, and Figure 2.5 shows the input matrix obtained from 16 sensors for the 166 subjects based on the H & Y scale for

the four stages of PD. Since the no of samples in the early stages of PD is significantly high, it can assist to reasonably capture the relationship between the input and output feature sets.

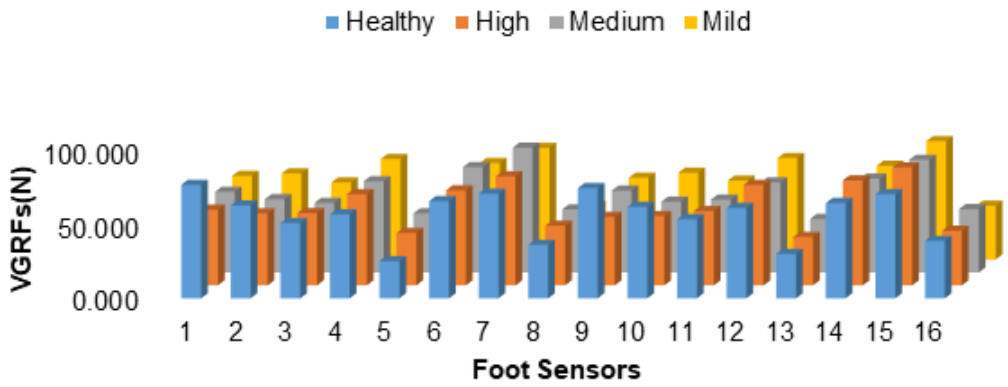


Figure 2.4: Mean plantar force of 16 sensors

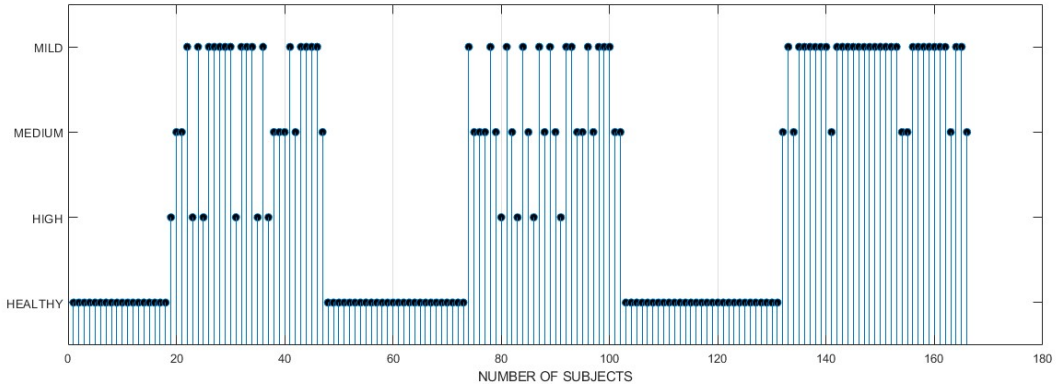


Figure 2.5: Classifier input pattern (16x166 matrix) with stages

2.4 Classification algorithms

2.4.1 Decision tree

Decision tree (DT) classifier, which is a supervised machine learning algorithm, is an effective and simple technique that formulates the classification model in the form of a tree structure. Dividing the dataset into smaller subsets

in the form of nodes and branches, DT iteratively determines the nonlinear relationship between the input and output of the system. In this study, DT classifies the 4 event responses namely healthy, mild, medium and high based on the H & Y scale. Figure 2.6 shows the tree structure of the DT used for PD classification.

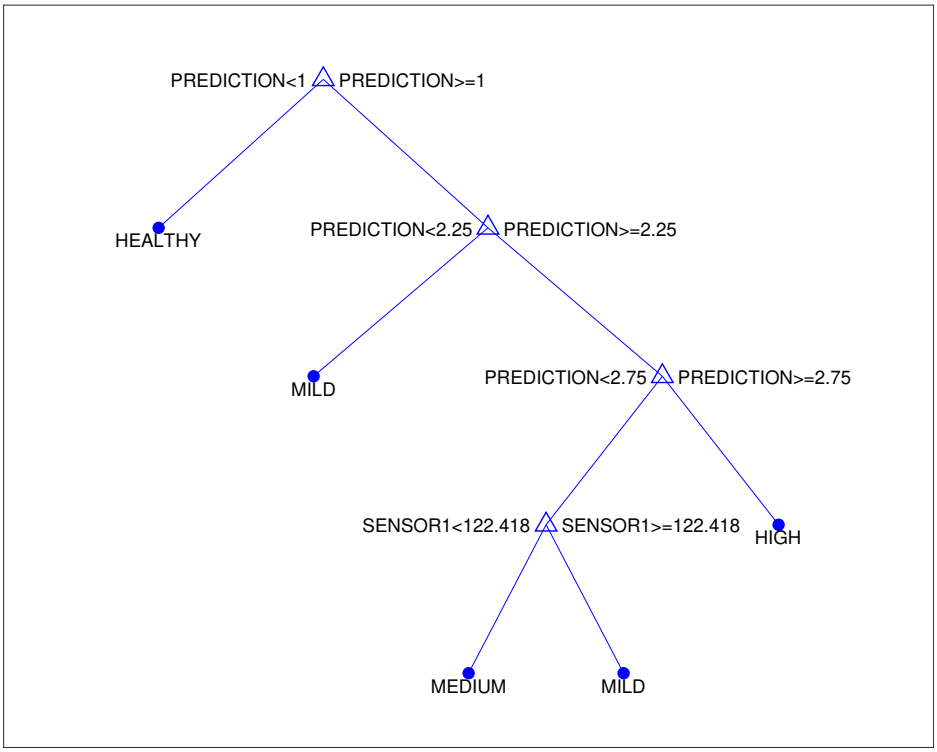


Figure 2.6: Tree structure of PD classification using DT algorithm

2.4.2 SVM Classifier

SVM uses a hyperplane to determine the decision boundaries which separate between data points of different classes. Aiming at formulating a hyperplane with the maximal distance between two classes, SVM works based on the maximum margin principle. The key advantage of SVM is that it can handle both linear and nonlinear classification problems (Luts et al. 2010). The fundamental idea behind SVM is to find a mapping between the original data

points from the input space to a higher dimensional feature space using the suitable kernel function. For a training data set $\{xi, yi\}_{i=1}^{n}$ with the input vectors being $x_i \in \mathbb{R}^d$ and the class labels being y_i, SVM attempts to map d-dimensional input vector x to the d_h dimensional feature space using the kernel function $\varphi(.) : \mathbb{R}^d \rightarrow \mathbb{R}^{d_h}$. The hyperplane which separates the classes in the feature space can be defined as $w^T \varphi(x) + b = 0$, with $b \in \mathbb{R}$ and w being an unknown vector which has the same dimension of $\varphi(x)$ (Ghaddar & Naoum-Sawaya 2018).

Figure 2.7 illustrates the SVM classifier for a binary classification problem. A data point in input vector x is allotted to the first class if $f(x)$ results in +1 otherwise to the second class if $f(x)$ yields -1. Particularly, when the separation surface is nonlinear, SVM uses the kernel functions to map the input vectors to a high dimensional feature space (Korkmaz et al. 2014). Oftentimes, as the perfect separation between two classes is not feasible, to account for the errors in the classification of data points the slack variables are utilized. Hence, the standard SVM problem can be defined as the following quadratic convex problem.

$$\min_{w,b} \frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i, \tag{2.1}$$

such that

$$y_i(w^T \varphi(x_i) + b) \geq 1 - \xi_i, \quad i = 1, 2, \ldots, N \tag{2.2}$$

$$\xi_i \geq 0 \quad i = 1, 2, \ldots, N \tag{2.3}$$

where C indicates the positive regularization constant, which defines the balance between maximizing the separation margin and minimizing the classification error.

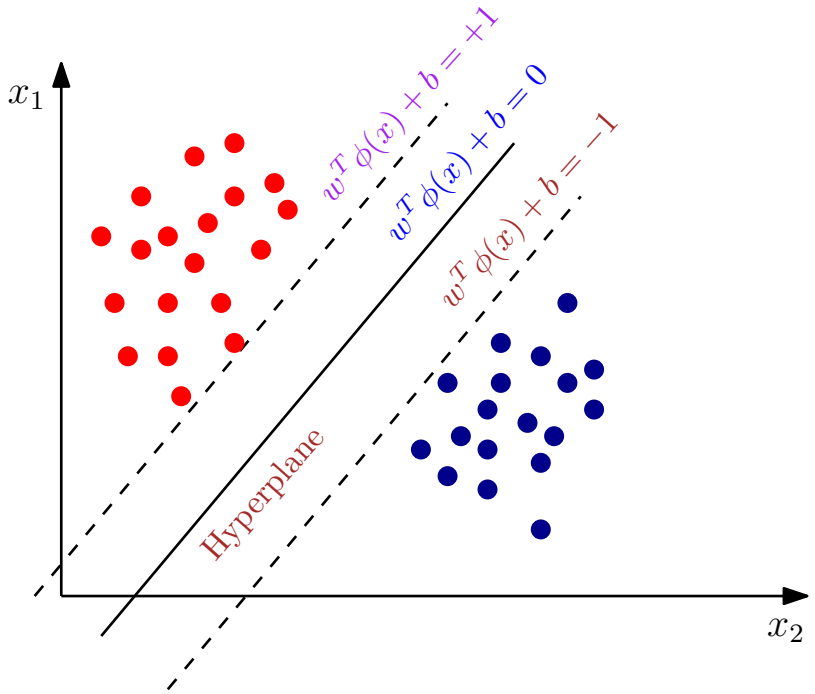


Figure 2.7: SVM classifier

2.4.3 Ensemble Classifier

Ensemble classifier is a learning model in which multiple classifiers are integrated to enhance the generalization capability of a single classifier. The idea of combining multiple classifiers was put forward by Huang & Suen 1993, and later in 1998, it was extended by Kittler 1998, who assessed the performance of multiple classifiers and presented an integrated theoretical framework. Normally, two major tasks are involved in creating a framework for successful ensemble classifier: 1. formulating a suitable set of classifiers and 2. identifying a combination technique, which aggregates different classifier outputs to find a suitable class. In this work, bootstrap aggregating (Bagging), which is a parallel learning technique, is used. Firstly, for a data set, which consists of *n* samples, *m* random samples are extracted and used for training the base classifiers. Then using a majority voting the results of base classifiers are combined to improve the accuracy of the classification framework.

2.4.4 Naive Bayes Classifier

Naive Bayes (NB) classifier is a probabilistic classifier which relies on two fundamental assumptions: 1. the attributes (features) are independent of each other and 2. each attribute has the same importance. The advantage of NB classifier is that it works well in multiclass prediction and requires less training data compared to other similar models such as logistic regression (Mughal & Kim 2018). Hence, it is one of the widely preferred classifier algorithms in medical expert systems for different clinical applications. For the given input dataset $X = [x_1, x_2, ...x_n]$ with n features, NB classifier determines the class C_k based on the Bayes theorem as follows.

$$p(C_k|X) = \frac{p(X|C_k)p(C_k)}{p(X)} \tag{2.4}$$

where $p(C_k|X)$ is the posterior probability, $p(C_k)$ is the marginal probability and $p(X|C_k)$ is the conditional probability. The data point with the highest conditional probability $p(C_k|X)$ is assigned to the class C_k as follows.

$$\hat{C} = argmax\ p(C_k)\prod_{i=1}^{n} p(X_i|C_k) \tag{2.5}$$

where $\hat{C}$ represents the predicted class for X with the given features $x_1, x_2, ..., x_n$.

2.5 Classifier performance metrics

The performance metrics used for assessing the classifier efficacy are accuracy, sensitivity, specificity, positive predictive value (PPV) (or) precision, negative predictive value (NPV) and F-score. Accuracy, which is the simplest and common measure to assess the classifier, is defined as the degree of correct predictions of the model based on true positive (TP), true negative (TN), false positive (FP) and false negative (FN) values.

$$\text{Accuracy}(\%) = \frac{TN+TP}{TN+TP+FN+FP} * 100\% \tag{2.6}$$

Sensitivity measures the proportion of true positives to the total number of actual positives. It indicates how good the diagnostic test can identify the normal (negative) condition.

$$\text{Sensitivity}(\%) = \frac{TP}{TP+FN} * 100\% \tag{2.7}$$

Specificity characterizes the maximum negative likelihood ratio (NLR) of true negative prediction to actual negatives (Abdulhay et al. 2018).

$$\text{Specificity}(\%) = \frac{TN}{TN+FP} * 100\% \tag{2.8}$$

PPV, which is also referred to as precision, represents the proportion of positive predictions to all actual positives. As it conveys how many of diagnostic test positives are true positives, the higher the PPV, the better the results.

$$\text{PPV}(\%) = \frac{TP}{TP+FP} * 100\% \tag{2.9}$$

NPV indicates the proportion of negative predictions to all actual negatives. It describes how many of test negatives are true negatives.

$$\text{NPV}(\%) = \frac{TN}{TN+FN} * 100\% \tag{2.10}$$

F-score, which conveys the accuracy of the model, is the weighted harmonic mean of precision and sensitivity.

$$\text{F-score}(\%) = 2 * \frac{\text{Precision*Sensitivity}}{\text{Precision+Sensitivity}} * 100\% \tag{2.11}$$

2.6 Results and Discussion

The proposed PD stage classification approach is implemented in MATLAB 2018 with the following parameter settings for the classifiers. In DT classifier, Gini's diversity index is chosen as a split criterion with the maximum number of splits being 100 and the surrogate decision splits per node being 10. For SVM, the linear kernel function with box constraint level of 3 and a kernel scale of 4 is configured. In the case of EC, the number of learners is assigned as 30, and the maximum number of splits is set to 165 with the learning rate of 0.1. Moreover, the bagged tree ensemble method is chosen for multi-class prediction. In Bayes classifier, the Gaussian kernel function with unbounded support vector is configured, and the multivariate multinomial predictor is set for categorical predictions. The classifier algorithms are run 10 times independently to get an average performance estimate.The training time of the four classifiers namely DT, SVM, EC and BC are 8.06s, 10.04s, 13.09s and 3.19s, respectively.

2.6.1 Performance assessment for statistical features

Figure 2.8 shows the scatter plot of each classifier with its respective stage classification based on the SL6 and SL7 sensor data, which have significant fluctuations for all stages of PD. The four stages of dataset include healthy-73, high-10, medium-28 and mild-55. The DT classifier excluding one medium class event, predicts all other events correctly, whereas in SVM, even though all the high events are properly classified, the prediction rate is relatively less in the case of medium and mild events. For instance, 3 medium and 2 mild data variables are unpredicted in the case of SVM. The EC out of the 73 healthy subjects, correctly identifies 72 as healthy subjects and gives one incorrect response. Similarly, in the high, medium and mild events, unpredicted values are 2, 3 and 2, respectively. Among all four classifiers, the BC provides the least prediction results with the incorrect classification in the case of healthy being 4

and medium being 25.

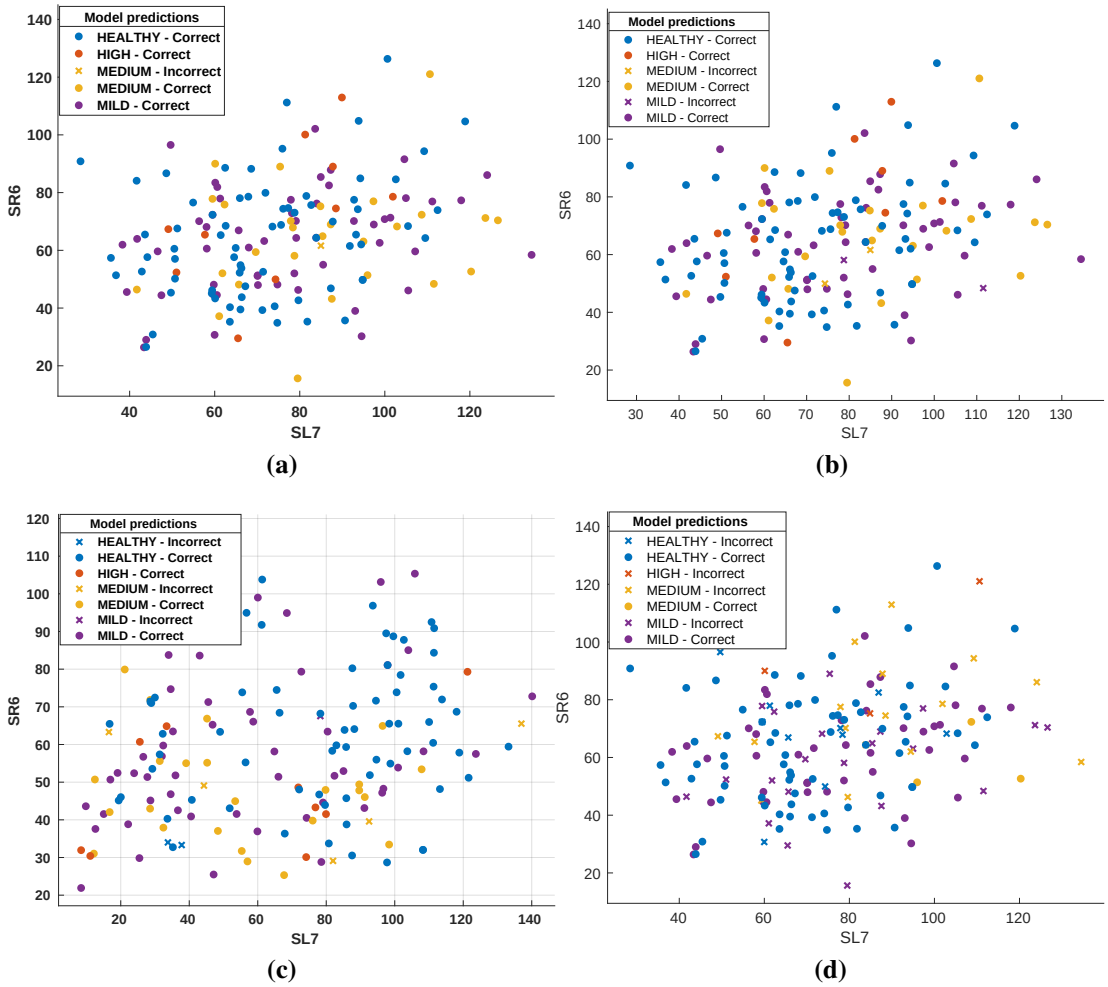


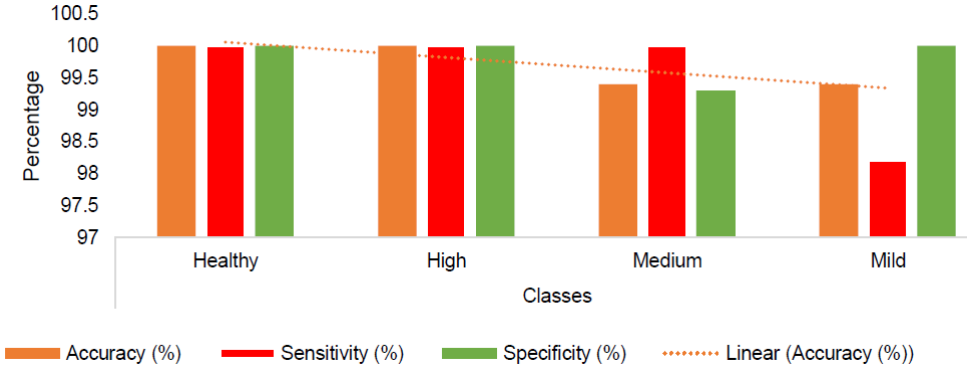
Figure 2.8: Scatter plot of (a) DT (b) SVM (c) EC (d) BC

Table 2.5 presents the classification accuracy along with the sensitivity and specificity for each stage of the subject for four classifiers. The three key performance metrics for each classifier is illustrated in Figure 2.9. Moreover, to visualize the cumulative performance of the classifiers for each stage of the subject, the confusion matrix, which is a table layout that gives the summary of the predicted results and the actual results, is shown in Figure 2.10. From the cumulative performance metrics for each classifier computed from the confusion matrix and presented in Table 2.6, we can read that in the case of DT, except

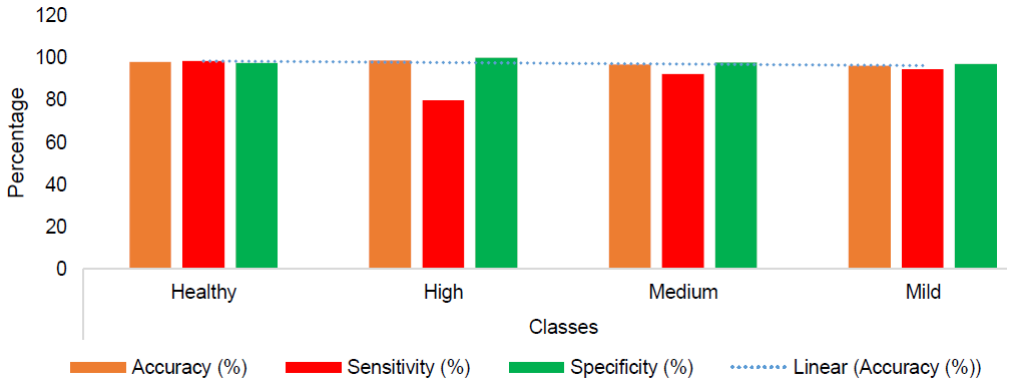
the medium event, all other events are effectively classified. Even though the misclassification rate in the case of medium event is 4 %, DT classifier offers an overall accuracy of 99.4 %. Moreover, with the least NPV of all classifiers, DT guarantees the lowest misclassification rate for all the four stages considered for classification. SVM which is the second best classifier gives the accuracy of 97.6%, followed by EC and BC with the accuracy of 95.1% and 69.7% respectively.

Table 2.5: Classifier performance metrics for four stages based on statistical features

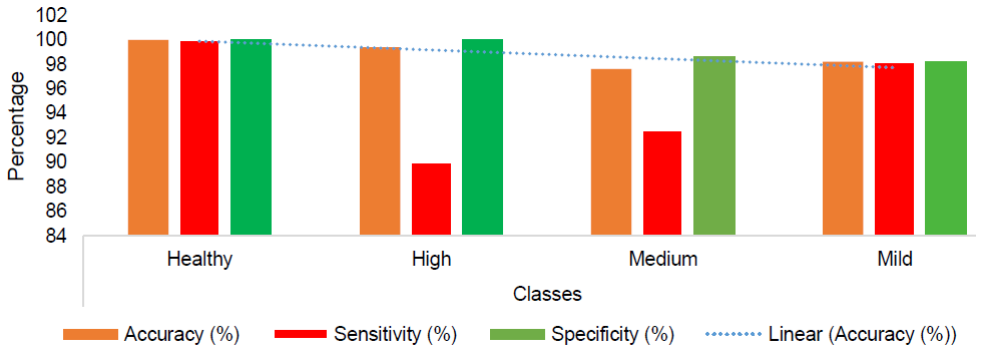
Classifier	Stage	TN	TP	FN	FP	Accuracy	Sensitivity	Specificity
DT	Healthy	73	93	0	0	100	100	100
	High	10	156	0	0	100	100	100
	Medium	27	138	1	0	99.4	100	99.3
	Mild	55	110	0	1	99.4	98.2	100
SVM	Healthy	73	93	0	0	100	100	100
	High	9	156	0	1	99.4	90.0	100
	Medium	25	137	2	2	97.6	92.6	98.6
	Mild	55	108	2	1	98.2	98.2	98.2
EC	Healthy	72	91	2	1	98.1	98.6	97.7
	High	8	156	0	2	98.8	80.0	100
	Medium	25	136	3	2	96.9	92.6	97.8
	Mild	53	107	3	3	96.3	94.6	97.2
BC	Healthy	69	83	10	4	91.6	94.5	89.2
	High	0	153	3	10	92.2	0	98.1
	Medium	3	123	16	24	76.0	11	88.5
	Mild	44	89	21	12	80.1	78.6	80.9



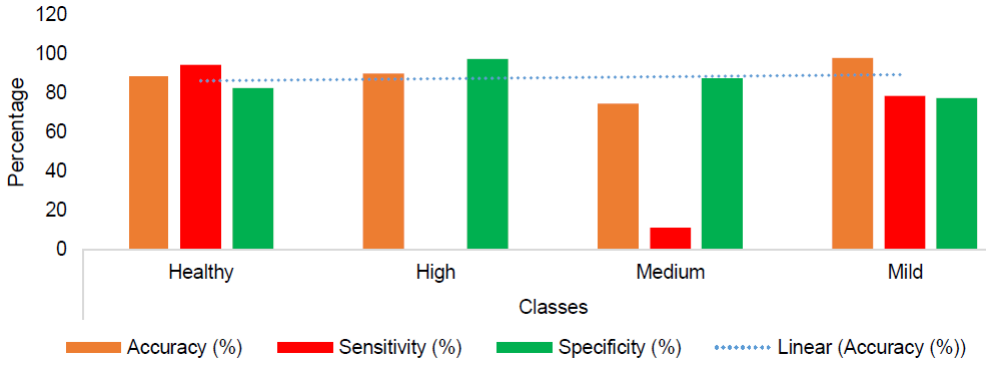
(a)



(b)



(c)



(d)

Figure 2.9: Performance metrics of classifiers for statistical features (a) DT (b) SVM (c) EC (d) BC

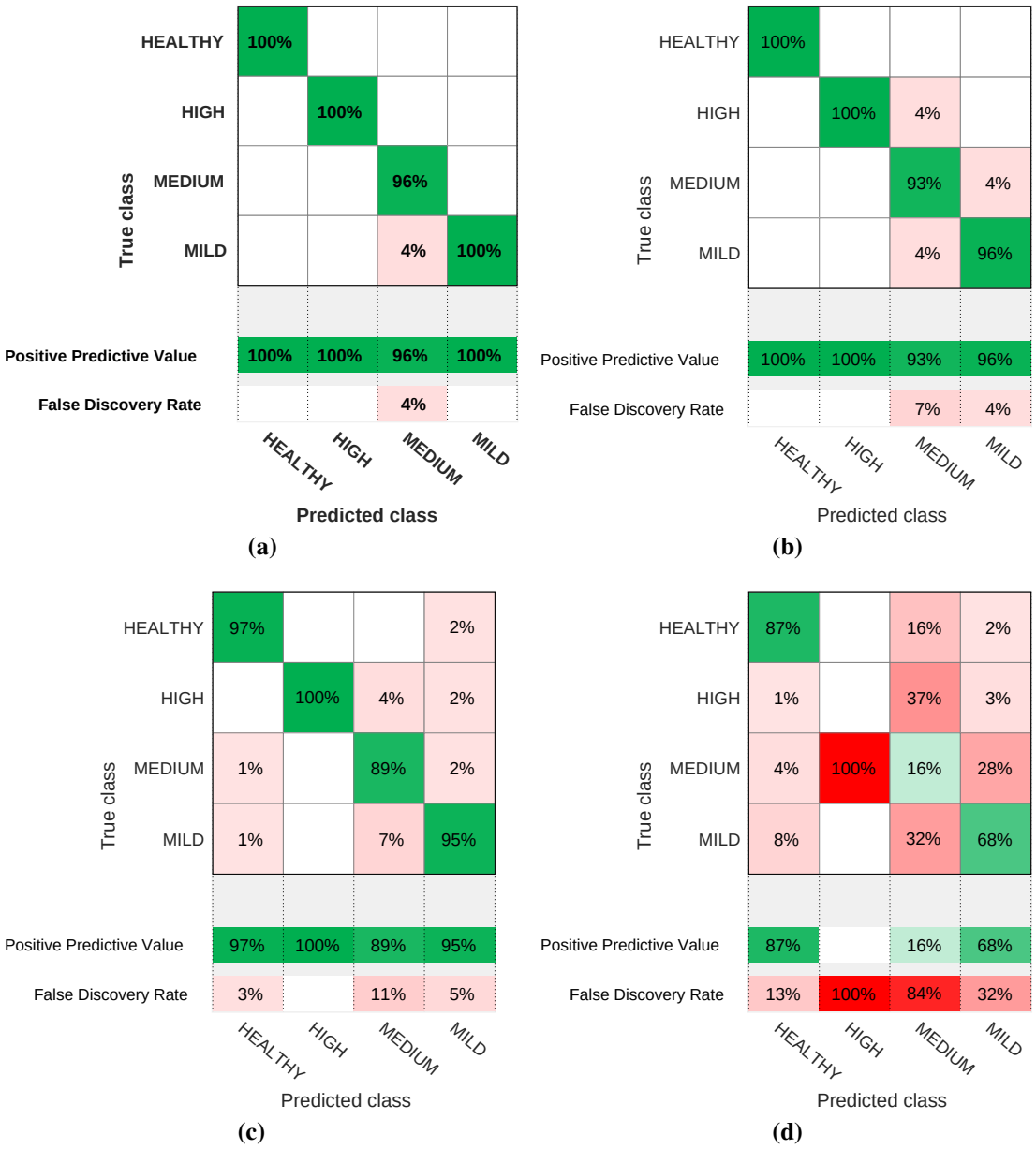


Figure 2.10: Statistical analysis – Confusion matrix (a) DT (b) SVM (c) EC (d) BC

Table 2.6: Cumulative performance metrics of classifiers for statistical features

Performance Measures	DT	SVM	EC	BC
Accuracy (%)	99.4	97.6	95.1	69.7
Sensitivity (%)	99.6	95.2	91.5	46.0
Specificity (%)	99.8	99.2	98.2	89.2
PPV (%)	99.0	97.8	94.6	67.6
NPV(%)	1	2.2	5.4	32.3
F-score (%)	99.25	97.3	94.6	70.7

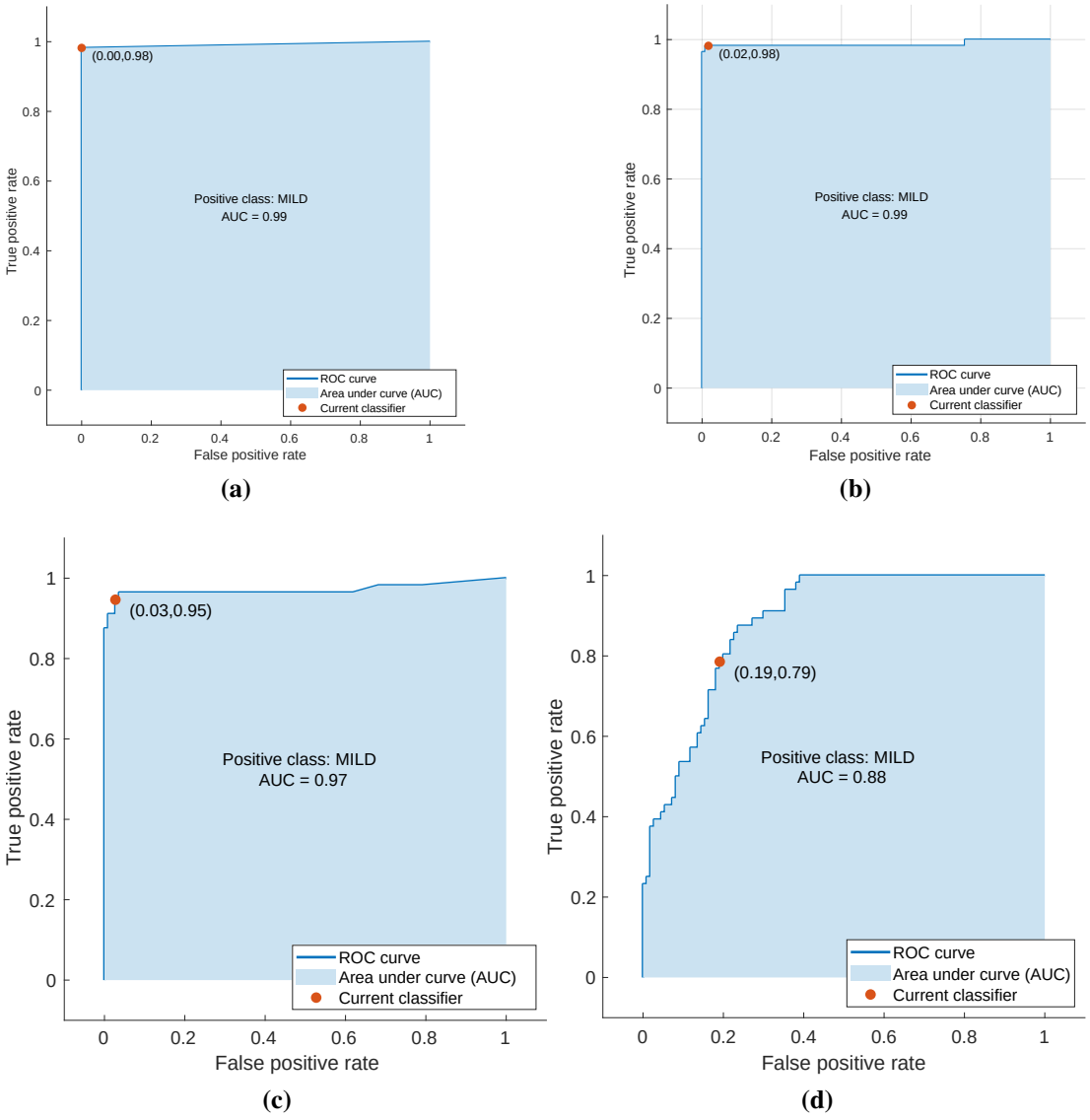


Figure 2.11: Statistical analysis – ROC for mild event (a) DT (b) SVM (c) EC (d) BC

In addition to the accuracy, sensitivity and specificity, an another important measure to assess the quality of classification models is area under curve (AUC) value. The receiver operating characteristic (ROC) curve, which is a plot between the false positive rate (FPR) and true positive rate (TPR), shows the trade-off between sensitivity and specificity. It indicates the degree of separability and conveys how effectively classifiers can distinguish the different classes. The larger the AUC, the better the accuracy of the classifier. For brevity, the AUC of mild event is illustrated in Figure 2.11. It is to be remarked that even though the classifiers effectively differentiate the stages of the subjects based on statistical features, the major limitation is that the VGRF sensor data may vary based on the weight of the subjects. The statistical analysis of VGRF sometime may not act as a real distinguisher of identifying the gait disorder in PD subjects. Hence, in order to address this issue in the following section we also analyze the walking pattern of the subjects using the kinematic analysis and extract 9 spatiotemporal features from the walking pattern and select the optimal biomarkers to differentiate the healthy subjects and PD subjects.

CHAPTER 3

KINEMATIC ANALYSIS OF GAIT PATTERN

3.1 Introduction

The kinematic gait analysis helps to assess progression of PD. PD affects the regular walking pattern of the subject and drastically reduces the number of footsteps, speed, step time, and stride interval. Hence, the study of gait cycle parameters such as step time, stride time, stance time, swing time, swing stance ratio, cadence, speed, step length and stride length, which act as significant biomarkers, help to understand the gait pattern disorder of PD subjects. Since feature selection plays a significant role in identifying the most discriminative features for efficient classification model, in this study, a correlation based feature selection technique is used to identify a set of informative and prominent gait attributes for efficient stage classification. Figure 3.1 illustrates various gait parameters of a subject during normal walk. The gait pattern contains 62% of stance interval and 38% of swing interval. In this work, the gait parameters are split into two components namely temporal and spatial features, which are briefed below.

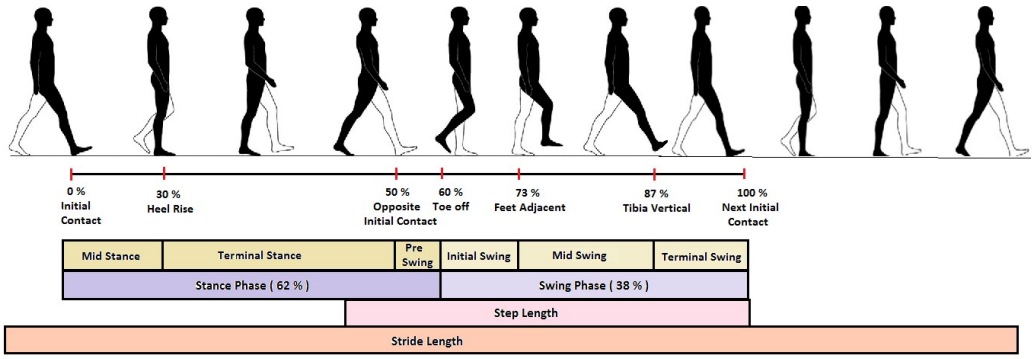


Figure 3.1: Different phase representation of gait pattern

3.2 Related Work

Supervised machine learning techniques can learn the inherent correlations between high dimensional spatiotemporal data using a training dataset and apply the knowledge gained during training to a new dataset to implement an automated stage classification of PD (Prashanth et al. 2016). Utility of machine learning algorithms in pathological gait has been studied using several machine learning techniques including decision tree (DT), SVM, principal component analysis (PCA), and random forest (RF). Some of the machine learning based PD gait classification techniques that have used wearable sensor data are briefly reviewed.

Introducing a hybrid model called, locally weighted random forest (LWRF), Aşuroğlu et al. 2018 presented a computational solution to quantitatively monitor the PD motor symptoms. Through regression analysis of ground reaction force (GRF) signals obtained from force resistive sensors (FSR), they showed that the PD severity level could be identified rather than a categorical result based on motor disorders. Prashanth & Roy 2018 presented diagnostic models to estimate the stage and severity of PD using the ML algorithms such as ordinal logistic regression (OLR), SVM and Random forest. Validating the performance on the dataset from Parkinson's Progression Markers Initiative (PPMI), which is a large-scale, comprehensive and multicentre to find PD progression biomarkers, they obtained the classification accuracy of 97.46%. Abdulhay et al. 2018 utilized a peak detection algorithm to extract the kinetic features including the heel and toe forces and analyzed the severity level of PD using fast Fourier transform of tremor signals. Through a pulse duration algorithm for acquiring the temporal features, they achieved an average classification accuracy of 92.7% using the medium Gaussian SVM. To normalize the subject's age, height, weight and walking speed, Wahid et al. 2015 used a multiple regression normalization strategy and identified the differences in spatiotempo-

ral gait parameters between healthy controls and PD patients. Subsequently, quantifying the raw gait data based on the kinematic analysis, they reported 92.6% classification accuracy using the random forest algorithm. Some of the other notable approaches reported in the literature in the direction of gait variability assessment are deep learning techniques (El Maachi et al. 2020), wavelet analysis (Joshi et al. 2017), deterministic learning theory (Zeng et al. 2016) and convolutional neural network (Zhao et al. 2018b).

In spite of the substantial amount of computational research on diagnosis of PD using gait analysis, the quantitative assessment of PD symptoms for improved stage classification needs finer investigation on hidden gait biomarkers. Hence, in this work, we extract the significant spatiotemporal features to identify the optimal feature vector for the ML algorithms. In the following section, we briefly define the spatiotemporal features considered for gait classification.

3.2.1 Temporal features

Step time: It is the time duration between heel strike of one leg and heel contact of the opposite leg. The approximate step time of healthy subject is 0.8 s.
Stance time: It is the time interval between the contact of the heel to the surface and the contact of the opposite leg when the toes do not touch the surface. Stance phase starts with the heel strike, and the mid stance is a complete body weight balance in one leg when another leg is in swing phase.
Swing time: Swing time is observed between the heel strike and toe off phase. Swing phase starts at the end of stance.
Swing stance ratio (SS ratio): It is a ratio between the swing time and stance time.
Stride time: It is the time duration between foot ground contact of successive instants of same foot.

Cadence: It is the total number of steps/unit time. The nominal values of cadence are: slow :60-70 steps/min, medium: 70-90 steps/min, and normal:90-115 steps/min.

Speed: The total distance covered / unit time gives the speed, and it is measured in m/sec. Typical values for male and female are 80 m/min, and 70 m/min, respectively.

3.2.2 Spatial features

Step length: It is the distance between the successive footsteps of heel contact of two foot steps. Typically, the right foot step length is equal to left foot step length.

Stride length: It is the distance between the successive foot of heel contact of the same foot, and it is also called as double step length.

Table 3.1 gives the 9 temporal and spatial features of 93 PD and 73 healthy subjects based on the sum of left and right feet sensor values. It can be noted that the swing time remains almost same for different stages of PD, whereas the stance time has relatively considerable fluctuations. The highest variation can be observed in the features such as cadence, stride length, stride time and speed. Figure 3.2 shows the temporal and spatial features of both healthy and PD subjects for different stages. The spatiotemporal features highlight that the cadence is the most influential factor which has larger variability due to the severity of PD. Even though the mild and medium stages of PD have similar cadence, compared to cadence of healthy control, which is 110.325, subjects with high severity of PD exhibit drastic reduction in cadence which is around 101.625. Similarly, the second most significant feature is the stride length, which reads for healthy-1.35, high-0.939, medium-1.115 and the mild-1.2.

Table 3.1: Temporal and spatial features of PD and healthy subjects

Spatiotemporal features	Healthy		Mild		Medium		High	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Swing Time (s)	0.399	±0.034	0.389	±0.051	0.378	±0.027	0.399	±0.041
Stance Time (s)	0.696	±0.060	0.847	±0.239	0.736	±0.083	0.713	±0.058
Stride Time (s)	1.095	±0.088	1.235	±0.269	1.114	±0.099	1.112	±0.087
SS Ratio (s)	0.575	±0.038	0.481	± 0.092	0.518	±0.053	0.561	±0.051
Step Time (s)	0.547	±0.044	0.618	±0.135	0.557	±0.049	0.556	±0.043
Step Length (m)	0.675	±0.068	0.469	±0.124	0.558	±0.098	0.600	±0.095
Stride Length (m)	1.35	±0.137	0.939	±0.248	1.115	±0.196	1.200	±0.191
Speed (m/s)	1.241	±0.160	0.808	±0.307	1.008	±0.183	1.083	±0.169
Cadence (steps/min)	110.325	±8.847	101.625	±21.521	108.481	±9.200	108.625	±9.006

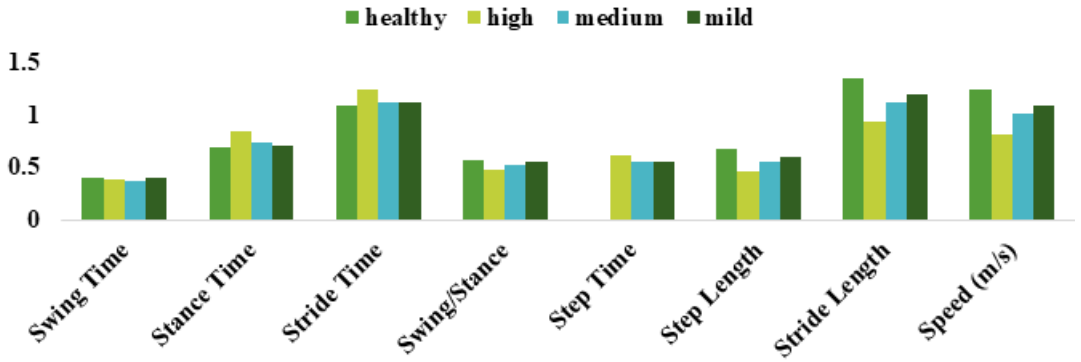


Figure 3.2: Temporal and spatial features of 4 classes

3.3 Proposed Methodology

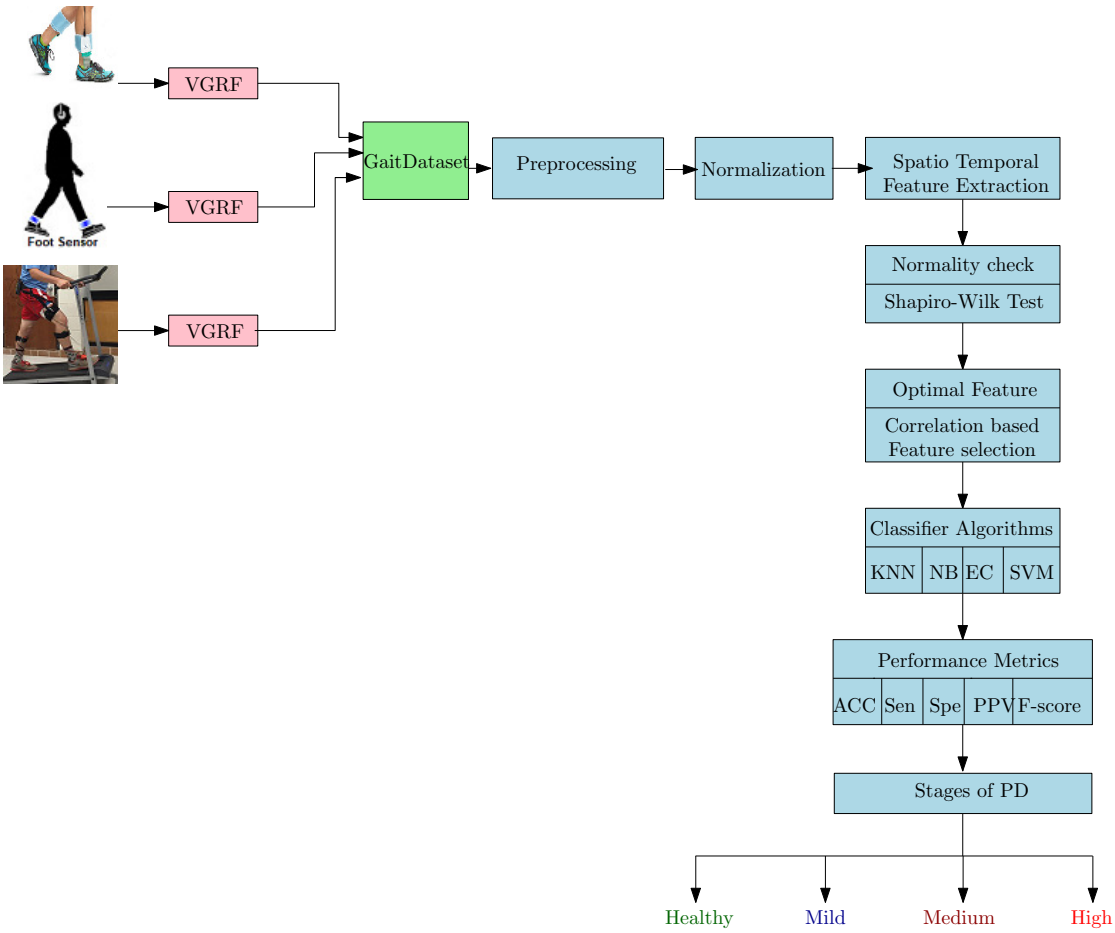


Figure 3.3: Proposed PD stage classification framework

Figure 3.3 shows the proposed schematic diagram for stage classification of PD based on H&Y scale. The major objective of this work is to analyze the gait variability of PD subjects from three different walking patterns in spatiotemporal domain and unveil the prominent biomarkers to distinguish the stages of PD based on motor symptoms. Unlike the previous study which used several gait features for severity rating, the proposed work identifies the significant spatiotemporal features like step time, stride time, stride length and cadence based on their rank correlation for early detection and classification. To avoid the gait initiation and termination effects, the first 20s and the last 10s data were removed. To minimize the dispersion in gait data and decorrelate the gait

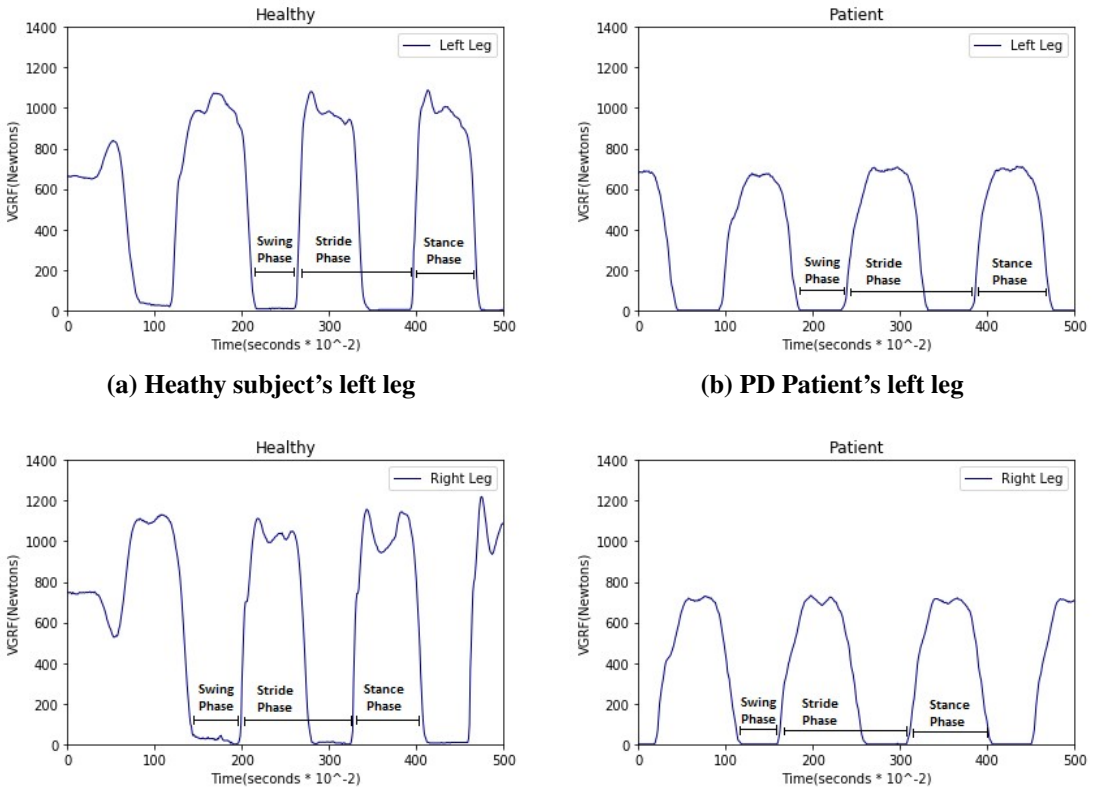
data from physical parameters like height and weight, the following multiple regression technique is employed.

$$y_i = \beta_0 + \sum_{i=1}^{n} \beta_j x_{i,j} + \varepsilon_i \tag{3.1}$$

where y_i indicates the dependent spatiotemporal gait feature of the i^{th} observation; $x_{i,j}$ denotes the j^{th} physical characteristics like walking speed, age, height and weight; β represents the unknown regression coefficients; and ε_i indicates the residual error for the i^{th} observation. After normalizing the gait data, the spatiotemporal features are extracted and the normal distribution of the gait features is assessed using the Shapiro-Wilk test with the confidence bound for the hypothesis test being 5%. For optimal feature vector selection, a Pearson rank correlation is computed for the spatiotemporal features. Finally, identifying the significant biomarkers based on rank correlation, the four supervised machine learning algorithms namely DT, SVM, EC and NB are employed to extract the hidden pattern in gait data and stage classification.

3.4 Classifier performance assessment

Figure 3.4 shows the kinematic features of healthy and PD subjects for both left and right legs. Comparing the left leg VGRF signals of healthy and PD subjects reveal that the signal amplitude for healthy are almost double as PD subjects amplitude. In addition, the stride phase of PD subjects is relatively less compared to healthy because of the movement disorder. To select optimal feature set from spatiotemporal features, a Pearson rank correlation based feature selection technique is used. Among the 15 spatiotemporal features, 9 features which have significant correlation are chosen.



(a) Heathy subject's left leg **(b) PD Patient's left leg**

(c) Heathy subject's right leg **(d) PD Patient's right leg**

Figure 3.4: Kinematic parameters of healthy subjects and PD patients

Figure 3.5, which shows the correlation matrix plot, reveals that the spatiotemporal features have both positive and negative correlation. For instance, the step time and stride time are highly positively correlated. However, the cadence and stance time are highly negatively correlated.

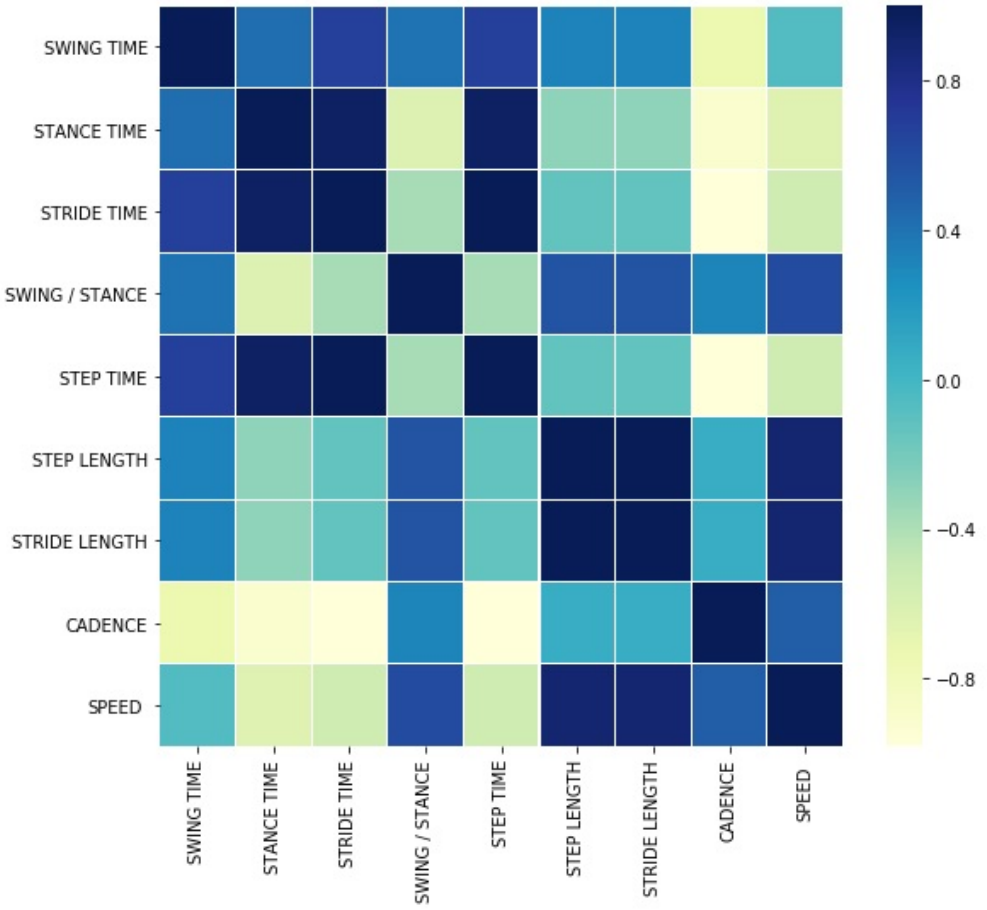
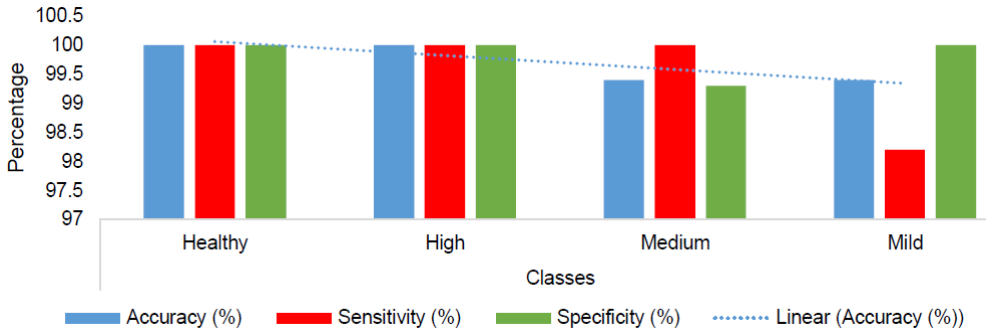


Figure 3.5: Correlation matrix plot

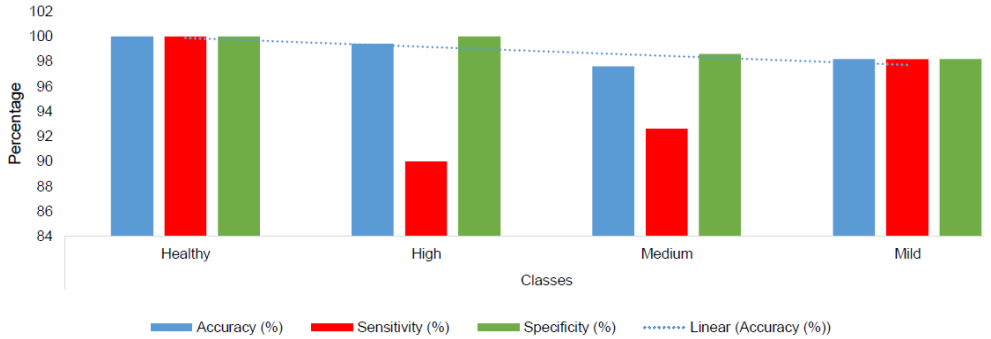
Table 3.2 presents the performance metrics for each stage of the classification assessed using the kinematic features, and the three metrics are illustrated in Figure 3.6. The confusion matrix, shown in Figure 3.7, highlights that the positive predictive value and false discovery rate of DT and SVM are same for the kinematic features. Excluding the medium stage of PD, all other stages are classified with 100% accuracy. Hence, DT and SVM offer the least misclassification rate of around 4%. However, the misclassification rate of BC, which is around 30%, is the highest at each stage.

Table 3.2: Kinematic features based performance metrics for four stages of classification

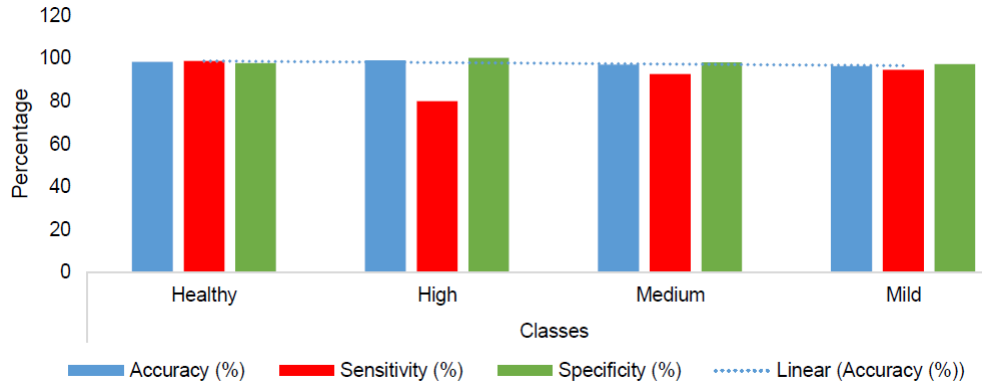
Classifier	Stage	TN	TP	FN	FP	Accuracy	Sensitivity	Specificity
DT	Healthy	73	93	0	0	100	100	100
	High	10	156	0	0	100	100	100
	Medium	27	139	1	0	99.4	100	99.3
	Mild	55	111	0	1	99.4	98.2	100
SVM	Healthy	73	93	0	0	100	100	100
	High	10	156	0	0	100	100	100
	Medium	27	138	1	0	99.4	100	99.3
	Mild	55	110	0	1	99.4	98.2	100
EC	Healthy	72	91	2	1	98.1	98.6	97.8
	High	8	156	0	2	98.8	80	100
	Medium	25	136	3	2	96.9	92.6	97.8
	Mild	53	107	3	3	96.3	94.6	97.2
BC	Healthy	69	83	10	4	89.2	94.5	82.5
	High	0	153	3	10	89.9	0	97.5
	Medium	3	123	16	24	74.4	11	87.6
	Mild	44	89	21	12	82.2	64.1	88.7



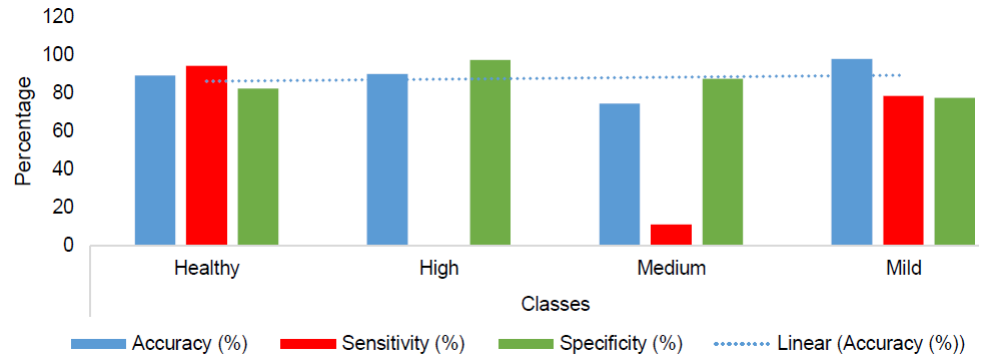
(a)



(b)



(c)



(d)

Figure 3.6: Performance metrics of classifiers - kinematic analysis (a) DT (b) SVM (c) EC (d) BC

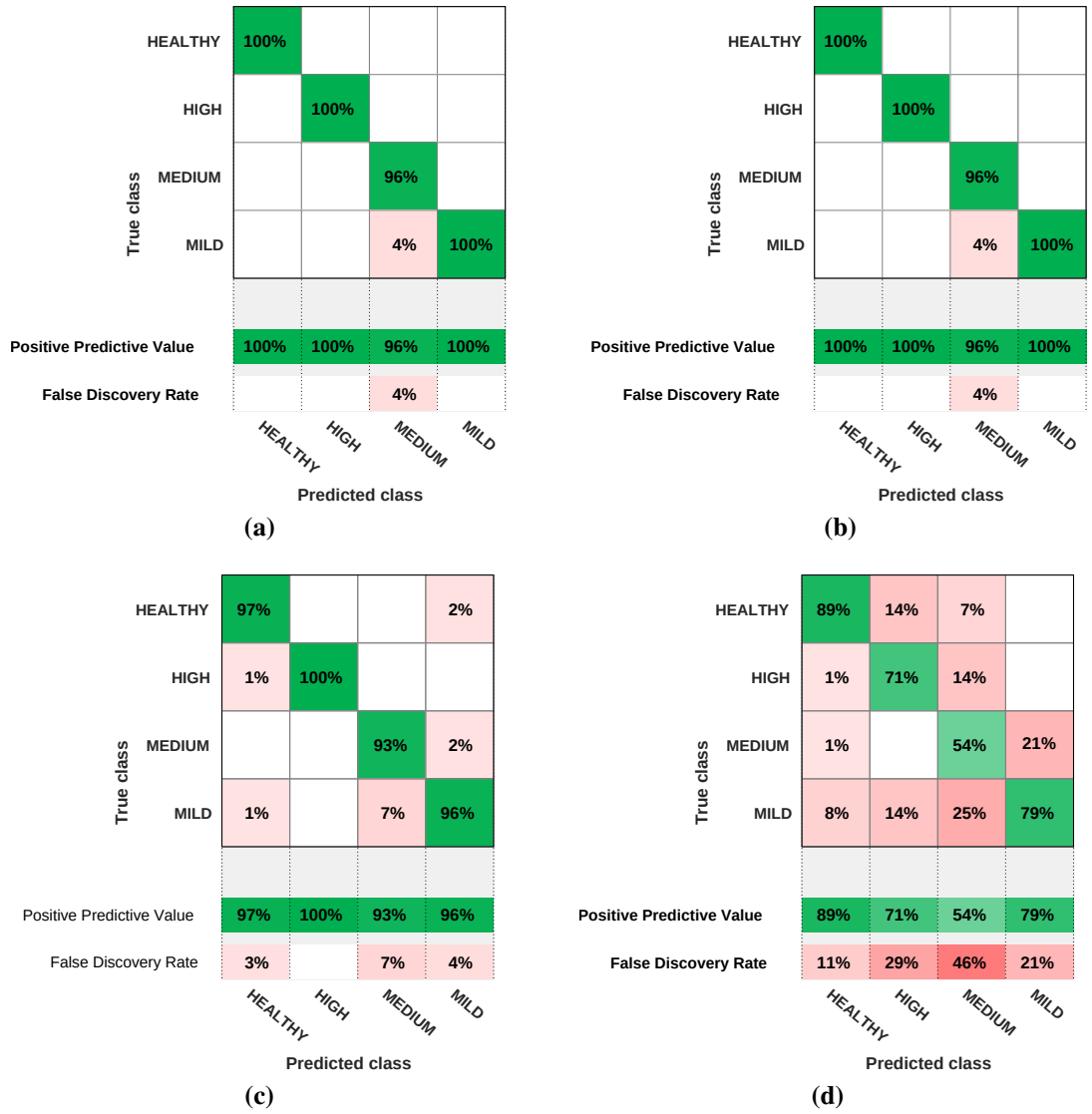


Figure 3.7: Kinematic analysis - Confusion matrix (a) DT (b) SVM (c) EC (d) BC

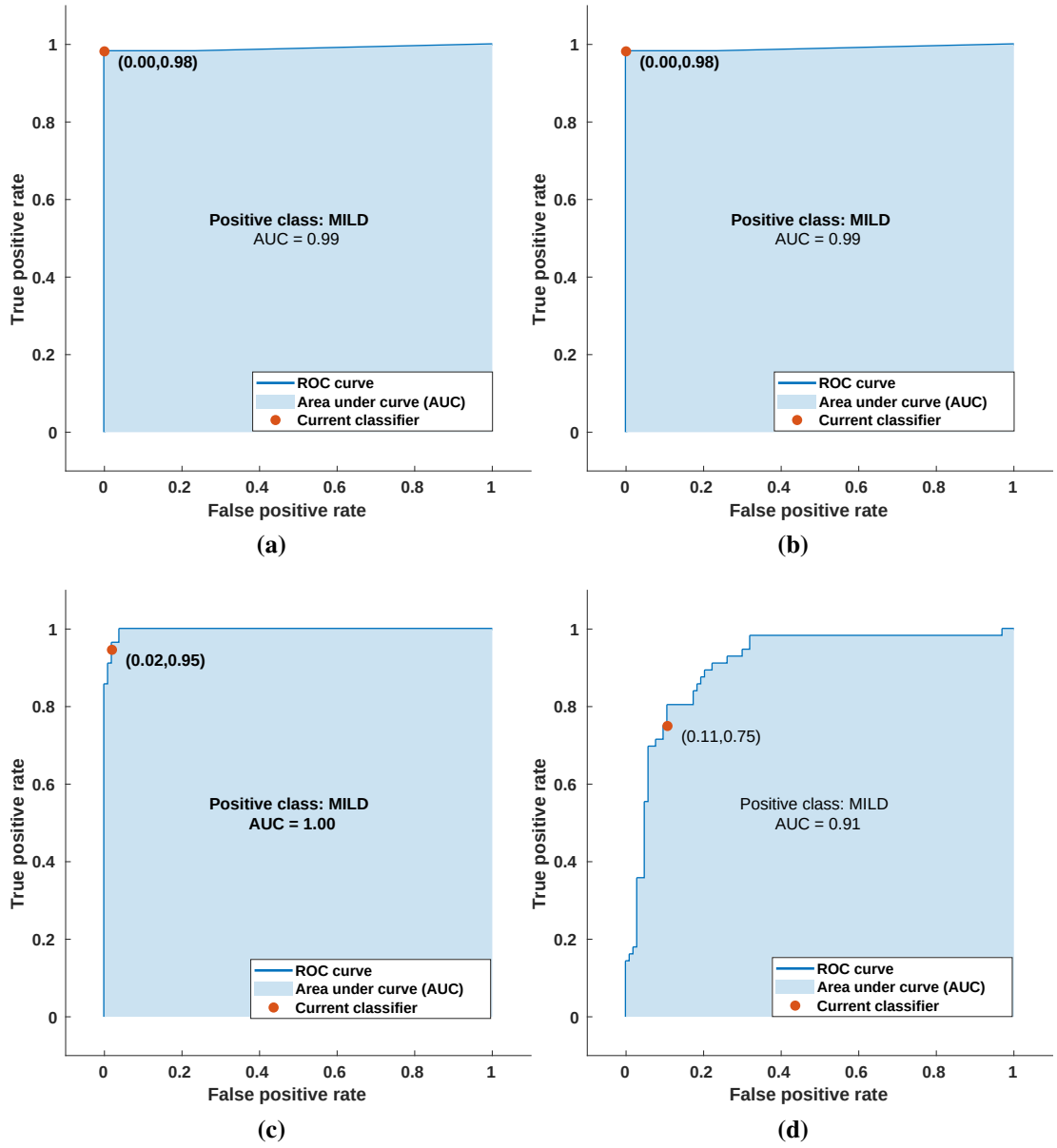


Figure 3.8: ROC curve for mild event (a) DT (b) SVM (c) EC (d) BC

Table 3.3: Cumulative performance evaluation of classifiers for kinematic features

Performance Measures	DT	SVM	EC	BC
Accuracy (%)	99.4	99.4	95.2	69.9
Sensitivity (%)	99.6	99.6	91.6	42.4
Specificity (%)	99.8	99.8	98.2	89.1
PPV (%)	99.0	99.0	95.3	65.7
NPV(%)	1	1	4.7	34.3
F-score (%)	99.25	99.25	93.1	72.1

Table 3.3 gives the cumulative performance metrics of each classifier for different stages of PD. Moreover, to highlight the misclassification for the mild stage of PD, the ROC and AUC of 4 classifiers are shown in Figure 3.8. To extend the AUC for a multi-class problem, we utilize the averaging pairwise comparison approach because it gives an overall measure of how effectively each class is differentiated from other classes.

Table 3.4: Comparison of accuracy of proposed and other reported approaches

References	Features	Classifiers	Accuracy
Khoury et al. 2019b	Time domain	K-NN, CART, SVM, and GMM	80-91 %
Elkurdi et al. 2018	Kinematic features	DT, SVM and KNN	86.19–92.9 %
Alam et al. 2017	Left and right VGRF signals	SVM, KNN, RFDTC	93.6 %
Martínez et al. 2018	Kinematic features	Linear Discriminant Model	64.1 %
Zeng & Wang 2015	Kinematic features	RBFNN	93.1 %
Zeng et al. 2016	Kinematic features	RBFNN	96.39 %
Perumal & Sankar 2016b	Time and Frequency domain	LDA, SVM and ANN	86.9 %
Prabhu et al. 2018	Time domain features	SVM, PNN	96.15 %
Proposed framework	Spatiotemporal features	DT	99.4 %

Finally, the performance validation of the proposed method to the state-of-the-art methods, which used the same gait dataset from Physionet, is reported in Table 3.4. It can be noted that the accuracy of the proposed approach is significantly better than those of the other methods and distinguishably the present work utilizes the minimum feature set formed using the spatiotemporal variables. Hence, utilizing the spatiotemporal features can significantly improve the classification accuracy and the proposed approach can serve as a tool for early diagnosis and severity rating of PD based on the clinical symptoms obtained through foot pressure analysis. In this direction, the major contributions of this work are as follows. 1. An automatic and non-invasive gait classification framework to predict the severity of PD using the spatiotemporal features obtained from the foot worn force sensor data is presented. 2. To minimize the dispersion in gait data and decorrelate the gait data from physical parameters like height and weight, a multiple regression normalization approach is used. 3.

The significant gait biomarkers from spatiotemporal features are obtained using the correlation based feature extraction technique . To avoid overfitting and improve the classification accuracy, 10 fold cross validation technique is used. 4. The performance of four supervised machine learning algorithms to predict the stages of PD based on H&Y scale has been assessed using the confusion matrix and area under curve.

3.5 Limitations

Even though the proposed kinematic analysis approach provides better classification accuracy compared to other methods reported on the same dataset, it is also important to mention the limitations of the proposed study. In the current study, we considered only the significant motor symptoms of PD for classification. However, to improve the prediction rate, the non-motor symptoms can also be considered. Moreover, in addition to the gait pattern, the tremor dataset can also be assessed for enhancing the accuracy of the stage classification. Finally, we have considered only the supervised linear classifiers for identifying the stages of PD. Nonetheless, the nonlinear classifiers can be employed for extracting the nonlinear relationships especially when the tremor data is considered along with the gait pattern.

CHAPTER 4

CNN FOR PD DIAGNOSIS

4.1 Introduction

From diagnostic point of view, gait analysis is an important approach because the PD patients manifest the gait abnormalities as one of the earliest symptoms(Alharthi et al. 2019). Hence, the gait assessment can be used for early diagnosis of PD. The Parkinsonian gait is primarily characterized by a large stride to stride variability, a slow gait cycle, a long stance phase and a flat foot-to-toe off. To identify the presence of PD, generally, the clinicians assess the walking pattern by visual inspection along with the UPDRS. Notwithstanding, the gait evaluation is highly subjective because it largely depends on the expertise of the clinician and there is no powerful tool available to predict the presence of PD (Alharthi & Ozanyan 2019; Schwab & Karlen 2019). Moreover, the progression of PD varies according to the age, health condition and other intrinsic factors. In order to alleviate the misdiagnosis problems due to manual assessment, automated systems are being designed to classify the PD from gait cycle. Even though gait analysis based PD diagnosis has attracted considerable attention, for quantitative assessment of PD and its stages, a finer investigation on the hidden gait biomarkers is needed for realizing an effective classifier tool that can assist the clinicians in their daily diagnosis of PD.

In literature, machine learning based feature extraction methods have been widely used for PD diagnosis.However, the manual preprocessing and feature extraction will always have limitations in their capacity to diagnose. Hence, in this work, to overcome such manual feature extraction, we put forward a novel

gait classification technique using deep convolutional neural network (DCNN) for automatic detection and severity rating of PD. The VGRF gait dataset obtained using foot worn sensors from Physionet is used for assessing the efficacy of the proposed approach (web 2019). Fusing the feature learning and classification tasks into one single entity, the DCNN avoids the need for hand-crafted feature extraction approach with the help of the convolution and pooling operations (Zhou et al. 2019). To test the performance, the experiments are implemented in Keras platform for different input frame sizes.

In this clinical context, the major contributions of this work are threefold. Firstly, we analyze 1D and 2D DCNN architectures for different frame size to assess the performance of one dimensional input signal method and two dimensional image input method in CNN framework. Secondly, the performance of the proposed DCNN architecture to classify the severity of PD is assessed based on modified H & Y scale using gait pattern obtained for three different walking scenarios such as 1. walking on a level ground 2. walking on a treadmill and 3. walking with rhythmic auditory stimulation. Thirdly, utilizing the stochastic gradient descent algorithm for CNN loss function optimization and L2 regularization technique to avoid data overfitting, we achieve an average stage classification accuracy of 98.45 % for PD stage prediction.

4.2 Related work

This section reviews related work on PD diagnosis especially based on gait signal analysis. Several methods to quantitatively assess the gait variability for PD diagnosis have been reported. Particularly, the gait assessment using the time-frequency domain, spatio-temporal domain using machine learning algorithms and the deep learning approach are discussed.

Kim et al. 2018 proposed a CNN structure for quantifying the sever-

ity of PD patients based on UPDRS scale. Using the custom designed wearable device fitted with a tri-axial accelerometer and a gyroscope, they collected tremor signals from 92 patients and transformed the accelerometer and gyroscope signals into a two-dimensional image representation for CNN input. They considered the categorical cross-entropy as the loss function and used 10 fold cross validation technique to avoid data overfitting. Through hyper-parameters of CNN that could discriminate the level of tremor severity, authors achieved a better classification accuracy of 85 % compared to other machine learning classifiers. However, the repeatability of the CNN model to quantify the tremor needs to be assessed when the change happens in patients state and environmental conditions.

To assess the neurological state of the patients and model the transitions during start and stop movements, Vásquez-Correa et al. 2018 proposed a deep learning architectures based multimodal assessment technique, fusing the information from three modalities: speech, handwriting and gait. The specific motor impairments in lower/upper limbs and in speech were examined to classify the stages of PD based on MDS-UPDRS scale. Based on the interpretation of the feature map obtained from CNN, they not only showed that the classification accuracy could be improved but also evaluated the robustness of their approach through the speech impairment analysis in three different languages: Spanish, German and Czech. It was reported that even though the gait and speech analysis could provide better results compared to handwriting assessment, to address the influence of the language further experiments on sentences, texts and spontaneous speech signals could be performed.

For diagnosing the presence of PD, Abdulhay et al. 2018 utilized a gait and tremor signals from Physionet and presented a peak detection and pulse duration measuring techniques to extract the spatiotemporal features. To pre-process the raw VGRF dataset, they employed a type II Chebyshev high pass filter and ob-

tained the kinematic features such as swing time, stride time, heel strike and toe off in order to build a feature set based on foot strike profile variability. Transforming the time series data into frequency domain using fast fourier transform (FFT) for distinguishing the amplitude variation, they used power spectral density of tremor signals to differentiate between tremor at rest and other tremors with respect to frequency characteristics. Medium tree and medium Gaussian support vector machine (SVM) were used to classify the healthy subjects and PD patients. However, this approach was limited to binary classification wherein only the presence of PD was diagnosed and the stages were not identified.

Pereira et al. 2018 put forward a hand written dynamics assessment technique for PD identification. Using the smart pen that captures the fingergrip, axial pressure, tilt and acceleration in X, Y and Z directions, they conducted six handwritten exams: 1. Drawing circles on a paper 2. Drawing spirals 3. Drawing in the air 4. Drawing meanders 5. Left-wrist movements and 6. Right-wrist movements. To capture different information from each exam and improve the classification accuracy, ensemble of CNNs, which were trained over six different handwriting exams, were used. By mapping the signals extracted from the time-series-based images drawn by control group and PD patients, they compared the performance of machine learning algorithms with those of the two distinct CNN architectures. Even though better classification accuracy was obtained using CNN architecture compared to machine learning approaches, the paper does not report any approach for reducing the dimensionality of the feature space.

For early detection and monitoring of PD, Perumal and Sankar investigated the effect of using both gait and tremor features obtained from the wearable sensors (Perumal & Sankar 2016a) . Extracting the kinetic and spatiotemporal features from gait cycle, they employed a linear discriminant analysis (LDA) classifier to distinguish between control subjects and PD patients.

In tremor analysis, to improve the frequency resolution and reduce the spectral leakage, a Haan window was utilized. One of the significant contributions of their work was that based on the frequency-domain characteristics they identified a PD tremor from other tremors due to atypical Parkinsonism. Testing the performance of the approach using the dataset from Physionet, they achieved an average classification accuracy of 86.9 %. Nevertheless, the approach is limited to binary classification problem and does not report the stages of PD based on any standard scale.

Several previous studies have reported results on the gait classification based on the VGRF dataset available in Physionet. Those studies have analyzed the hand-crafted features using conventional machine learning algorithms for classification. However, in spite of the great potential of deep learning in promoting bioinformatics and healthcare research, there are no significant results available in the literature on applying the deep learning methodologies for gait variability analysis and PD diagnosis. Hence, the major focal point of this work is to analyze the performance of the convolutional neural networks for classification based on both 1D and 2D architectures.

4.3 Methodology

4.3.1 CNN Overview

CNN, which derives a biological inspiration from the visual cortex, is a class of deep feed-forward artificial neural networks (El Maachi et al. 2020). The two major limitations of shallow neural networks which led to the development of CNN are as follows. Firstly, in shallow neural network the number of parameters is huge. For instance, consider a hidden layer consisting of N_1 nodes and the subsequent output layer having N_2 nodes. Then, the number of parameters between these nodes is $N_1 \times N_2$ which not only results in

large number of parameters but also causes overfitting issue and slows down the training and testing processes (Fawaz et al. 2019). Secondly, the shallow neural network ignores the correlation between the input features because it considers each input feature independently. Hence, to overcome these two issues CNN was put forward with two interesting features: sparse connectivity and weight sharing (Camps et al. 2018). CNN was introduced by LeCun in 1989. Hand written character recognition was the first successful real-world application of CNN demonstrated by LeCun through their proposal of LeNet in 1998 (Cao et al. 2018). However, due to black-box based processing, CNN did not receive much of research attention until the significant contribution by Krizhevsky et al in 2012 (Li et al. 2019). In the ImageNet Challenge in 2012, Krizhevsky et al. put forward AlexNet, which was similar to architecture in LeNet but deeper, for the large scale visual recognition competition and outperformed all of their competitors by achieving a winning top-5 test error rate of 15.3 % compared to the second best of 26.2 %. Indeed, it was the breakthrough in the area of deep learning and subsequently, other architectures based on CNN such as VGGNet (Simonyan & Zisserman 2014), GoogleNet (Szegedy et al. 2015), DenseNet (Huang et al. 2017), and Residual Net (He et al. 2016) were proposed.

In this work, the CNN model is employed for extracting the hidden biomarkers from gait pattern for effective severity rating of PD. Figure 4.1 shows the proposed DCNN architecture for predicting the stages of PD based on modified H & Y scale. The time series data obtained from the VGRF sensors is given as an input to DCNN. The DCNN has four types of layers namely convolution, pooling, fully connected and softmax.

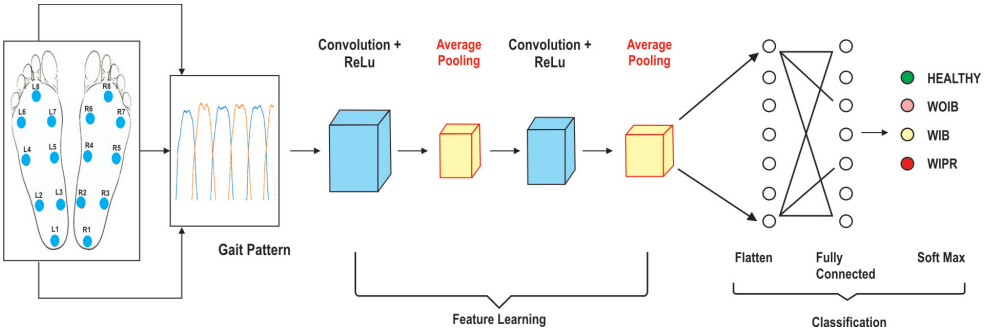


Figure 4.1: Proposed DCNN architecture for gait classification

4.3.2 DCNN architecture for gait classification

i. Convolutional layer: It consists of set of learnable filters that move over the input data based on the kernel function to capture the significant features. The input is convolved with the kernel function based on the following equation.

$$c_m = \sum_{n=0}^{N-1} f_n k_{m-n} \tag{4.1}$$

where k, c, N and f indicate input signal, convolved output, number of data points in k and filter, respectively. The subscript n denotes the n^{th} parameter of the filter vector and m indicates the m^{th} output element to be computed. When n varies from 0 to $N-1$, the filter function f_n is convolved with the input signal k_{m-n}. After the convolution operation, the convolved matrix is given to the rectified linear unit (ReLU) activation function to account for nonlinearities in time series data. ReLU, as illustrated in Figure 4.2, is a piecewise linear function which basically outputs 0 when it gets a negative input value and the value itself in any other case. ReLU does not saturate and is computationally very efficient (Arifoglu & Bouchachia 2019). Moreover, it converges faster than other nonlinear thresholding functions like $sigmoid/tanh$. The mathematical representation of ReLU is as follows.

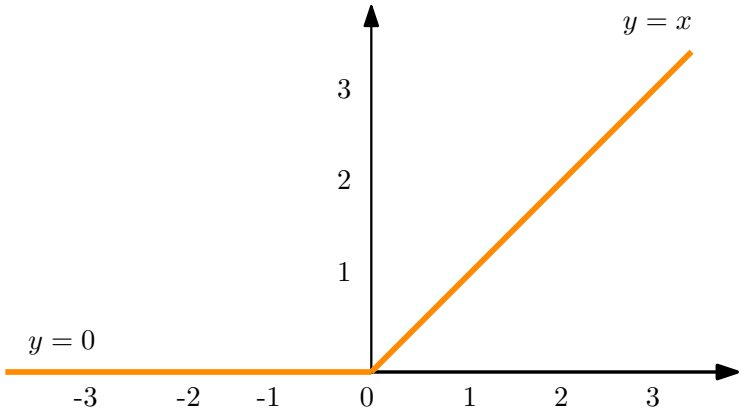


Figure 4.2: ReLU activation function

$$ReLU(z) = max(0,z) = \begin{cases} z & z > 0 \\ 0 & z \leq 0 \end{cases} \tag{4.2}$$

ii. Pooling layer: The feature map obtained from the nonlinear threshold function passes through the pooling layer to downsample the feature map/activation map and extract the most representative features. Basically, the pooling layer minimizes the number of parameters to be learnt and the computations to be performed in the network (He et al. 2020). Providing a summary statistic of nearby outputs, pooling layer majorly performs dimensionality reduction and guarantees translation invariance. The hyperparameters of pooling layer are filter size (f_s), stride window (s) and type of pooling, which are commonly either max pooling or average pooling.

iii. Fully connected (FC) layer: It connects each neuron in one layer to every neuron in other layer. In general, the DCNN process is classified into two major groups namely feature extraction and classification (Basha et al. 2020). The convolution layer along with the pooling layer is responsible for feature extraction and the fully connected layer and softmax layer handle the decision making process (San-Segundo et al. 2019). For instance, the convolution layer

receives an input dimension of $HXWXD$ and generates a feature map of dimension $H'XW'XD'$. The resultant output activation function map is flattened into a single column vector, and fed as an input to the FC layer. The output of the nodes in FC layer is determined as follows.

$$x_i = b_i + \sum_j \omega_{ji} y_j \tag{4.3}$$

where ω and b indicate the weights and biases, respectively. x and y are the output of the current and previous layers. Subsequently, the output from the FC layer is fed into softmax layer to predict the classes.

iv. Softmax layer: It converts the class scores from FC layer into a probability mass function (Xia et al. 2018). Taking a K-dimensional vector of arbitrary real-valued score z, the softmax function in (4.4) squashes it to a K-dimensional vector $f(z)$ with values ranging from 0 to 1.

$$f_j(z) = \frac{e^{z_j}}{\sum_{j=1}^{K} e^{z_k}} \tag{4.4}$$

$$H_{p'}(p) = -\sum_i p'_i log(p_i) \tag{4.5}$$

$$p_i \in (0,1) : \sum_i p_i = 1 \, \forall i \tag{4.6}$$

where p is the predicted probability distribution, p' is the true distribution and i indicates the different classes. For the multi-class classification, the softmax layer utilizes the cross entropy function given (4.5). Typically, CNN training iteration comprises forward and backward passes (Acharya et al. 2018). Forward pass results in a loss which computes the discrepancy between the truth and the current predictions. Then, backward pass through the backpropagation determines the gradient and the negative of the gradient indicates the steepest descent direction (Saba et al. 2019). For the optimization of DCNN loss function, the following SGD technique, which updates the parameters using Jacobian matrix,

is employed.

$$\omega_{t+1} = \omega_t - \eta \frac{\partial z}{\partial \omega_t} \tag{4.7}$$

where η is the learning rate. Unlike the gradient descent, which iteratively determines the gradient of the loss function L with respect to model parameters ω and updates them, the SGD computes the gradient of the loss function and updates for mini-batch of n training samples x_i and labels y_i. One of the key advantages of SGD is that even with a large training dataset, the computational time for each update does not grow exponentially (Arcos-García et al. 2018). To prevent overfitting in DCNN training, the L_2 regularization technique, which is also called ridge regression, is used. Equation 5.8 represents the regularized objective function.

$$\hat{L}(\omega, X) = L(\omega, X) + \lambda R(\omega) = L(\omega, X) + \lambda \sum_i |\omega_i^2| \tag{4.8}$$

where λ is the regularization strength, which adjusts the relative contribution of the norm penalty term $R(\omega)$ with respect to the standard loss function $L(w, X)$. Since $R(\omega)$ is convex, when it is added to the convex loss function, the augmented function is still be convex.

Figure 4.3 shows the proposed DCNN architecture based on 1D and 2D CNN for discriminating the stride variability which can help to classify the stages of PD. The raw gait data is preprocessed before applying it to the DCNN. For 1D, two frame sizes such as 300×1 and 500×1 are considered. Similarly, for 2D, the two frame sizes considered are 300×18 and 500×18 because the gait dataset consists of 16 VGRF sensor data and 2 cumulative values of left and right sensors.

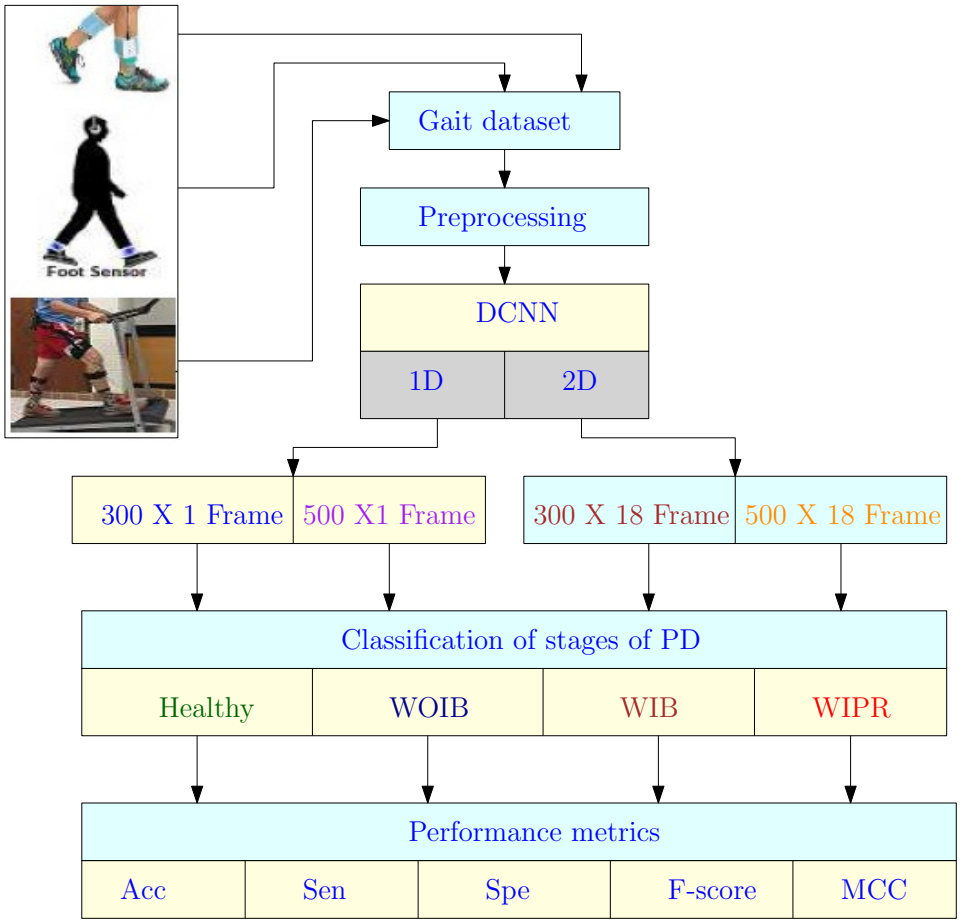


Figure 4.3: DCNN framework for PD severity prediction

4.4 Experimental Results and Discussion

Each gait sample contains 18 columns of data which include the 16 VGRF sensor signals along with 2 cumulative values from each foot. To remove the gait initiation and end-up effects, the first 20s and the last 10s data are removed, and the pre-processed gait signal is fed into DCNN. The datasets Ga(Toledo et al. 2005b) , Ju(Hausdorff et al. 2007) and Si(Yogev et al. 2005) contain sample size of about 13500, 11730, and 7700, respectively. Since each dataset has different sample size, to make them with equal length, the datasets are divided into input segments of 300 and 500 frames. In this approach, we have chosen 300×1 and 500×1 input frames for training the 1D DCNN with

the segment size 73,224 and 43,704, respectively. Table 4.1 gives the configuration parameters of 1D DCNN, which contains 4 convolutional layers, 4 average pooling layers, 2 fully connected layers and a softmax layer. Intermediate layers utilize ReLU as an activation function because ReLU results in reduced likelihood of the gradient vanishing. In fully connected layer, 50 input and 4 output layers are configured based on the four classes of PD. To measure the difference between the predictions and targets, the categorical cross-entropy was chosen as the loss function for training the DCNN. The deep learning network adjusts its weights adaptively to minimize the loss using SGD. The learning rate for SGD was initialized to 0.001 and the weights were assigned to random numbers with a uniform distribution. To avoid, the data overfitting, L_2 regularization technique is used, and the softmax layer is employed to normalize the DCNN outputs, which sum up to one. For validation, the gait dataset is divided into 80 % and 20 % for training and testing, respectively. The network is trained for up to 120 epochs, and the experiments are conducted in Keras library using TensorFlow backend.

Table 4.1: 1D DCNN configuration

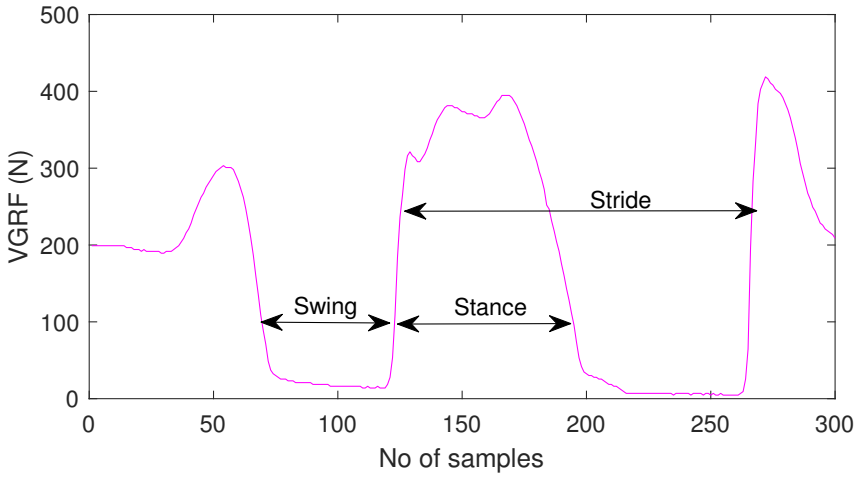
Model layers	Filter	Activation	Layer output		Parameters	Stride
			300X1	500X1		
Conv1	12	ReLU	300X12	300X12	48	(1,1)
Max-Pooling	2	-	150X12	150X12	0	(2,2)
Conv2	24	ReLU	150X24	150X24	888	(1,1)
Max-Pooling	2	-	75X24	75X24	0	(2,2)
Conv3	48	ReLU	75X48	75X48	3504	(1,1)
Max-Pooling	2	-	37X48	37X48	0	(2,2)
Conv4	96	ReLU	37X96	37X96	13920	(1,1)
Max-Pooling	2	-	18X96	18X96	0	(2,2)
Fully connected	100	ReLU	50	50	86450	-
Fully connected output	4	Softmax	4	4	36	-

To assess the significance of left and right pairs of VGRF data to classify the stages of PD, firstly we tested from L1 &R1 to L8 & R8. Table 4.2 gives the accuracy, sensitivity and specificity for each pair configuration of 1D DCNN. The performance metrics reveal that among all the pairs the L8 & R8 combination gives the highest accuracy, sensitivity and specificity. The key reason for this increased accuracy with the L8 and R8 combinations is that among the 8 sensors positioning, L8 and R8 sensors are the initial point of contact with the ground while walking. Hence, in both stride phase and stance phase, L8 and R8 generate relatively higher plantar pressure and results in better feature representation. However, when the respective left and right pairs of VGRF signals are independently used, the accuracy is relatively less, which can be attributed to minimum number of segments in the input frame. Hence, the combination of all 18 VRGF signals is utilized for training the DCNN so that the network can implicitly learn more significant biomarkers from the gait pattern.

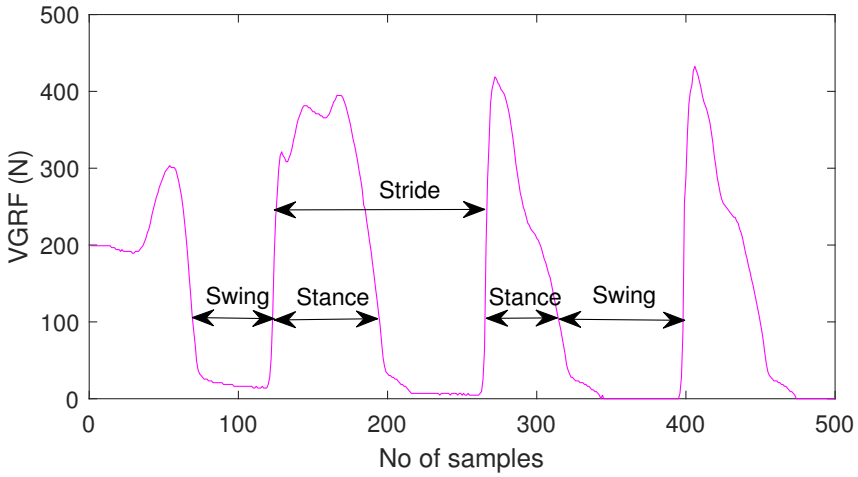
Firstly, for 1D DCNN, two input frames 300×1 (F_1) and 500×1(F_2) are tested to assess the impact of the input frame size on the classifier performance. Figure 4.4 illustrates the VGRF input data for 300 and 500 frames. It is worth noting that the VGRF plot is the signal acquired from one of 16 sensors. Figure 4.4(a) manifests that the gait cycle comprises one stride phase and swing phase. However, Figure 4.4(b) that corresponds to F_2 input frame consists of two swing phases and two stride to stride variabilities. Hence, the larger input frame contains more kinematic features that can assist the deep learning network to identify the prominent biomarkers during the training phase.

Table 4.2: Segment level impact of input signals on PD Diagnosis

VGRF Inputs	Acc(%)	Sen(%)	Spe(%)
L_1 & R_1	64.14	60.59	83.21
L_2 & R_2	63.16	55.89	82.54
L_3 & R_3	59.44	51.81	80.03
L_4 & R_4	64.15	59.41	83.30
L_5 & R_5	68.73	63.74	85.68
L_6 & R_6	65.05	58.83	83.70
L_7 & R_7	66.55	62.59	84.82
L_8 & R_8	71.63	64.71	87.36

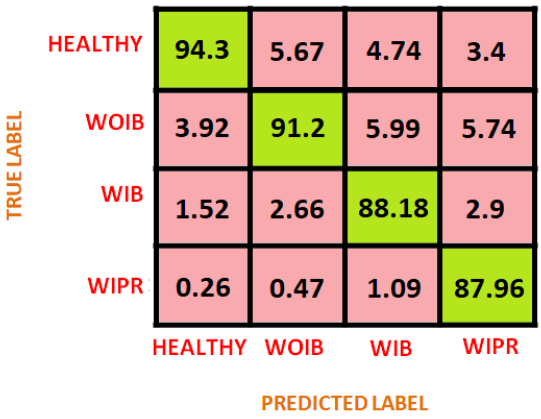


(a)

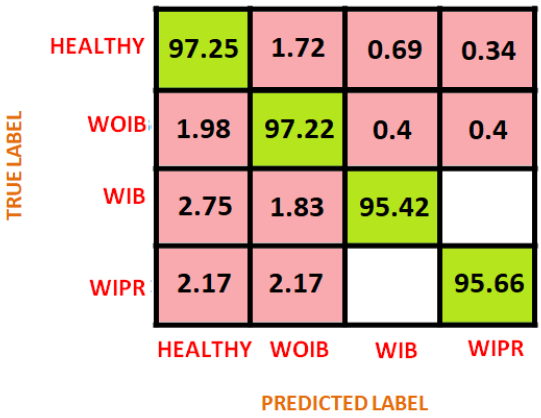


(b)

Figure 4.4: Gait time series plot - 1D CNN (a) 300X1 (b) 500X1



(a)



(b)

Figure 4.5: Confusion matrix - 1D DCNN (a) 300X1 (b) 500X1

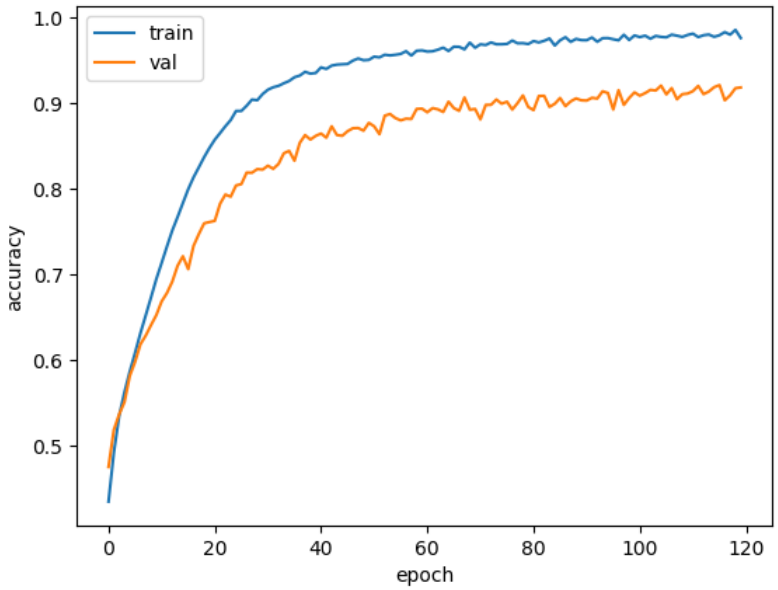
In order to quantify the performance of the classifier model, the performance metrics such as the accuracy, sensitivity and specificity is computed from the confusion matrix shown in Figure 4.5. The diagonal values of the confusion matrix indicates the maximum accuracy of each class. For instance, the 1D DCNN with F_1 input frame predicts the "healthy" class with an accuracy of 94.3 whereas the F_2 input frame predicts the same class with the accuracy of 97.25. To further quantify the classifier performance, six metrics used for objective performance comparison are given in Table 4.3. The specificity of all the classes in the case of F_2 input matrix is higher than those of similar classes in F_1 input matrix. Hence, the DCNN has demonstrated that it can effectively

exclude the patients without the disease being recognized as PD. Likewise, it can be noted that the F-score is relatively same among the classes, which corroborates that the proposed approach is relatively consistent in predicting the classes.

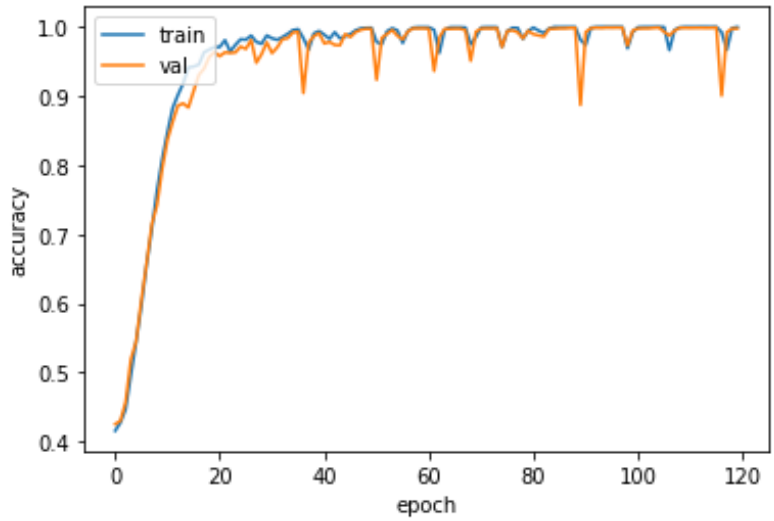
Table 4.3: Performance metrics for 1D DCNN

Performance metrics	300X1				500X1			
	Healthy	WOIB	WIB	WIPR	Healthy	WOIB	WIB	WIPR
Acc (%)	96.31	96.88	98.37	98.29	96.69	97.39	98.53	99.55
Sen (%)	98.43	95.23	91.06	84.16	97.00	95.55	95.1	95.24
Spe (%)	94.73	97.78	99.69	99.63	96.45	98.42	99.13	99.84
PPV (%)	93.31	95.91	98.19	95.51	95.41	97.12	95.41	97.56
F-score	0.95	0.96	0.94	0.89	0.96	0.96	0.95	0.96
MCC	0.93	0.93	0.94	0.89	0.93	0.94	0.95	0.96

Figure 4.6, which shows the accuracy plots for two input frames, highlights that compared to F_1, the overfitting is significantly reduced in the case of F_2 input frame. Furthermore, the maximum training accuracy of 300 and 500 frames are 0.97 and 0.98, respectively. From Figure 4.7 that illustrates the loss function for both the input frames we can note that the F_2 frame has the minimum loss function of 0.061 compared to F_1 frame that has 0.592.

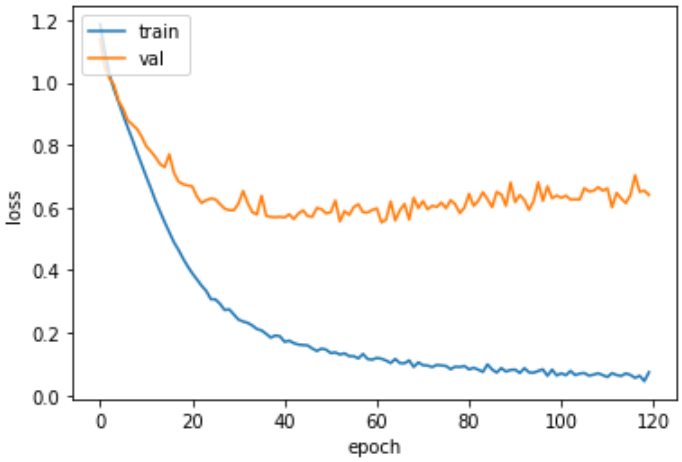


(a)

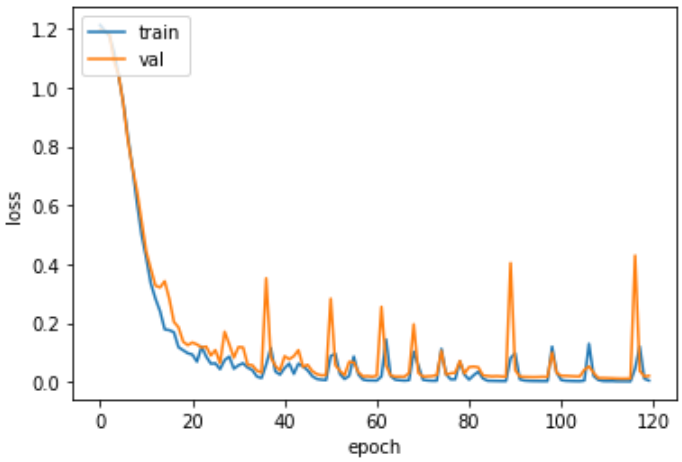


(b)

Figure 4.6: Accuracy plot - 1D DCNN (a) 300X1 (b) 500X1



(a)



(b)

Figure 4.7: Loss plot - 1D DCNN (a) 300X1 (b) 500X1

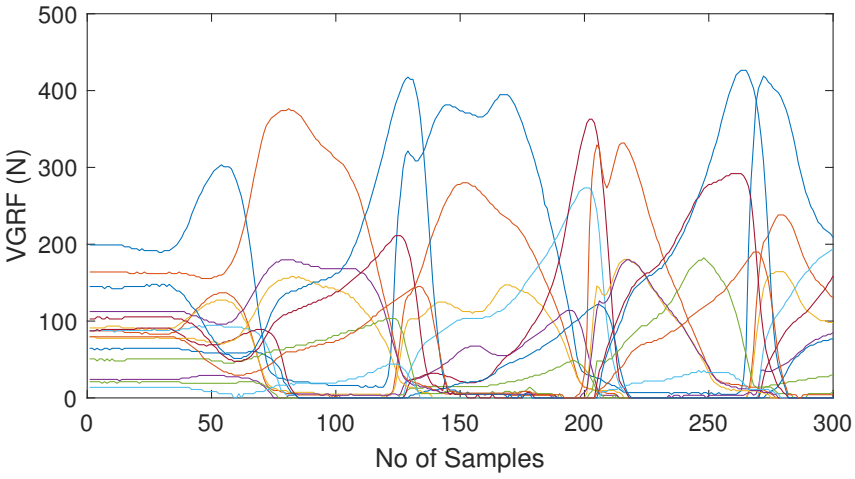
4.4.1 2D DCNN performance analysis

The fundamental difference between 1D and 2D DCNNs is that the array input in 1D is replaced with the matrix input in 2D. Hence, The gait cycle dataset is organized in a 2D matrix, and the 2D DCNN performance is tested for two frame configurations: F×18 (F:300,500). Figure 4.8 shows the 2D inputs frames of 300×18 (F_3) and 500×18 (F_4) segments. Unlike 1D DCNN, which

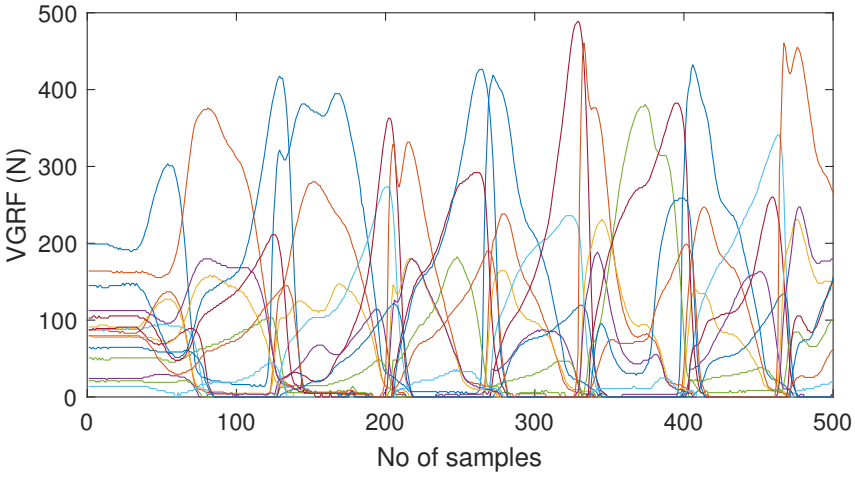
takes each foot worn sensor signal individually as a vector, the input frame for 2D DCNN reads the 16 sensor data along with two cumulative values of both left and right foot for each epoch. Table 4.4 gives the layer configuration of 2D convent, which comprises 3 convolutional layers and 3 average pooling layers. The filter size for the 3 convolutional layers are 12, 24, and 48 and for the fully connected layer the filter size is 100. The kernel functions in the convolutional layer have a stride window of (1,1). For dimensionality reduction and extracting the key features in the pooling layer, the stride window of (2,2) is used. To regularize the network during training, for each mini-batch, some of the neurons from the fully connected layers are randomly dropped, thereby enabling the classifier model to learn from a subset of input features rather than using the entire input feature set.

Table 4.4: 2D DCNN configuration

Model layers	Filter	Activation	Layer output		Parameters	Stride
			300X18	500X18		
Conv1	12	ReLU	50X12X12	50X15X12	984	(1,1)
Average-Pooling	2	-	25X6X12	25X7X12	0	(2,2)
Conv2	24	ReLU	25X6X24	25X7X24	2616	(1,1)
Average-Pooling	2	-	12X3X24	12X3X24	0	(2,2)
Conv3	48	ReLU	12X3X48	12X3X48	10416	(1,1)
Average-Pooling	2	-	6X1X48	6X1X48	0	(2,2)
Fully connected	100	ReLU	100	100	28900	-
Fully connected output	4	Softmax	4	4	404	-

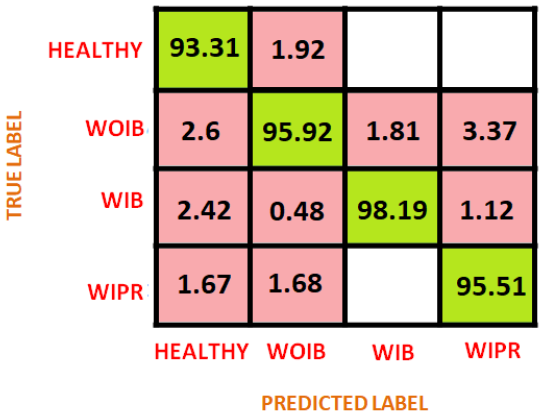


(a)

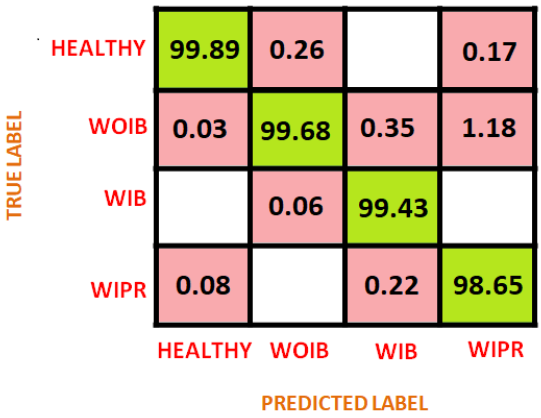


(b)

Figure 4.8: 18 VGRF input pattern - 2D DCNN (a) 300X18 (b) 500X18



(a)



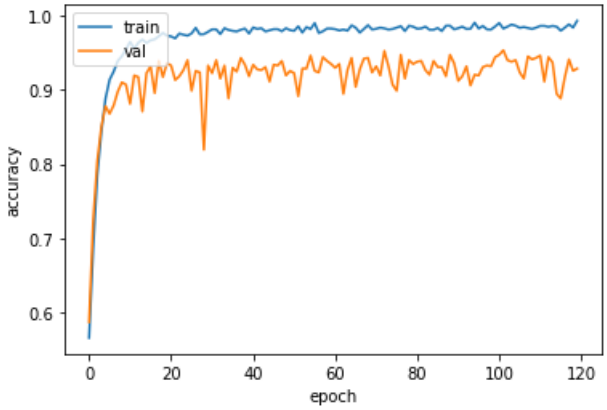
(b)

Figure 4.9: Confusion matrix - 2D DCNN (a) 300X18 (b) 500X18

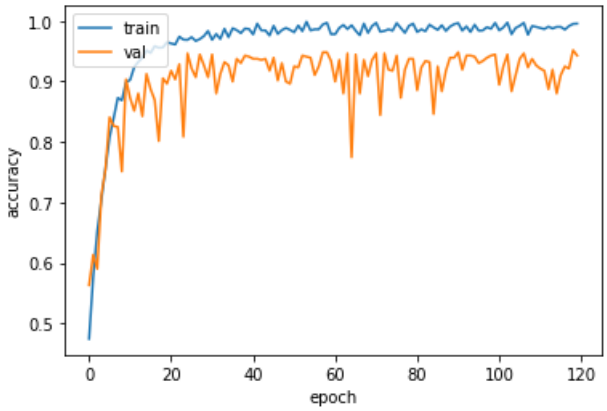
To evaluate the effective stage classification of 2D DCNN, the confusion matrix for the two input frame configurations F_3 and F_4 is shown in Figure 4.9. For all four classes, the misclassification rate in the case of F_4 frame is less than that of the F_3 frame. From Table 4.5, which gives the performance metrics of 2D DCNN, it can be noted that the 500 frames have better MCC of 0.96 compared to 0.89 in the case of 300 frames.

Table 4.5: Performance metrics for 2D DCNN

Performance metrics	300X18				500X18			
	Healthy	WOIB	WIB	WIPR	Healthy	WOIB	WIB	WIPR
Acc (%)	94.48	93.82	96.12	98.57	97.55	97.83	98.83	99.41
Sen (%)	93.12	90.97	89.23	93.44	97.24	97.23	95.41	95.65
Spe (%)	95.55	95.33	97.54	98.98	97.77	98.17	99.48	99.68
PPV (%)	94.30	91.19	88.18	87.96	96.91	96.85	97.19	95.65
F-score (%)	0.93	0.91	0.87	0.90	0.97	0.97	0.96	0.95
MCC	0.89	0.86	0.86	0.89	0.95	0.95	0.96	0.95

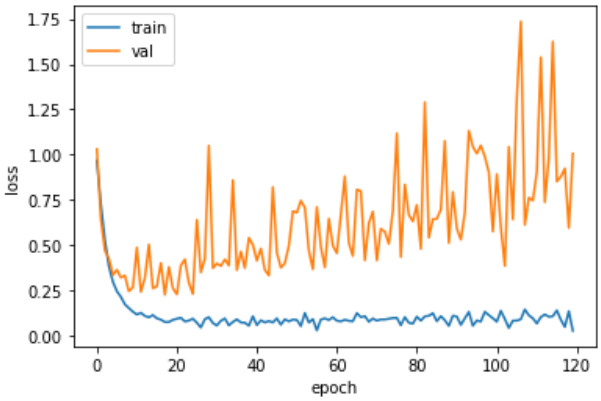


(a)

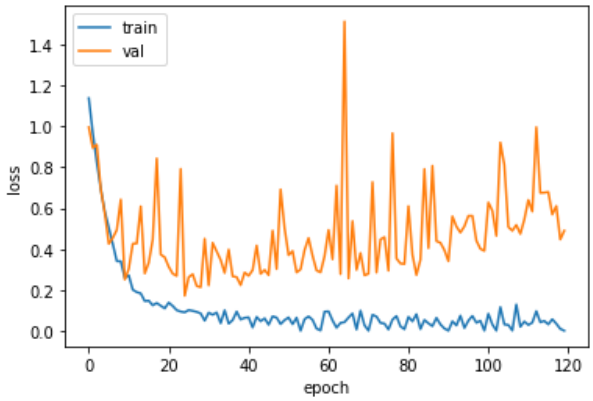


(b)

Figure 4.10: Accuracy plot - 2D DCNN (a) 300X18 (b) 500X18



(a)



(b)

Figure 4.11: Loss plot - 2D DCNN (a) 300X18 (b) 500X18

Figure 4.10 illustrates the accuracy plot of 2D DCNN for all the walking patterns corresponding to the two input frames. It can be noted that validation accuracy increases as the training accuracy increases. Moreover, even though both the test inputs converge quickly, the F_4 has better validation accuracy compared to F_3 input frame. Figure 4.11, which illustrates the loss function plots, substantiate that for the 2D CNN, the loss function value in F_3 is 0.76, whereas for F_4 is 0.33. Hence, these values substantiate that the prediction accuracy improves with the increase in the segment size. Moreover, to assess the performance of the DCNN in classifying the four stages of PD, the statistical distribution of the performance metrics across different classes is computed.

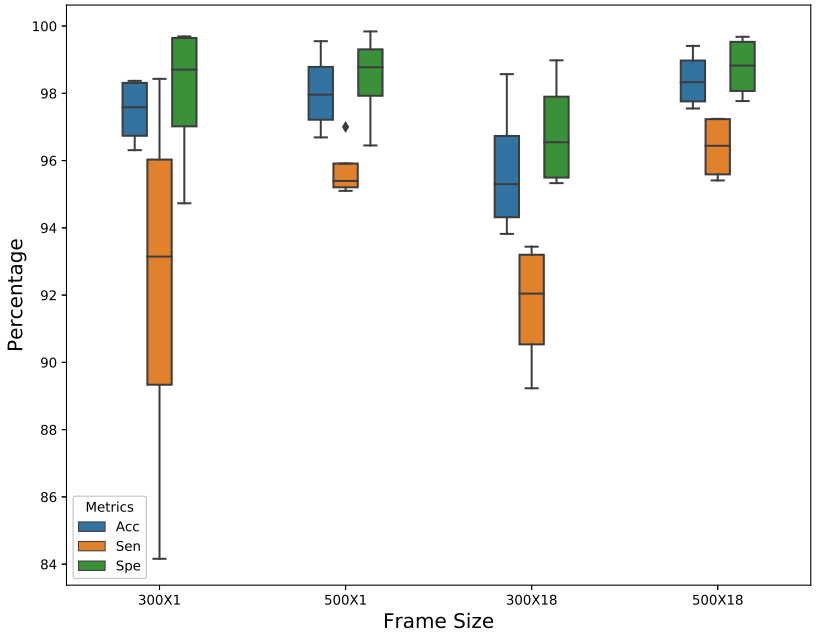


Figure 4.12: Key performance metrics of 1D and 2D DCNN for different frame sizes

Figure 4.12 illustrates the boxplot of accuracy, sensitivity and specificity distributions for all four input patterns under study. Compared to 1D DCNN with F_2 input frame that has the accuracy of 98.01 %, 2D DCNN with F_4 input frame offers better accuracy of 98.45 %. Similarly, the sensitivity of the F_2 and F_4 input frames are 95.72 % and 96.63 %, respectively. From the boxplot, we can note that the image input method with enhanced frame size offers better performance than the time series based input method with 1D DCNN. The key reason for the improved performance with 2D DCNN is that large frame size contains more discriminative biomarkers for input classification. Moreover, unlike 1D DCNN which processes each VGRF sensor data individually, the 2D DCNN, which receives the fixed window of all 16 VGRF sensor time series data, gets both spatial and temporal information. Hence, the 2D can learn more discriminative features to classify the PD gait. From these performance met-

rics, it is evident that the 2D DCNN with higher frame size offers better stage classification.

Finally, Table 4.6 gives the comparison of the proposed approach with those of the previous approaches, which have used gait pattern for PD diagnosis. We can note that the proposed approach outperforms the other approaches and is effective in classifying the four stages of PD. Furthermore, it can be noted that unlike the machine learning based approach that need the manual feature selection, the DCNN does not require the hand-crafted features and extracts the important biomarkers automatically from the gait pattern.

Table 4.6: Comparison of proposed approach with other reported methods

References	Features	Classifiers	Acc (%)	Sen (%)	Spe (%)
Ertuğrul et al. 2016	Local binary patterns	MLP	88.8	88.9	82.2
Perumal & Sankar 2016a	Spatiotemporal	LDA	87.5	72.0	81.0
Wu et al. 2017	Entropy parameters	SVM	84.48	72.41	96.5
Zeng et al. 2016	Deterministic learning	RBF NN	96.39	96.77	95.89
Khoury et al. 2019a	Spatiotemporal	RF	90.91	85.35	88.35
Proposed	CNN based features	DCNN	97.1	96.38	98.77

CHAPTER 5

LSTM FOR PD DIAGNOSIS

Deep learning has a huge potential in healthcare for uncovering the hidden patterns from large volume of clinical data to diagnose different diseases. This work presents a novel deep learning architecture based long short term memory (LSTM) network for severity rating of Parkinson's disease (PD) using gait pattern. Unlike machine learning (ML) algorithms, the LSTM network avoids the need for hand crafted features and learns the long-term temporal dependencies in the gait cycle for robust diagnosis of PD. The primary advantage of the LSTM network is that it solves the vanishing gradient problem by introducing the memory blocks in place of self-connected hidden units, thereby deciding when to learn new information. Three distinct gait datasets containing vertical ground reaction force (VGRF) recordings for different walking scenarios are used for training the LSTM network.

To avoid data overfitting, the proposed approach utilizes dropout and L2 regularization techniques. For solving the cost function, Adam, a stochastic gradient-based optimizer, is employed and the severity of PD is categorized based on unified Parkinson's disease rating scale (UPDRS) and Hoehn and Yahr (H&Y) scale. The experimental results reveal that Adam optimized LSTM network can effectively learn the gait kinematic features and offer an average accuracy of 98.6% for binary classification and 96.6% for multi-class classification, with an accuracy improvement of 3.9% in comparison with the related techniques.

5.1 Motivation

Recently, considerable results have been reported on the efficacy of ML algorithms to predict the stages of PD using gait analysis. However, the major limitation of the ML algorithms is that they require hand-crafted features for efficient classification. Generally, the hand-crafted features are not only computationally intensive because of high dimension but also not robust. Hence, the deep learning algorithms that do not require the manual hand-crafted features are currently being explored for robust and reliable diagnosis. Automatic stage classification of PD using deep neural networks (DNN) has gained considerable attention in the recent years. Nevertheless, the potential of DNN in classifying the severity of PD based on gait analysis has not been fully exploited, especially in terms of acquiring the hidden patterns from VGRF signals and diagnosing the severity of PD based on H & Y scale. Hence, in this work, a class of recurrent neural network (RNN), called LSTM, is explored to learn the long-term temporal dependencies in gait pattern for robust diagnosis of PD severity.

LSTM solves the vanishing gradient problem in vannila RNN by introducing the memory blocks in place of self-connected hidden units. These memory units enable the learning network to decide when to learn new information and when to discard old information. In this direction, the major contributions of this paper are as follows. 1. We present a gait based automatic and non-invasive PD stage diagnosis system using LSTM classifier. Firstly, the binary classification of PD is addressed by exploiting the long-term temporal gait sequence. Subsequently, the severity rating of PD is identified based on UPDRS and H & Y scale. 2. The performance of LSTM classifier is tested on the VGRF gait dataset, which includes three different walking scenarios. Moreover, to avoid data overfitting, the proposed approach utilizes dropout and L2 regularization techniques. 3. Adam optimizer, which has less memory requirement and few hyper parameters tuning, is employed to solve the cost function.

The LSTM classifier has achieved an average accuracy of 96.6% for multi-class classification and provides 3.9% improvement in overall accuracy compared to the existing techniques.

5.2 Related work

Zhao et al. 2018c proposed a two-channel model which fuses LSTM and CNN to automatically rate the PD severity from gait pattern. To capture the spatio-temporal features from gait data acquired using the force sensors, they designed the hybrid model for both binary and multi-class classifications and achieved the accuracy of 98.61%. Based on free-speech in uncontrolled background conditions, Braga et al. 2019, proposed a ML based methodology to detect the early signs of PD. To optimize the performance of three ML classifiers namely SVM, Random forest and Neural Network, they employed grid-search and learning curves. Moreover, to validate the performance of the classifiers, in addition to the dataset which the authors collected, the UCI dataset was also used.

Ashour et al. 2020 presented the LSTM based patient dependent model for freezing of gait (FOG) detection from multiple body worn sensors positioned at different places like hip, knee and ankle. Measuring the freezing and non-freezing signals using accelerometer from 9 patients, they firstly extracted the features using the DWT combined with FFT and employed the traditional ML algorithms such as SVM and ANN for classification. Secondly, they utilized LSTM based classifier model for distinguishing freezing and non-freezing cases in PD. Even though an average accuracy of 83% was achieved, the paper fails to report the stage classification of PD. In addition, the experiments were conducted on limited number of test subjects.

In another study, Zhao et al. 2018a, utilized dual channel LSTM for

diagnosing the neurodegenerative diseases (ND) namely, amyotrophic lateral sclerosis (ALS), Huntington's disease (HD) and PD. To model the gait dynamics of ND, they combined the force series with the time series and achieved a maximum classification accuracy of 97.43%. Nevertheless, in the foot switch system, only one sensor was used in each foot for acquiring gait pattern, leading to partial gait dynamics acquisition.

In another direction, Oktay & Kocer 2020 presented a convolutional LSTM technique to differentiate between Parkinsonian tremor (PT) and essential tremor (ET). Capturing the tremor using leap motion controller, which consists of a 4D camera, they attempted to extract the significant features using CNN and utilized the LSTM network for classifying the ET and PT. Nonetheless, the normalization technique and the approach followed to avoid the overfitting problem were not reported.

To overcome the limitations of hand-crafted features in ML approaches, El Maachi et al. 2020. proposed a 1D deep neural network (DNN) for diagnosing the PD using motor and non-motor symptoms. For assessing the performance of the 1D convolutional neural network (1D Convnet), they utilized the VGRF dataset from Physionet and classified the stages of PD based on UPDRS. Even though the binary classification accuracy was significantly high, the maximum stage classification accuracy was limited to 85.3 %.

For differentiating the healthy controls from PD patients, extracting the kinematic features from VGRF signals and performing frequency analysis for tremor, Abdulhay et al. 2018 implemented two ML techniques such as medium tree and Gaussian support vector machine (SVM). The fast Fourier transform (FFT) of tremor signal was applied for identifying the frequency distribution. Even though the average accuracy of 92.7 % was achieved, the paper reported only the binary classification and failed to assess the stages of PD.

On a different approach, Senturk 2020 utilized a speech signal, which is a non-motor symptom, for early diagnosis of PD using three ML algorithms namely classification and regression trees (CART), SVM and artificial neural network (ANN). For determining the most relevant features from the speech signal, recursive feature elimination (RFE) and feature importance techniques were employed. Despite the significant improvement in SVM classifier performance due to RFE method, the approach proposed in that paper required hand-crafted features for classifying the PD subjects and healthy controls.

It is important to mention that considerable results have been reported in the literature on diagnosing PD using ML approaches. However, one of the major limitations of using ML algorithms is that the classifier needs the hand-crafted features for categorizing the stages of PD. Hence, considerable attention is being paid towards deep learning methods to avoid such manual feature extraction and learn the significant features automatically using the deep neural networks. The key reason for superior classification performance by deep neural networks is their modular design in selecting the architecture, normalization schemes and activation functions. Even though deep learning has been largely applied in bioinformatics, little work has been reported for identifying the severity of PD, which requires finer investigation for multi-class classification. In this direction, the focal point of this work is to explore LSTM technique to not only diagnose the presence of PD as a (binary classification problem) but also classify the stages of PD based on H & Y scale.

5.3 Modified H&Y scale

The H&Y scale originally proposed in 1967 contains scales 1 to 5. However, the movement disorder society (MDS) task force, introduced 0.5 increments for some clinical tests based on Parkinsonian dysfunction and compromised balance. Precisely, for describing the intermediate course of the disease,

in the conventional H&Y scale, two additional stages 1.5 and 2.5 were introduced, resulting in the widely accepted modified H & Y scale. Table 5.1 gives the functionality and stages of PD based on the modified H & Y scale.

Table 5.1: PD severity rating based on the modified H & Y scale

Scale	Functionality	Stage
1	Unilateral involvement only	No functional disability(NFD)
1.5	Unilateral and axial involvement	NFD but Initial Phase
2	Bilateral	Without Impairment of Balance(WOIB)
2.5	Mild bilateral disease with recovery on pull test	With Impairment of Balance(WIB)
3	Mild to moderate bilateral disease	With Impaired Postural Reflexes (WIPR)
4	Severe disability	Still able to walk or stand unassisted
5	Confined to bed or wheelchair bound	Completely Disabled

5.4 Methodology

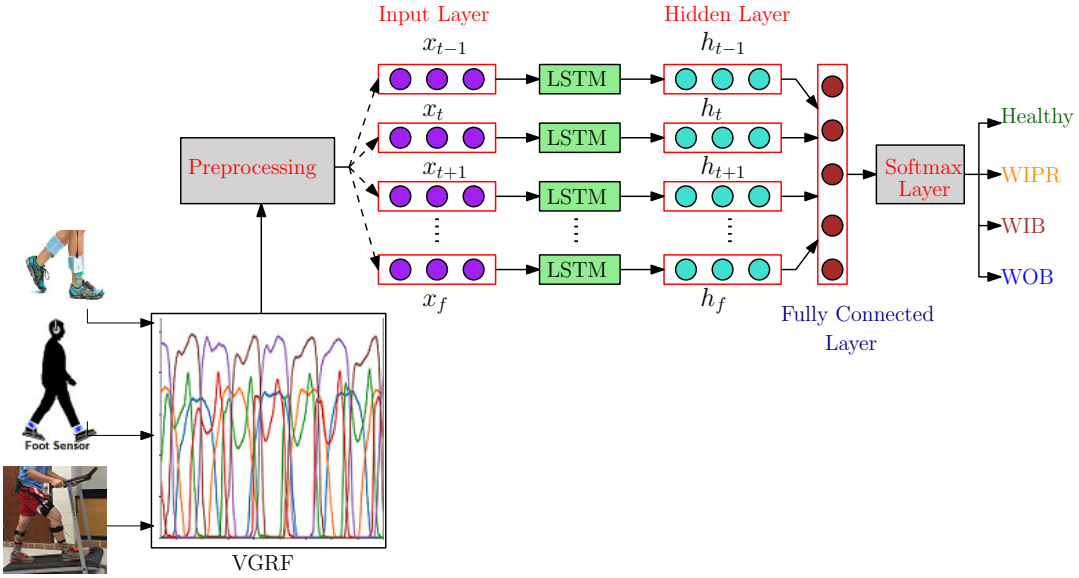


Figure 5.1: Proposed LSTM classifier for PD diagnosis

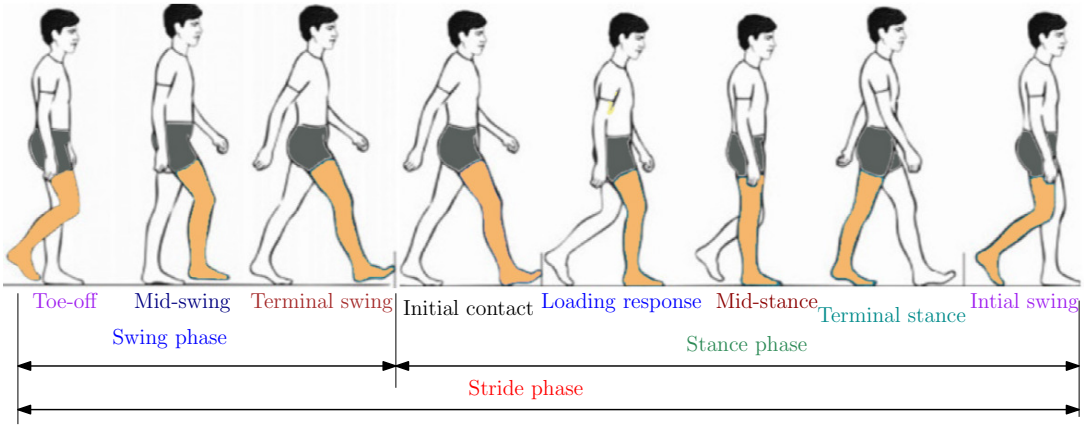


Figure 5.2: Kinematic gait parameters

Figure 5.1 shows the block diagram of the proposed LSTM based PD severity prediction approach. The pre-processed VGRF signal, which is divided into segments, is given to the LSTM input layer. Gait cycle, as illustrated in Figure 5.2, consists of significant spatiotemporal features to differentiate between the healthy and PD patients (Moltedo et al. 2018). The two primary phases of gait cycle namely swing and stance phases constitute 60% and 40% of total gait cycle. Capturing these spatiotemporal features through VGRF sensors help to identify the significant biomarkers for effective classifications. In the next section,we briefly present the LSTM classifier along with the regularization technique.

5.4.1 LSTM Overview

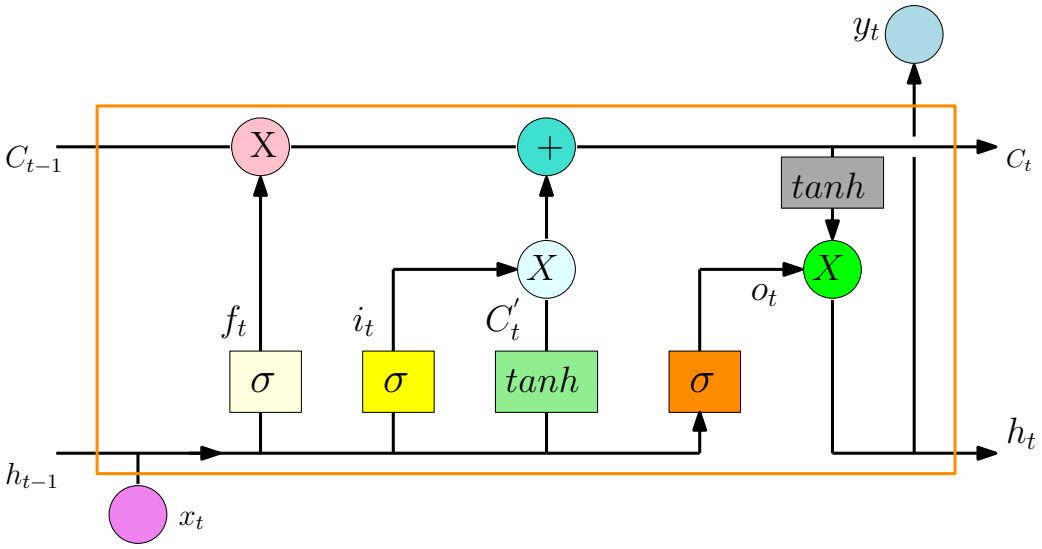


Figure 5.3: LSTM block diagram with gating mechanism

To overcome the major limitations of vannila RNN, such as exploding and vanishing gradient problems, Hochreiter and Schmidhuber put forward a LSTM, which has the capability to learn the long-term dependencies among sequential data sets (Yildirim 2018; Zheng et al. 2020). To enhance the learning process and mitigate vanishing gradient problem, unlike RNN, LSTM consists of memory cells in place of each traditional node in the hidden layer, thereby the information can be stored and accessed over a long period of time. Typically, LSTM networks are widely used in time series prediction such as machine translation, air pollution forecasting, weather forecasting and speech recognition (Hosny et al. 2020).

Figure 5.3 shows the architecture of LSTM, which consists of memory blocks and memory cells in the hidden layer. The memory block contains three gate units namely, input gate, forget gate and output gate. To avoid the negative effects due to unrelated inputs, the multiplicative gate units are utilized (Wang & Chen 2019). The forget gate determines the quantity of information to be retained or forgotten from the memory cell. The activation function of the

forget gate is computed as follows.

$$f_t = \sigma(W_{hf} h_{t-1} + W_{xf} x_t + b_f) \tag{5.1}$$

where b is the bias vector,h_{t-1} is the previous block output, x_t is the input sequence;w_{hf} and w_{xf} represent the weight matrices for the output vector of precedent cell and input vector of current cell, respectively, in the forget gate. σ is the sigmoid function given by

$$\sigma(x) = (1 + e^{-x})^{-1} \tag{5.2}$$

The input layer determines what information can be stored in the memory cell from the current input vector as follows.

$$i_t = \sigma(W_{hi} h_{t-1} + W_{xi} x_t + b_i) \tag{5.3}$$

where W_{hi} and W_{xi} represent the weight matrices for the output vector of precedent cell and input vector of current cell, respectively, in the input layer. b_i depicts the bias in the input gate layer. Finally, the output gate layer, which decides which output can be passed in the current time step, is defined as

$$o_t = \sigma(W_{ho} h_{t-1} + W_{xo} x_t + b_o) \tag{5.4}$$

where W_{ho} and W_{xo} indicate the weight matrices for the output vector of precedent cell and input vector of current cell, respectively, in the output gate layer. Hence, integrating the forget gate and input, the current cell state is computed by

$$C_t = f_t \odot C_{t-1} + i_t \odot C_t^{'} \tag{5.5}$$

where C_t is the cell state at time step t and $\odot$ is the Hadamard product that indicates element-wise multiplication of vector. $f_t \odot C_{t-1}$ and $i_t \odot C_t^{'}$ decide the information to be inherited from the precedent cell state and current input,

respectively (Arslan & Sekertekin 2019). Hence, $C^{'}$ is determined based on the tanh activation function.

$$C_t^{'} = \tanh(W_{hC}h_{t-1} + W_{xC}x_t + b_C) \tag{5.6}$$

The hidden state is determined by multiplying the output gate with the current cell state.

$$o_t = h_t \odot \tanh(C_t) \tag{5.7}$$

5.4.2 Reducing overfitting

DNN with a large number of learnable parameters are susceptible to overfitting problems, particularly when trained using a relatively small dataset. Consequently, the DNN model lacks the generalization ability to classify the new test sample (He et al. 2020). The possible reasons for overfitting in deep learning algorithms are less training data, noisy input data and high dimensional input space. Hence, regularization techniques are used to reduce the overfitting issue in the DNN model. In this application, we have used the L2 regularization and dropout techniques to address the overfitting issue.

5.4.3 L2 Regularization

L2 regularization is a process of changing the loss function to penalize the weights which are large, consequently reducing the generalization error. Basically, L2 regularization, also called ridge regression, adds a function of weights to the loss function to avoid the weights from growing too large because large weight makes the DNN model more sensitive to noise in the input

data. The cost function used in L2 regularization is as follows.

$$\hat{L}(\omega, X) = L(\omega, X) + \lambda R(\omega) = L(\omega, X) + \lambda \sum_i |\omega_i^2| \tag{5.8}$$

where $L(w, X)$ can be a cross entropy loss function. λ indicates the regularization strength that has fine influence over the total loss function $\hat{L}(\omega, X)$. It is important to note that as $R(\omega)$ is convex, when it is attached to the convex loss function, the resultant function will still be convex.

5.4.4 Dropout

Unlike L2 regularization method which modifies the cost function, dropout alters the connections, thereby dropping neurons with p probability. Dropout follows the concept of co-adaptation avoidance among the hidden nodes of DNN by randomly dropping out chosen hidden nodes (Ha et al. 2019). Hence, the fundamental idea behind dropout is that while training a DNN, some parts of the model is dropped out stochastically in each iteration, thereby training exponential number of models and combining them into a single model during validation. The forward propagation with dropout is represented as

$$z_j = \sum_i W_{ij} d_i x_i + b_i \tag{5.9}$$

where d_i is a vector of independent Bernoulli random variables with probability p. If d_i is zero, the input node x_i will be dropped out.

5.4.5 Softmax Layer

The final layer of the DNN classifier is the softmax layer, which de-

termines the probability of the input that belongs to a particular class.

$$h_\theta(x) = \begin{pmatrix} p(y=1|x;\theta) \\ p(y=2|x;\theta) \\ \vdots \\ p(y=C|x;\theta) \end{pmatrix} = \frac{1}{\sum_{i=1}^{C} e^{(\theta_i^T x)}} \begin{pmatrix} e^{(\theta_1 x)} \\ e^{(\theta_2 x)} \\ \vdots \\ e^{(\theta_C x)} \end{pmatrix} \tag{5.10}$$

The following cross entropy loss function, also called negative log likelihood, which measures the dissimilarity between predicted distribution q and true label distribution p, is used to update the w and b during training.

$$H(p,q) = -\sum_x p(x) log\, q(x) \tag{5.11}$$

Integrating the aforementioned steps, the pseudo code of the proposed LSTM network classifier for PD diagnosis is given in Algorithm 5.1.

Algorithm 5.1 PD prediction using LSTM

Input:d-dimensional gait pattern x, trained LSTM model
Output: Predicted stages of PD
Initialization: No of gait signals =18,No of classes (C)= 4, d=12000, M=500
procedure (PD-LSTM(x, C,LSTM))
    Select gait segment length (L)
    Partition the VGRF signals into M segments (300×1, 500×1)
    while $i \leq M$ **do**
        $z_t = g(W_z x_t + R_z y_{t-1} + b_z)$ ▷ LSTM Input
        $i_t = \sigma(W_{hi} h_{t-1} + W_{xi} x_t + b_i)$ ▷ Input gate
        $f_t = \sigma(W_{hf} h_{t-1} + W_{xf} x_t + b_f)$ ▷ Forget gate
        $C_t = f_t \odot C_{t-1} + i_t \odot C_t^{'}$ ▷ Cell
        $o_t = \sigma(W_{ho} h_{t-1} + W_{xo} x_t + b_o)$ ▷ Output gate
        $o_t = h_t \odot \tanh(C_t)$ ▷ LSTM Output
        $s_t = h_t(y_t)$ ▷ FC Layer
    end while
    $E = AP(s_t, s_{t-1}, ...s_{t-M})$; ▷ Average pooling
    Determine $P_C = \{P_1, ..., P_c\} \leftarrow$ Softmax(E)
    Compute IHP $\leftarrow$ Support ($max(P_C)$) ▷ Index of highest probability
    $\hat{y} = IHP$; ▷ Predicted classes
end procedure

5.4.6 Adam Optimization

Adam optimization algorithm, which is a combination of root mean square propogation (RMSProp) and gradient descent with momentum algorithms, is one of the widely used algorithms to train DNNs (Fei et al. 2020). Adam, which is not an acronym but the named coined by the authors from "adaptive moment estimation", can be substituted with the standard stochastic gradient descent algorithm for updating the DNN weights (Chang et al. 2019). It is a first order stochastic optimization technique that maintains the exponential moving averages of gradient and squared gradient. This method has less memory requirement and requires only few hyper-parameters tuning. Moreover, the parameter updates guarantee bounded norm. Hence, Adam optimization has been widely used in several DNN applications (Kingma & Ba 2014). Algorithm 5.2

presents the Adam optimization for solving the stochastic cost function. The fundamental idea behind the Adam optimization is as follows. Firstly, the decaying averages of past gradient and its square are calculated.

$$m_t = \beta_1.m_{t-1} + (1-\beta_1).g_t \tag{5.12}$$

$$m_t = \beta_1.m_{t-1} + (1-\beta_1).g_t \tag{5.13}$$

where m_t and v_t indicate the 1^{st} and 2^{nd} moments of the gradients, respectively. Since, the procedure begins with m_t and v_t being initialized as vectors of 0's, resulting in moment estimates biased towards zero, to correct the initialization bias, the first and second moments are updated as follows.

$$\hat{m}_t = m_t/(1-\beta_1^t) \tag{5.14}$$

$$\hat{v}_t = v_t/(1-\beta_2^t) \tag{5.15}$$

Then, to update the parameters, the following Adam update rule is used.

$$\theta_t = \theta_{t-1} - \alpha.\hat{m}_t/(\sqrt{\hat{v}_t}+\varepsilon) \tag{5.16}$$

Algorithm 5.2 Adam optimization

Require: stepsize α, exponential decay rates for the moment estimates $\beta_1, \beta_2 \in [0,1]$, stochastic objective function $f(\theta)$ with parameters θ
Initialize: Parameter vector θ, timetep $t = 1, ..., T$
Initialize: 1^{st} moment and 2^{nd} moment vectors: $m_0 = 0$, $v_0 = 0$
while until θ_t converges **do**
    $i = i + 1$
    Find gradients $g_t \leftarrow \Delta_\theta f_t(\theta_{t-1})$
    Update 1^{st} moment estimate $m_t \leftarrow \beta_1.m_{t-1} + (1-\beta_1).g_t$
    Update 2^{nd} raw moment estimate $v_t \leftarrow \beta_2.v_{t-1} + (1-\beta_2).g_t^2$
    Calculate bias corrected first moment estimate $\hat{m}_t \leftarrow m_t/(1-\beta_1^t)$
    Calculate bias corrected second moment estimate $\hat{v}_t \leftarrow v_t/(1-\beta_2^t)$
    Update parameters $\theta_t \leftarrow \theta_{t-1} - \alpha.\hat{m}_t/(\sqrt{\hat{v}_t}+\varepsilon)$
    Return θ_t
end while

5.5 Results and Discussion

Each gait pattern from the Physionet consists of 18 columns of data, including 16 VGRF sensor signals and 2 cumulative values of left and right feet sensors. Firstly, to mitigate the gait initiation and end-up effects, 10s data from the start and 20s data at the end are removed. The three datasets Ga, JU and Si consist of data samples of 13500, 11730 and 7700 respectively. Table 5.2 gives the parameter configuration of LSTM network, which contains 4 LSTM layers, 4 dropout layers, followed by a fully connected layer and a softmax layer. For multi-class stage prediction, a cross entropy loss function is used. The LSTM network is trained for 50 epochs and the experiments are implemented using Keras library with TensorFlow backend. 80% of the dataset is used for training the LSTM network and 20% of the dataset is used for testing.

The LSTM network consists of 50 hidden units and utilizes "ReLU" nonlinear activation function for efficient gradient propagation. To avoid the overfitting, L_2 regularization is used along with the dropout layers. To regularize the dataset and improve the model performance the dropout rate is set to 0.2. To minimize the loss function, the following hyper-parameters of Adam optimization algorithm are configured: the learning rate $\alpha = 0.001$, the decay rates $\beta_1 = 0.9,\ \beta_2 = 0.99$, constant $\varepsilon = 10^{-8}$ and the threshold $\delta = 10^{-5}$. Finally, the softmax layer predicts the probability of the input sequence which belongs to any of the four classes namely healthy, WOIB, WIB, and WIPR.

Table 5.2: LSTM Configuration

Layers	Type of layer	Properties	Parameters
Layer 1	Sequence input	1 dimension	-
Layer 2	LSTM	50 units, (N:300,500)input shape	13800
Layer 3	Dropout	0.2 dropout rate	0
Layer 4	LSTM	50 units	20200
Layer 5	Dropout	0.2 dropout rate	0
Layer 6	LSTM	50 units	20200
Layer 7	Dropout	0.2 dropout rate	0
Layer 8	LSTM	50 units	20200
Layer 9	Dropout	0.2 dropout rate	0
Layer 10	Fully connected	50 layer output	0
Layer 11	Dense	ReLU activation, 25 units	1125025
Layer 12	Dense	Softmax-4 classes	104

5.5.1 Binary classification

Firstly, the performance of the LSTM network to predict the presence of PD is formulated as a binary classification problem. To visualize the prediction rate of the LSTM classifier, the confusion matrix (CM), which is also called an error matrix, is illustrated in Figure 5.4. The diagonal entries of CM represent the correct prediction of the classes and the non-diagonal values indicate the misclassification rate. We can read from Figure 5.4, which illustrates the CM of two class classification problem, that the LSTM network classifies the healthy and PD patients with an accuracy of 98.6%. The misclassification in the case of PD and healthy are 1 % and 0.4%, respectively. Figure 5.5a, which shows the accuracy plot, highlights that the gap between the validation and training is significantly less, thereby avoiding the overfitting issue. Moreover, the loss function plot, illustrated in Figure 5.5b, gives the quantitative loss measure at

each epoch to identify the direction in which the LSTM is trained towards classifying the input pattern. The minimum loss function value of around 0.1 is an indication that the model converges considerably quickly and both training and validation performances are quite similar. The convergence behaviour of the network underscores that the mean squared error (MSE) is a good match for the prediction.

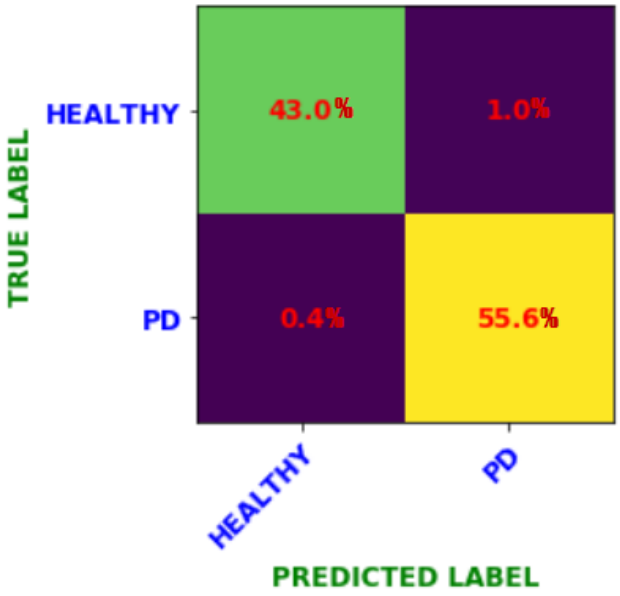
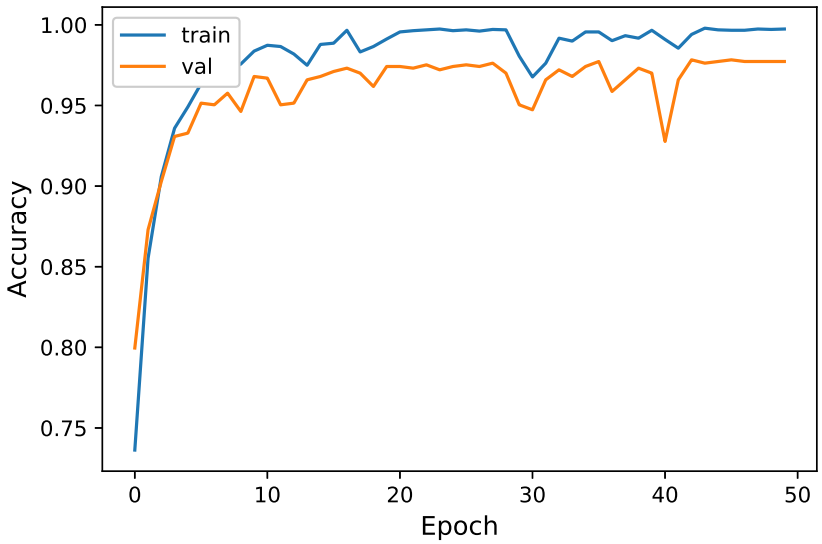
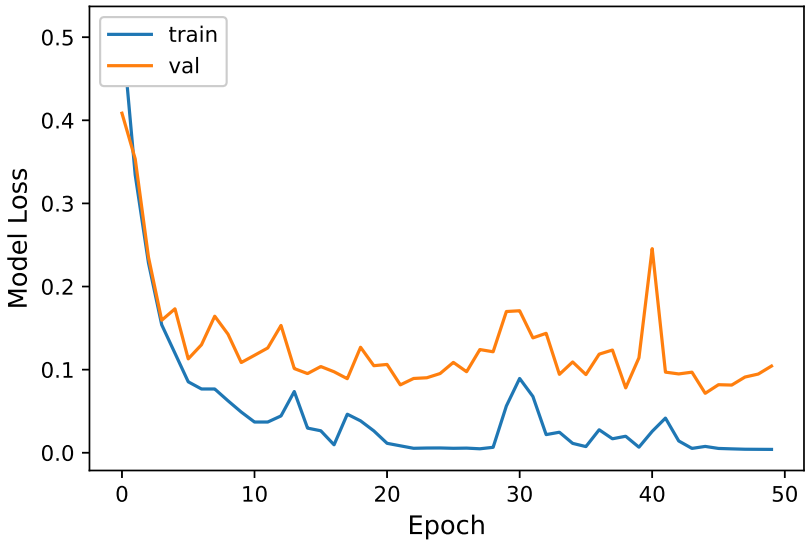


Figure 5.4: Confusion matrix for binary classification



(a)



(b)

Figure 5.5: Accuracy and loss function plots for binary classification

5.5.2 Multi-class classification

Even though the LSTM model provides high accuracy in binary classification, the progression of the PD needs to be evaluated so as to decide the severity of the PD for further medication/therapy. Hence, we divide the gait time

series data into segments of 300 $\times 1(F_1)$ and 500 $\times 1(F_2)$ and provide them as an input to LSTM network. Figure 5.6 shows the VGRF time series data for 300$\times$1 and 500 $\times$1 frames. Compared to F1 frame the occurrence of swing and stride are more in F2, which can assist the deep learning model to train the network effectively.

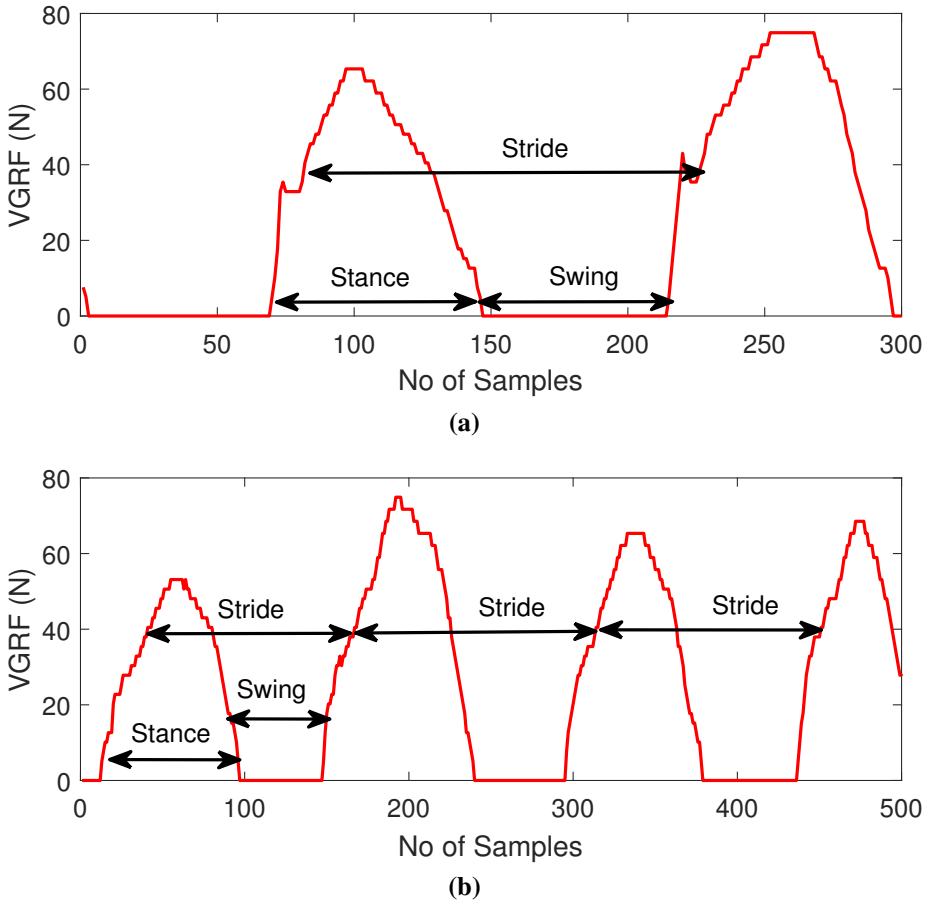
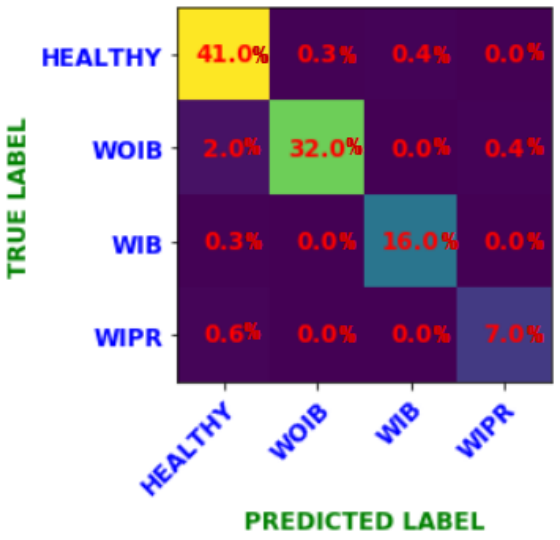
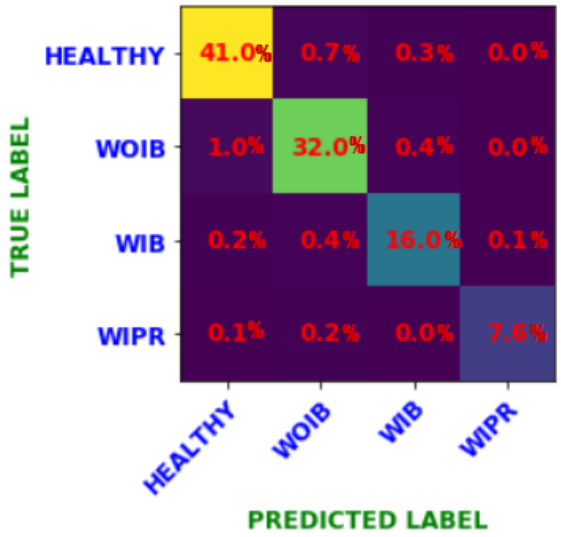


Figure 5.6: VGRF input frame-(a) 300$\times$1 (b)500$\times$1



(a)



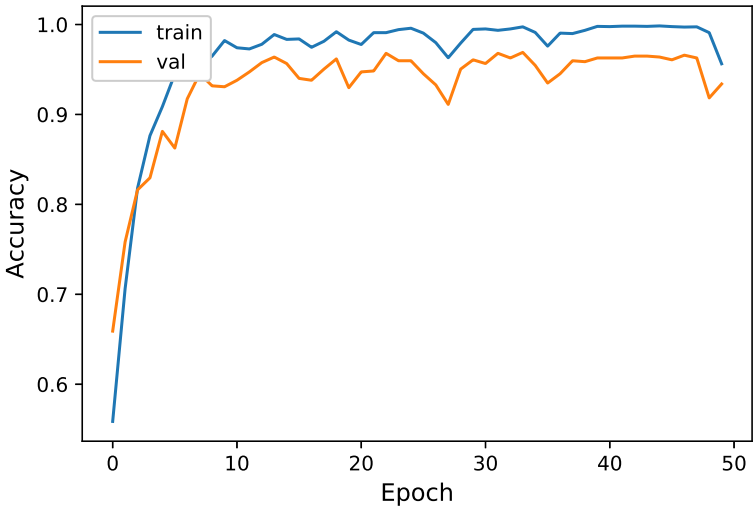
(b)

Figure 5.7: Confusion matrix (a) 300X1 (b) 500X1

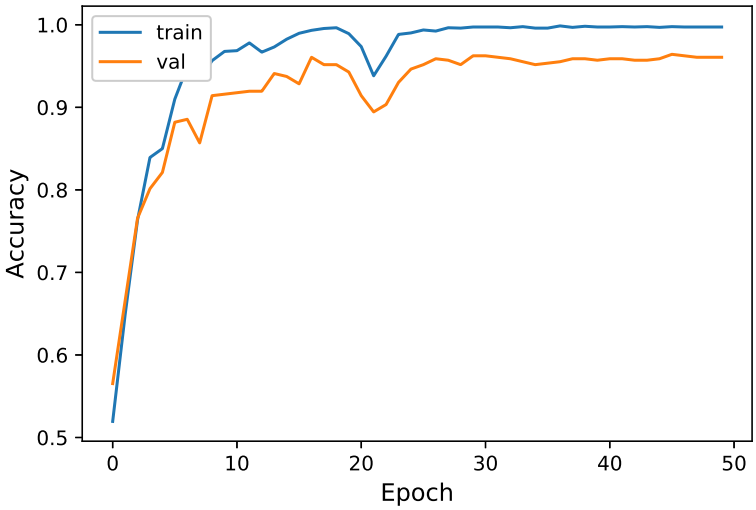
Figure 5.7 shows the CM for multi-class classification problem. We can note that for the input segment F_1, the classifier predicts the healthy and WIPR classes with an accuracy of 96.00% and 97.01%, respectively. The misclassification rate in the case of WIB and WOIB are 0.7% and 2.7%. In the case of input segment F_2, we can notice a marginal increase in the classifier performance with the highest accuracy being 96.60%. The misclassification rate is

also slightly reduced in the case of F_2 frame because the network can exploit the repetitiveness of the gait kinematic features such as swing time, stride time and stance.

Table 5.3 gives the six performance metrics of LSTM classifier for both F_1 and F_2 frames. Comparing the average specificity of $F1$ and F_2 frames, we can note that the latter has significant improvement of 1.2. Interestingly, the F score of WIPR for F_2 frame is significantly higher than that of F_1 frame because when the subject is in WIPR stage, the movement of the patient is generally slower compared to other stages. Hence, the larger frame segment captures more number of kinematic features and yields better F-score. Moreover, to highlight the multi-class classification performance during training and validation phases, the accuracy and loss function plots are illustrated in Figure 5.8 and 5.9, respectively. The combination of L_2 regularization technique and the dropout layer significantly reduces the overfitting, which is evident from the accuracy plot. For graphically depicting the dispersion of accuracy, sensitivity and specificity of the LSTM classifiers for two frames, the box plot is illustrated in Figure 5.10. The box plot highlights the significant improvement in the classifier performance for the increased frame size to effectively classify the stages of PD. Table 5.4 gives the comparison of performance of the proposed approach with those of the other state-of-the-art methods that have used VGRF dataset. It is evident that the proposed approach offers better accuracy and specificity.

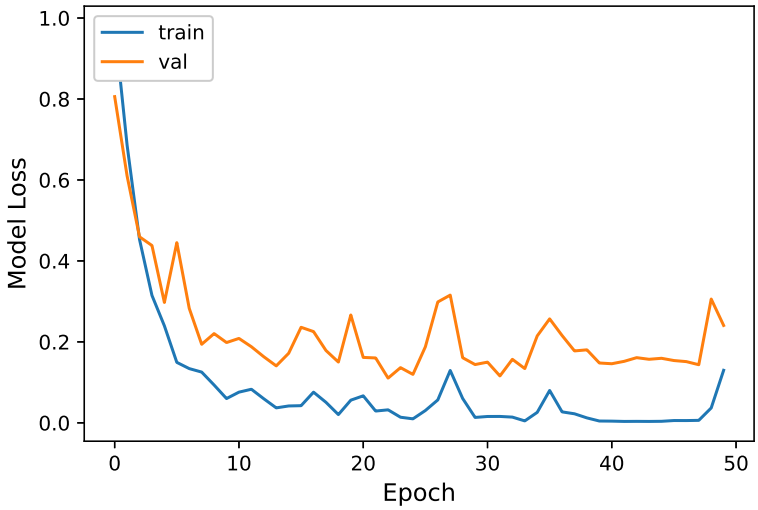


(a)

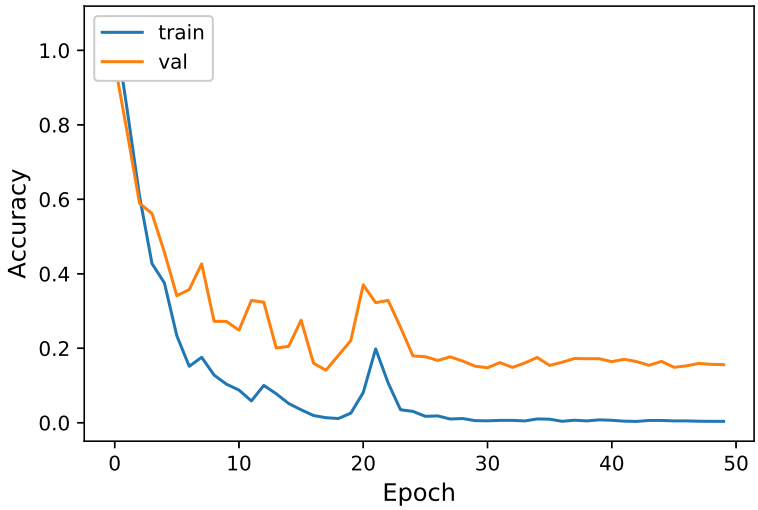


(b)

Figure 5.8: Accuracy plot (a) 300X1 (b) 500X1



(a)



(b)

Figure 5.9: Loss plot (a) 300X1 (b) 500X1

Table 5.3: LSTM network classification performance metrics

Parameters	300X1				500X1			
	Healthy	WOIB	WIB	WIPR	Healthy	WOIB	WIB	WIPR
Acc (%)	96.00	95.98	96.04	96.01	96.80	96.70	98.00	98.70
Sen (%)	96.02	95.40	94.62	90.03	96.20	95.20	94.60	96.20
Spe (%)	96.40	96.10	97.80	98.98	97.3	97.60	98.70	98.70
PPV (%)	95.00	95.70	92.10	91.20	96.20	95.90	93.10	96.21
F-score (%)	95.50	95.20	92.12	91.00	96.21	95.60	94.10	96.40
MCC	0.94	0.92	0.93	0.96	0.97	0.96	0.96	0.97

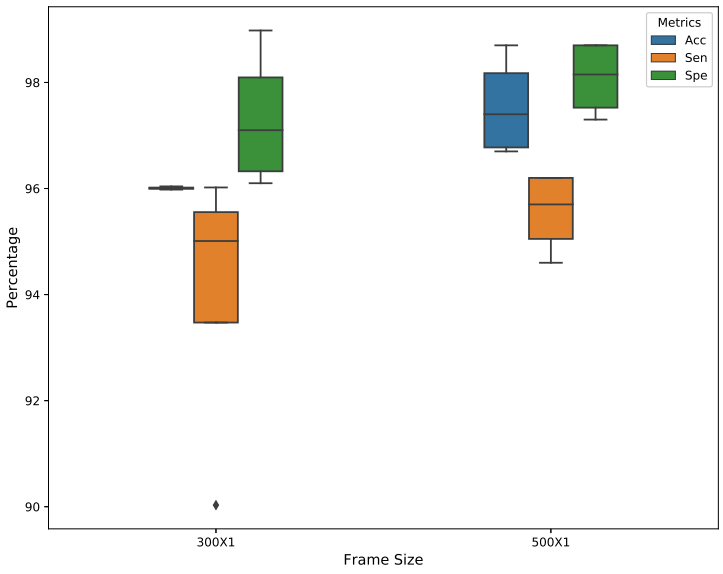


Figure 5.10: Box plot for accuracy, sensitivity and specificity

Table 5.4: Comparison of proposed approach with other reported approaches

References	Classifiers	Acc (%)	Sen (%)	Spe (%)
Paragliola & Coronato 2018	CNN	92.00	85.10	88.21
Ashour et al. 2020	LSTM	83.89	88.72	84.89
Oktay & Kocer 2020	LSTM	90.00	95.24	86.96
Khoury et al. 2019a	RF	90.91	85.35	88.35
Abdulhay et al. 2018	SVM	92.7	96.13	92.21
Veeraragavan et al. 2020	ANN	87.10	67.90	90.50
Proposed	Adam-LSTM (Multi-class)	96.60	96.20	98.08
	Adam-LSTM (Binary)	98.60	98.23	99.10

5.6 Summary and future scope

This work has presented a LSTM network for early and non-invasive diagnosis of PD based on gait pattern. LSTM, which is largely suited for sequential data analysis, has the capability to extract the long-term dependencies from time series data. Hence, we have utilized the VGRF gait time series dataset from Physionet for differentiating the healthy control and PD patients. Since three input datasets have different sample size, to normalize the input pattern to LSTM network, the input frame has been divided into two different segments and the experiments are conducted using Tensorflow backend. The dropout and L2 regularization techniques used in this approach has significantly reduced the data overfitting, which is evident from the accuracy and loss function plots. Particularly, the PD diagnosis is formulated as a binary class and multi-class classification problems and an Adam optimized LSTM is used to identify the severity of PD based on UPDRS and H&Y scale. The LSTM classifier performance

is assessed using six key performance metrics and the results substantiate that the proposed approach performs better than the state-of-the-art techniques that used gait based PD diagnosis. The major limitation of the proposed approach is the computational time, which is around five hours for 50 epochs in the case of multi-class classification. Hence, as a future work, we aim to explore the optimal selection of input frame and number of epochs for improved accuracy and reduced time complexity. Moreover, as this study is based on only gait analysis, it is indeed limited to people who are able to walk and fall under the H &Y scale 1-3. Hence, as a further investigation, both motor and non-motor symptoms can be assessed for PD prognosis of people who are either wheel chair bounded or need assistance in walking. Moreover, from the deep learning algorithms stand point, even though LSTM offers very good performance in the gait time series classification, the two major limitations of LSTM are the long training time and high sensitivity to random weight initialization. Hence, currently, significant research attention is also paid to developing a light weight CNN-LSTM model that can reduce the training time and is robust against weight initialization effects.

CHAPTER 6

CONCLUSIONS

To assist the neurologists in PD diagnosis and minimize the misclassification rate in quantifying the stages of PD, a gait classification framework using supervised machine learning algorithms is put forward. Assessing the gait abnormality based on the VRGF dataset using the statistical and kinematic analyses, we have extracted optimal diagnostic biomarkers from spatiotemporal domain and implemented a 10 fold cross validation technique to avoid data overfitting. Moreover, the performance of the classifiers is assessed using the confusion matrix and the ROC curves. Unlike most of the previous machine learning based approaches which are binary classification problem that detects only the presence of PD, the proposed approach can perform a multi-class classification and quantify the stages of PD. Experimental validation substantiates that compared to several state-of-the-art methods, the proposed approach, which uses optimal spatiotemporal domain features, can offer better prediction of stages of PD based on the H & Y scale. Moreover, to avoid the hand-crafted feature selection in the machine learning algorithms, we have also explored two deep learning models such as CNN and LSTM for stage classification of PD.

Experimental results substantiate that the proposed gait classification framework provides an average stage classification accuracy of 98.4 % and 96.6 % for CNN and LSTM, respectively. Even though the proposed gait classification frameworks provide better classification accuracy compared to other methods reported on the same dataset, it is also important to mention the limitations of the proposed study. In the current study, we considered only the significant motor symptoms of PD for classification. However, to improve the prediction

rate, the non-motor symptoms can also be considered. Moreover, in addition to the gait pattern, the tremor dataset can also be assessed for enhancing the accuracy of the stage classification.

REFERENCES

1. Dataset accessed on 10 October 2019, 'Gait in Parkinson's Disease: Available online:', https://www.physionet.org/content/gaitpdb/1.0.0/.

2. Abdulhay, E., Arunkumar, N., Narasimhan, K., Vellaiappan, E., & Venkatraman, V. 2018, 'Gait and tremor investigation using machine learning techniques for the diagnosis of parkinson disease', Future Generation Computer Systems, vol. 83, pp. 366–373.

3. Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., Adeli, H., & Subha, D. P. 2018, 'Automated EEG-based screening of depression using deep convolutional neural network', Computer methods and programs in biomedicine, vol. 161, pp. 103–113.

4. Alam, M. N., Garg, A., Munia, T. T. K., Fazel-Rezai, R., & Tavakolian, K. 2017, 'Vertical ground reaction force marker for parkinson's disease', PloS one, vol. 12, no. 5, pp. e0175951.

5. Alharthi, A. S. & Ozanyan, K. B. 2019, 'Deep learning for ground reaction force data analysis: Application to wide-area floor sensing', 2019 IEEE 28th International Symposium on Industrial Electronics (ISIE), pp 1401–1406,IEEE.

6. Alharthi, A. S., Yunas, S. U., & Ozanyan, K. B. 2019, 'Deep learning for monitoring of human gait: a review', IEEE Sensors Journal, vol. 19, no. 21, pp. 9575–9591.

7. Amoroso, N., La Rocca, M., Monaco, A., Bellotti, R., & Tangaro, S. 2018, 'Complex networks reveal early MRI markers of parkinson's disease', Medical image analysis, vol. 48, pp. 12–24.

8. Arcos-García, Á., Alvarez-Garcia, J. A., & Soria-Morillo, L. M. 2018, 'Deep neural network for traffic sign recognition systems: An analysis of spatial transformers and stochastic optimisation methods', Neural Networks, vol. 99, pp. 158–165.

9. Arifoglu, D. & Bouchachia, A. 2019, 'Detection of abnormal behaviour for dementia sufferers using convolutional neural networks', Artificial intelligence in medicine, vol. 94, pp. 88–95.

10. Arslan, N. & Sekertekin, A. 2019, 'Application of long short-term memory neural network model for the reconstruction of modis land surface temperature images', Journal of Atmospheric and Solar-Terrestrial Physics, vol. 194, pp. 105100.

11. Ashour, A. S., El-Attar, A., Dey, N., Abd El-Kader, H., & Abd El-Naby, M. M. 2020, 'Long short term memory based patient-dependent model for fog detection in parkinson's disease', Pattern Recognition Letters, vol. 131, pp. 23–29.

12. Aşuroğlu, T., Açıcı, K., Erdaş, Ç. B., Toprak, M. K., Erdem, H., & Oğul, H. 2018, 'Parkinson's disease monitoring from gait analysis via foot-worn sensors', Biocybernetics and Biomedical Engineering, vol. 38, no. 3, pp. 760–772.

13. Balaji, E., Brindha, D., & Balakrishnan, R. 2020, 'Supervised machine learning based gait classification system for early detection and stage classification of parkinson's disease', Applied Soft Computing, pp. 106494.

14. Basha, S. S., Dubey, S. R., Pulabaigari, V., & Mukherjee, S. 2020, 'Impact of fully connected layers on performance of convolutional neural networks for image classification', Neurocomputing, vol. 378, pp. 112–119.

15. Battineni, G., Chintalapudi, N., & Amenta, F. 2019, 'Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM)', Informatics in Medicine Unlocked, vol. 16, pp. 100200.

16. Braga, D., Madureira, A. M., Coelho, L., & Ajith, R. 2019, 'Automatic detection of parkinson's disease based on acoustic analysis of speech', Engineering Applications of Artificial Intelligence, vol. 77, pp. 148–158.

17. Camps, J., Sama, A., Martin, M., Rodriguez-Martin, D., Perez-Lopez, C., Arostegui, J. M. M., Cabestany, J., Catala, A., Alcaine, S., Mestre, B., et al. 2018, 'Deep learning for freezing of gait detection in parkinson's disease patients in their homes using a waist-worn inertial measurement unit', Knowledge-Based Systems, vol. 139, pp. 119–131.

18. Cao, C., Liu, F., Tan, H., Song, D., Shu, W., Li, W., Zhou, Y., Bo, X., & Xie, Z. 2018, 'Deep learning and its applications in biomedicine', Genomics, proteomics & bioinformatics, vol. 16, no. 1, pp. 17–32.

19. Chang, Z., Zhang, Y., & Chen, W. 2019, 'Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform', Energy, vol. 187, pp. 115804.

20. Chen, P.-H., Wang, R.-L., Liou, D.-J., & Shaw, J.-S. 2013, 'Gait disorders in parkinson's disease: assessment and management', International Journal of Gerontology, vol. 7, no. 4, pp. 189–193.

21. Creaby, M. W. & Cole, M. H. 2018, 'Gait characteristics and falls in parkinson's disease: A systematic review and meta-analysis', Parkinsonism & related disorders, vol. 57, pp. 1–8.

22. Das, R. 2010, 'A comparison of multiple classification methods for diagnosis of parkinson disease', Expert Systems with Applications, vol. 37, no. 2, pp. 1568–1572.

23. El Maachi, I., Bilodeau, G.-A., & Bouachir, W. 2020, 'Deep 1d-convnet for accurate parkinson disease detection and severity prediction from gait', Expert Systems with Applications, vol. 143, pp. 113075.

24. Elkurdi, A., Soufian, M., Nefti-Meziani, S., et al. 2018, 'Gait speeds classifications by supervised modulation based machine-learning using kinect camera', Medical Research and Innovations, vol. 2, no. 4, pp. 1–6.

25. Ertuğrul, Ö. F., Kaya, Y., Tekin, R., & Almalı, M. N. 2016, 'Detection of parkinson's disease by shifted one dimensional local binary patterns from gait', Expert Systems with Applications, vol. 56, pp. 156–163.

26. Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L., & Muller, P.-A. 2019, 'Deep learning for time series classification: a review', Data Mining and Knowledge Discovery, vol. 33, no. 4, pp. 917–963.

27. Fei, Z., Wu, Z., Xiao, Y., Ma, J., & He, W. 2020, 'A new short-arc fitting method with high precision using adam optimization algorithm', Optik, pp. 164788.

28. Figueiredo, J., Santos, C. P., & Moreno, J. C. 2018, 'Automatic recognition of gait patterns in human motor disorders using machine learning: A review', Medical engineering & physics, vol. 53, pp. 1–12.

29. Ghaddar, B. & Naoum-Sawaya, J. 2018, 'High dimensional data classification and feature selection using support vector machines', European Journal of Operational Research, vol. 265, no. 3, pp. 993–1004.

30. Goetz, C. G., Poewe, W., Rascol, O., Sampaio, C., Stebbins, G. T., Counsell, C., Giladi, N., Holloway, R. G., Moore, C. G., Wenning, G. K., et al. 2004, 'Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations the Movement Disorder Society Task Force on rating scales for Parkinson's disease', Movement disorders, vol. 19, no. 9, pp. 1020–1028.

31. Gupta, U., Bansal, H., & Joshi, D. 2020, 'An improved sex-specific and age-dependent classification model for parkinson's diagnosis using handwriting measurement', Computer Methods and Programs in Biomedicine, vol. 189, pp. 105305.

32. Ha, C., Tran, V.-D., Van, L. N., & Than, K. 2019, 'Eliminating overfitting of probabilistic topic models on short and noisy text: The role of dropout', International Journal of Approximate Reasoning, vol. 112, pp. 85–104.

33. Hausdorff, J. M., Lowenthal, J., Herman, T., Gruendlinger, L., Peretz, C., & Giladi, N. 2007, 'Rhythmic auditory stimulation modulates gait variability in parkinson's disease', European Journal of Neuroscience, vol. 26, no. 8, pp. 2369–2375.

34. He, K., Zhang, X., Ren, S., & Sun, J. 2016, 'Deep residual learning for image recognition', Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778.

35. He, T., Liu, Y., Yu, Y., Zhao, Q., & Hu, Z. 2020, 'Application of deep convolutional neural network on feature extraction and detection of wood defects', Measurement, vol. 152, pp. 107357.

36. Hosny, M., Zhu, M., Gao, W., & Fu, Y. 2020, 'A novel deep LSTM network for artifacts detection in microelectrode recordings', Biocybernetics and Biomedical Engineering.

37. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. 2017, 'Densely connected convolutional networks', Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4700–4708.

38. Huang, Y. S. & Suen, C. Y. 1993, 'The behavior-knowledge space method for combination of multiple classifiers', IEEE computer society conference on computer vision and pattern recognition, pp 347–347,Institute of Electrical Engineers Inc (IEEE).

39. Joshi, D., Khajuria, A., & Joshi, P. 2017, 'An automatic non-invasive method for parkinson's disease classification', Computer methods and programs in biomedicine, vol. 145, pp. 135–145.

40. Khoury, N., Attal, F., Amirat, Y., Oukhellou, L., & Mohammed, S. 2019a, 'Data-driven based approach to aid parkinson's disease diagnosis', Sensors, vol. 19, no. 2, pp. 1–27.

41. Khoury, N., Attal, F., Amirat, Y., Oukhellou, L., & Mohammed, S. 2019b, 'Data-driven based approach to aid Parkinson's disease diagnosis', Sensors, vol. 19, no. 2, pp. 242.

42. Kim, H. B., Lee, W. W., Kim, A., Lee, H. J., Park, H. Y., Jeon, H. S., Kim, S. K., Jeon, B., & Park, K. S. 2018, 'Wrist sensor-based tremor severity quantification in parkinson's disease using convolutional neural network', Computers in biology and medicine, vol. 95, pp. 140–146.

43. Kingma, D. P. & Ba, J. 2014, 'Adam: A method for stochastic optimization', arXiv preprint arXiv:1412.6980.

44. Kittler, J. 1998, 'Combining classifiers: A theoretical framework', Pattern analysis and Applications, vol. 1, no. 1, pp. 18–27.

45. Korkmaz, S., Zararsiz, G., & Goksuluk, D. 2014, 'Drug/nondrug classification using support vector machines with various feature selection strategies', Computer methods and programs in biomedicine, vol. 117, no. 2, pp. 51–60.

46. Lee, S.-H. & Lim, J. S. 2012, 'Parkinson's disease classification using gait characteristics and wavelet-based feature extraction', Expert Systems with Applications, vol. 39, no. 8, pp. 7338 – 7344.

47. Li, Y., Huang, C., Ding, L., Li, Z., Pan, Y., & Gao, X. 2019, 'Deep learning in bioinformatics: Introduction, application, and perspective in the big data era', Methods, vol. 166, pp. 4–21.

48. Luts, J., Ojeda, F., Van de Plas, R., De Moor, B., Van Huffel, S., & Suykens, J. A. 2010, 'A tutorial on support vector machine-based methods for classification problems in chemometrics', Analytica Chimica Acta, vol. 665, no. 2, pp. 129–145.

49. Martínez, M., Villagra, F., Castellote, J. M., & Pastor, M. A. 2018, 'Kinematic and kinetic patterns related to free-walking in parkinson's disease', Sensors, vol. 18, no. 12, pp. 4224.

50. Mirelman, A., Bonato, P., Camicioli, R., Ellis, T. D., Giladi, N., Hamilton, J. L., Hass, C. J., Hausdorff, J. M., Pelosin, E., & Almeida, Q. J. 2019, 'Gait impairments in Parkinson's disease', The Lancet Neurology.

51. Moltedo, M., Baček, T., Verstraten, T., Rodriguez-Guerrero, C., Vanderborght, B., & Lefeber, D. 2018, 'Powered ankle-foot orthoses: the effects of the assistance on healthy and impaired users while walking', Journal of neuroengineering and rehabilitation, vol. 15, no. 1, pp. 86.

52. Mu, Y., Liu, X., & Wang, L. 2018, 'A pearson's correlation coefficient based decision tree and its parallel implementation', Information Sciences, vol. 435, pp. 40–58.

53. Mughal, M. O. & Kim, S. 2018, 'Signal classification and jamming detection in wide-band radios using Naive Bayes classifier', IEEE Communications Letters, vol. 22, no. 7, pp. 1398–1401.

54. Oktay, A. B. & Kocer, A. 2020, 'Differential diagnosis of parkinson and essential tremor with convolutional LSTM networks', Biomedical Signal Processing and Control, vol. 56, pp. 101683.

55. Oung, Q. W., Muthusamy, H., Basah, S. N., Lee, H., & Vijean, V. 2018, 'Empirical wavelet transform based features for classification of parkinson's disease severity', Journal of medical systems, vol. 42, no. 29, pp. 1–17.

56. Paragliola, G. & Coronato, A. 2018, 'Gait anomaly detection of subjects with Parkinson's disease using a deep time series-based approach', IEEE Access, vol. 6, pp. 73280–73292.

57. Pereira, C. R., Pereira, D. R., Rosa, G. H., Albuquerque, V. H., Weber, S. A., Hook, C., & Papa, J. P. 2018, 'Handwritten dynamics assessment through con-

volutional neural networks: An application to parkinson's disease identification', Artificial intelligence in medicine, vol. 87, pp. 67–77.

58. Perumal, S. V. & Sankar, R. 2016a, 'Gait and tremor assessment for patients with parkinson's disease using wearable sensors', Ict Express, vol. 2, no. 4, pp. 168–174.

59. Perumal, S. V. & Sankar, R. 2016b, 'Gait and tremor assessment for patients with Parkinson's disease using wearable sensors', ICT Express, vol. 2, no. 4, pp. 168 – 174 Special Issue on Emerging Technologies for Medical Diagnostics.

60. Prabhu, P., Karunakar, A., Anitha, H., & Pradhan, N. 2018, 'Classification of gait signals into different neurodegenerative diseases using statistical analysis and recurrence quantification analysis', Pattern Recognition Letters.

61. Prashanth, R. & Roy, S. D. 2018, 'Novel and improved stage estimation in Parkinson's disease using clinical scales and machine learning', Neurocomputing, vol. 305, pp. 78–103.

62. Prashanth, R., Roy, S. D., Mandal, P. K., & Ghosh, S. 2016, 'High-accuracy detection of early Parkinson's disease through multimodal features and machine learning', International journal of medical informatics, vol. 90, pp. 13–21.

63. Ricciardi, C., Amboni, M., De Santis, C., Improta, G., Volpe, G., Iuppariello, L., Ricciardelli, G., D'Addio, G., Vitale, C., & Barone, P. 2019, 'Using gait analysis' parameters to classify parkinsonism: a data mining approach', Computer methods and programs in biomedicine, pp. 105033.

64. Saba, L., Biswas, M., Kuppili, V., Godia, E. C., Suri, H. S., Edla, D. R., Omerzu, T., Laird, J. R., Khanna, N. N., Mavrogeni, S., et al. 2019, 'The present and future of deep learning in radiology', European journal of radiology, vol. 114, pp. 14–24.

65. Sakar, B. E., Isenkul, M. E., Sakar, C. O., Sertbas, A., Gurgen, F., Delil, S., Apaydin, H., & Kursun, O. 2013, 'Collection and analysis of a parkinson speech

dataset with multiple types of sound recordings', IEEE Journal of Biomedical and Health Informatics, vol. 17, no. 4, pp. 828–834.

66. San-Segundo, R., Gil-Martín, M., D'Haro-Enríquez, L. F., & Pardo, J. M. 2019, 'Classification of epileptic EEG recordings using signal transforms and convolutional neural networks', Computers in biology and medicine, vol. 109, pp. 148–158.

67. Schwab, P. & Karlen, W. 2019, 'Phonemd: Learning to diagnose parkinson's disease from smartphone data', Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp 1118–1125.

68. Senturk, Z. K. 2020, 'Early diagnosis of parkinson's disease using machine learning algorithms', Medical Hypotheses, vol. 138, pp. 109603.

69. Shrivastava, P., Shukla, A., Vepakomma, P., Bhansali, N., & Verma, K. 2017, 'A survey of nature-inspired algorithms for feature selection to identify parkinson's disease', Computer methods and programs in biomedicine, vol. 139, pp. 171–179.

70. Simonyan, K. & Zisserman, A. 2014, 'Very deep convolutional networks for large-scale image recognition', arXiv preprint arXiv:1409.1556.

71. Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. 2015, 'Going deeper with convolutions', 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 1–9.

72. Toledo, S., Giladi, N., Peretz, C., Herman, T., Gruendlinger, L., & Hausdorff, J. M. 2005a, 'Effect of gait speed on gait rhythmicity in Parkinson's disease: variability of stride time and swing time respond differently', Journal of neuroengineering and rehabilitation, vol. 2, no. 1, pp. 23.

73. Toledo, S., Giladi, N., Peretz, C., Herman, T., Gruendlinger, L., & Hausdorff, J. M. 2005b, 'Treadmill walking as an external pacemaker to improve gait

rhythm and stability in parkinson's disease', Movement disorders: official journal of the Movement Disorder Society, vol. 20, no. 9, pp. 1109–1114.

74. Tropea, T. F. & Chen-Plotkin, A. S. 2018, 'Unlocking the mystery of biomarkers: A brief introduction, challenges and opportunities in Parkinson disease', Parkinsonism & related disorders, vol. 46, pp. S15–S18.

75. Turner, A. & Hayes, S. 2019, 'The classification of minor gait alterations using wearable sensors and deep learning', IEEE Transactions on Biomedical Engineering, vol. 66, no. 11, pp. 3136–3145.

76. Vásquez-Correa, J. C., Arias-Vergara, T., Orozco-Arroyave, J., Eskofier, B., Klucken, J., & Nöth, E. 2018, 'Multimodal assessment of parkinson's disease: a deep learning approach', IEEE journal of biomedical and health informatics, vol. 23, no. 4, pp. 1618–1630.

77. Veeraragavan, S., Gopalai, A. A., Gouwanda, D., & Ahmad, S. A. 2020, 'Parkinson's disease diagnosis and severity assessment using ground reaction forces and neural networks', Frontiers in Physiology, vol. 11.

78. Wahid, F., Begg, R. K., Hass, C. J., Halgamuge, S., & Ackland, D. C. 2015, 'Classification of parkinson's disease gait using spatial-temporal gait features', IEEE journal of biomedical and health informatics, vol. 19, no. 6, pp. 1794–1802.

79. Wang, S. & Chen, S. 2019, 'Application of the long short-term memory networks for well-testing data interpretation in tight reservoirs', Journal of Petroleum Science and Engineering, vol. 183, pp. 106391.

80. Wu, Y., Chen, P., Luo, X., Wu, M., Liao, L., Yang, S., & Rangayyan, R. M. 2017, 'Measuring signal fluctuations in gait rhythm time series of patients with parkinson's disease using entropy parameters', Biomedical Signal Processing and Control, vol. 31, pp. 265–271.

81. Xia, Y., Zhang, J., Ye, Q., Cheng, N., Lu, Y., & Zhang, D. 2018, 'Evaluation of deep convolutional neural networks for detection of freezing of gait in parkinson's disease patients', Biomedical Signal Processing and Control, vol. 46, pp. 221–230.

82. Yildirim, Ö. 2018, 'A novel wavelet sequence based on deep bidirectional LSTM network model for ecg signal classification', Computers in biology and medicine, vol. 96, pp. 189–202.

83. Yogev, G., Giladi, N., Peretz, C., Springer, S., Simon, E. S., & Hausdorff, J. M. 2005, 'Dual tasking, gait rhythmicity, and parkinson's disease: which aspects of gait are attention demanding?', European journal of neuroscience, vol. 22, no. 5, pp. 1248–1256.

84. Zeng, W., Liu, F., Wang, Q., Wang, Y., Ma, L., & Zhang, Y. 2016, 'Parkinson's disease classification using gait analysis via deterministic learning', Neuroscience letters, vol. 633, pp. 268–278.

85. Zeng, W. & Wang, C. 2015, 'Classification of neurodegenerative diseases using gait dynamics via deterministic learning', Information Sciences, vol. 317, pp. 246 – 258.

86. Zhang, Y. & Ma, Y. 2019, 'Application of supervised machine learning algorithms in the classification of sagittal gait patterns of cerebral palsy children with spastic diplegia', Computers in biology and medicine, vol. 106, pp. 33–39.

87. Zhao, A., Qi, L., Dong, J., & Yu, H. 2018a, 'Dual channel LSTM based multi-feature extraction in gait for diagnosis of neurodegenerative diseases', Knowledge-Based Systems, vol. 145, pp. 91–97.

88. Zhao, A., Qi, L., Li, J., Dong, J., & Yu, H. 2018b, 'A hybrid spatio-temporal model for detection and severity rating of parkinson's disease from gait data', Neurocomputing, vol. 315, pp. 1–8.

89. Zhao, A., Qi, L., Li, J., Dong, J., & Yu, H. 2018c, 'A hybrid spatio-temporal model for detection and severity rating of parkinson's disease from gait data', Neurocomputing, vol. 315, pp. 1–8.

90. Zheng, X., Chen, W., You, Y., Jiang, Y., Li, M., & Zhang, T. 2020, 'Ensemble deep learning for automated visual classification using EEG signals', Pattern Recognition, vol. 102, pp. 107147.

91. Zhou, T., Ruan, S., & Canu, S. 2019, 'A review: Deep learning for medical image segmentation using multi-modality fusion', Array, pp. 100004.

LIST OF PUBLICATIONS

Journal

1. Balaji E, Brindha D, Balakrishnan R (2020), 'Supervised machine learning based gait classification system for early detection and stage classification of Parkinson's disease', Applied Soft Computing, vol. 94, pp.1-15. (Impact factor: 6.725)
2. Balaji E, Brindha D, Vinodh Kumar E, Umesh K (2021), 'Data-driven gait analysis for diagnosis and severity rating of Parkinson's disease', Medical Engineering & Physics, vol. 91, pp.54-64. (Impact factor: 2.24)
3. Balaji E, Brindha D, Vinodh Kumar E, and Vikrama R (2021), 'Automatic and non-invasive Parkinson's disease diagnosis and severity rating using LSTM network', Applied Soft Computing, vol. 108, pp.1-14. (Impact factor: 6.725)