Human Gait Analysis using Spatiotemporal Data Obtained from Gait Videos

Zur Erlangung des akademischen Grades eines

DOKTORS DER INGENIEURWISSENSCHAFTEN (Dr.-Ing.)

von der KIT-Fakultät für

Elektrotechnik und Informationstechnik

des Karlsruher Instituts für Technologie (KIT)

genehmigte

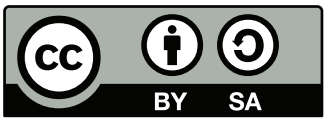
DISSERTATION

von

Anees Qumar Abbasi, M.Phil.

geb. in Bagh AJK, Pakistan

Tag der mündlichen Prüfung:	19. Januar 2023
Referent:	Prof. Dr. rer. nat. Werner Nahm
Korreferent:	Prof. Dr. rer. nat. Wilhelm Stork



Abstract

With the development of deep learning techniques, deep neural network (NN)-based methods have become the standard for vision tasks such as tracking human motion and pose estimation, recognizing human activity, and recognizing faces. Deep learning techniques have improved the design, implementation, and deployment of complex and diverse applications, which are now being used in a wide variety of fields, including biomedical engineering. The application of computer vision techniques to medical image and video analysis has resulted in remarkable results in recognizing events. The inbuilt capability of convolutional neural network (CNN) in extracting features from complex medical images, coupled with long short term memory network (LSTM)'s ability to maintain the temporal information among events, has created many new horizons for medical research. Gait is one of the critical physiological areas that can reflect many disorders associated with aging and neurodegeneration. A comprehensive and accurate gait analysis can provide insights into human physiological conditions. Existing gait analysis techniques require a dedicated environment, complex medical equipment, and trained staff to collect the gait data. In the case of wearable systems, such a system can alter cognitive abilities and cause discomfort for patients.

Additionally, it has been reported that patients usually try to perform better during the laboratory gait test, which may not represent their actual gait. Despite technological advances, we continue to encounter limitations when it comes to measuring human walking in clinical and laboratory settings. Using current gait analysis techniques remains expensive and time-consuming and makes it difficult to access specialized equipment and expertise.

Therefore, it is imperative to have such methods that could give long-term data about the patient's health without any dual cognitive tasks or discomfort while using wearable sensors. Hence, this thesis proposes a simple, easy-to-deploy, inexpensive method for gait data collection. This method is based on recording walking videos using a smartphone camera in a home environment under free conditions. Deep NN then processes those videos to extract the gait events after classifying the positions of the feet. The detected events are then further used to quantify various spatiotemporal parameters of the gait, which are important for any gait analysis system.

In this thesis, walking videos were used that were captured by a low-resolution smartphone camera outside the laboratory environment. Many deep learning-based NNs were implemented to detect the basic gait events like the foot position in respect of the ground from those videos. In the first study, the architecture of AlexNet was used to train the model from scratch using walking videos and publicly available datasets. An overall accuracy of 74% was achieved with this model. However, the LSTM layer was included with the same architecture in the next step. The inbuilt capability of LSTM regarding the temporal information resulted in improved prediction of the labels for foot position, and an accuracy of 91% was achieved. However, there is hardship in predicting true labels at the last stage of the swing and the stance phase of each foot.

In the next step, transfer learning is used to get the benefit of already trained deep NNs by using pre-trained weights. Two famous models inceptionresnetv2 (IRNV-2) and densenet201 (DN-201) were used with their learned weights for re-training the NN on new data. Transfer learning-based pre-trained NN improved the prediction of labels for different feet' positions. It especially reduced the variations in the predictions in the last stage of the gait swing and stance phases. An accuracy of 94% was achieved in predicting the class labels of the test data. Since the variation in predicting the true label was primarily one frame, it could be ignored at a frame rate of 30 frames per second.

The predicted labels were used to extract various spatiotemporal parameters of the gait, which are critical for any gait analysis system. A total of 12 gait parameters were quantified and compared with the ground truth obtained by observational methods. The NN-based spatiotemporal parameters showed a high correlation with the ground truth, and in some cases, a very high correlation was obtained. The results proved the usefulness of the proposed method. The parameter's value over time resulted in a time series, a long-term gait representation. This time series could be further analyzed using various mathematical methods. As the third contribution in this dissertation, improvements were proposed to the existing mathematical methods of time series analysis of temporal gait data. For this purpose, two refinements are suggested to existing entropy-based methods for stride interval time series analysis. These refinements were validated on stride interval time series data of normal and neurodegenerative disease conditions downloaded from the publicly available databank PhysioNet. The results showed that our proposed method made a clear degree of separation between healthy and diseased groups.

In the future, advanced medical support systems that utilize artificial intelligence, derived from the methods introduced here, could assist physicians in diagnosing and monitoring patients' gaits on a long-term basis, thus reducing clinical workload and improving patient safety.

Kurzfassung

Mit der Entwicklung von Deep-Learning-Techniken sind Deep-acNN-basierte Methoden zum Standard für Bildverarbeitungsaufgaben geworden, wie z. B. die Verfolgung menschlicher Bewegungen und Posenschätzung, die Erkennung menschlicher Aktivitäten und die Erkennung von Gesichtern. Deep-Learning-Techniken haben den Entwurf, die Implementierung und den Einsatz komplexer und vielfältiger Anwendungen verbessert, die nun in einer Vielzahl von Bereichen, einschließlich der Biomedizintechnik, eingesetzt werden. Die Anwendung von Computer-Vision-Techniken auf die medizinische Bild- und Videoanalyse hat zu bemerkenswerten Ergebnissen bei der Erkennung von Ereignissen geführt. Die eingebaute Fähigkeit von convolutional neural network (CNN), Merkmale aus komplexen medizinischen Bildern zu extrahieren, hat in Verbindung mit der Fähigkeit von long short term memory network (LSTM), die zeitlichen Informationen zwischen Ereignissen zu erhalten, viele neue Horizonte für die medizinische Forschung geschaffen. Der Gang ist einer der kritischen physiologischen Bereiche, der viele Störungen im Zusammenhang mit Alterung und Neurodegeneration widerspiegeln kann. Eine umfassende und genaue Ganganalyse kann Einblicke in die physiologischen Bedingungen des Menschen geben. Bestehende Ganganalyseverfahren erfordern eine spezielle Umgebung, komplexe medizinische Geräte und geschultes Personal für die Erfassung der Gangdaten. Im Falle von tragbaren Systemen kann ein solches System die kognitiven Fähigkeiten beeinträchtigen und für die Patienten unangenehm sein.

Außerdem wurde berichtet, dass die Patienten in der Regel versuchen, während des Labortests bessere Leistungen zu erbringen, was möglicherweise nicht ihrem tatsächlichen Gang entspricht. Trotz technologischer Fortschritte stoßen wir bei der Messung des menschlichen Gehens in klinischen und Laborumgebungen nach wie vor an Grenzen. Der Einsatz aktueller Ganganalyseverfahren ist nach wie vor teuer und zeitaufwändig und erschwert den Zugang zu Spezialgeräten und Fachwissen.

Daher ist es zwingend erforderlich, über Methoden zu verfügen, die langfristige Daten über den Gesundheitszustand des Patienten liefern, ohne doppelte kognitive Aufgaben oder Unannehmlichkeiten bei der Verwendung tragbarer Sensoren. In dieser Arbeit wird daher

eine einfache, leicht zu implementierende und kostengünstige Methode zur Erfassung von Gangdaten vorgeschlagen. Diese Methode basiert auf der Aufnahme von Gehvideos mit einer Smartphone-Kamera in einer häuslichen Umgebung unter freien Bedingungen. Deep neural network (NN) verarbeitet dann diese Videos, um die Gangereignisse zu extrahieren. Die erkannten Ereignisse werden dann weiter verwendet, um verschiedene räumlich-zeitliche Parameter des Gangs zu quantifizieren, die für jedes Ganganalysesystem wichtig sind.

In dieser Arbeit wurden Gangvideos verwendet, die mit einer Smartphone-Kamera mit geringer Auflösung außerhalb der Laborumgebung aufgenommen wurden. Viele Deep-Learning-basierte NNs wurden implementiert, um die grundlegenden Gangereignisse wie die Fußposition in Bezug auf den Boden aus diesen Videos zu erkennen. In der ersten Studie wurde die Architektur von AlexNet verwendet, um das Modell anhand von Gehvideos und öffentlich verfügbaren Datensätzen von Grund auf zu trainieren. Mit diesem Modell wurde eine Gesamtgenauigkeit von 74% erreicht. Im nächsten Schritt wurde jedoch die LSTM-Schicht in dieselbe Architektur integriert. Die eingebaute Fähigkeit von LSTM in Bezug auf die zeitliche Information führte zu einer verbesserten Vorhersage der Etiketten für die Fußposition, und es wurde eine Genauigkeit von 91% erreicht. Allerdings gibt es Schwierigkeiten bei der Vorhersage der richtigen Bezeichnungen in der letzten Phase des Schwungs und der Standphase jedes Fußes.

Im nächsten Schritt wird das Transfer-Lernen eingesetzt, um die Vorteile von bereits trainierten tiefen NNs zu nutzen, indem vortrainierte Gewichte verwendet werden. Zwei bekannte Modelle, inceptionresnetv2 (IRNV-2) und densenet201 (DN-201), wurden mit ihren gelernten Gewichten für das erneute Training des NN auf neuen Daten verwendet. Das auf Transfer-Lernen basierende vortrainierte NN verbesserte die Vorhersage von Kennzeichnungen für verschiedene Fußpositionen. Es reduzierte insbesondere die Schwankungen in den Vorhersagen in der letzten Phase des Gangschwungs und der Standphase. Bei der Vorhersage der Klassenbezeichnungen der Testdaten wurde eine Genauigkeit von 94% erreicht. Da die Abweichung bei der Vorhersage des wahren Labels hauptsächlich ein Bild betrug, konnte sie bei einer Bildrate von 30 Bildern pro Sekunde ignoriert werden.

Die vorhergesagten Markierungen wurden verwendet, um verschiedene räumlich-zeitliche Parameter des Gangs zu extrahieren, die für jedes Ganganalysesystem entscheidend sind. Insgesamt wurden 12 Gangparameter quantifiziert und mit der durch Beobachtungsmethoden gewonnenen Grundwahrheit verglichen. Die NN-basierten räumlich-zeitlichen Parameter zeigten eine hohe Korrelation mit der Grundwahrheit, und in einigen Fällen wurde eine sehr hohe Korrelation erzielt. Die Ergebnisse belegen die Nützlichkeit der vorgeschlagenen Methode. Der Wert des Parameters über die Zeit ergab eine Zeitreihe, eine langfristige Darstellung

des Ganges. Diese Zeitreihe konnte mit verschiedenen mathematischen Methoden weiter analysiert werden.

Als dritter Beitrag in dieser Dissertation wurden Verbesserungen an den bestehenden mathematischen Methoden der Zeitreihenanalyse von zeitlichen Gangdaten vorgeschlagen. Zu diesem Zweck werden zwei Verfeinerungen bestehender entropiebasierter Methoden zur Analyse von Schrittintervall-Zeitreihen vorgeschlagen. Diese Verfeinerungen wurden an Schrittintervall-Zeitseriendaten von normalen und neurodegenerativen Erkrankungen validiert, die aus der öffentlich zugänglichen Datenbank PhysioNet heruntergeladen wurden. Die Ergebnisse zeigten, dass die von uns vorgeschlagene Methode eine klare Trennung zwischen gesunden und kranken Gruppen ermöglicht.

In Zukunft könnten fortschrittliche medizinische Unterstützungssysteme, die künstliche Intelligenz nutzen und von den hier vorgestellten Methoden abgeleitet sind, Ärzte bei der Diagnose und langfristigen Überwachung des Gangs von Patienten unterstützen und so die klinische Arbeitsbelastung verringern und die Patientensicherheit verbessern.

Acknowledgments

In the name of Allah, the most Merciful, the most Gracious. It is all Allah's blessings that have enabled me to complete the thesis such that I have been able to demonstrate critical curiosity, constructive insight, and steadfastness.

My deepest gratitude is extended to my esteemed supervisor, Prof. Dr. rer. nat. Werner Nahm for accepting me as a Ph.D. student in his group and for his guidance, mentorship, scientific enthusiasm, and support during challenging times.

I am highly thankful to Prof. Dr. rer. nat. Wilhelm Stork for being my first contact in Germany, my reason for joining KIT, and for agreeing to act as the second examiner.

During these 4 years, I met with many colleagues in the Institute of Biomedical Engineering (IBT), KIT, and they all deserve special thanks because everyone guided me and was there to help me in any matter. A very big thanks to the colleagues and friends who read some chapters of my thesis and gave me helpful feedback: Lorena Krames, Matthias Schaufelberger, Alexander Anton, Miriam Weiss, Mahwish Fatima, and Sajjad Hussain. Ron and his team started this project; I am thankful to them and Cagla as well, who supported me during data collection.

Furthermore, I would take this opportunity to thank the German academic exchange service (DAAD) for their assistance and support during my stay in Germany. I am also thankful to Karlsruhe house of young scientists (KHYS) for providing additional funding to complete my Ph.D. If my university in Pakistan had not permitted me and provided me with funding for my Ph.D., I would not have been able to complete it. Big thanks to all the administration and my colleagues in the WUAJK, Bagh AJK, Pakistan. A big credit goes to the Government of Pakistan and the Higher Education Commission (HEC) for funding my Ph.D.

A very big thanks to all my friends who made my stay in Germany and Karlsruhe very comfortable despite being away from my family. I am thankful to Vikram, Usman, and Bilal, who were with me from my first day in Germany until now. Our long video calls, and tours across Europe, helped me cope with the challenges I faced while completing my Ph.D. Many thanks to all my friends in Karlsruhe, who were very supportive and always available for

me. I couldn't mention all names here, but I am especially thankful to my big brother Sajjad Hussain who was with me from my first day in Karlsruhe.

At this moment, I am indebted to my late parents, who always supported and encouraged me throughout my life. I wish you were alive at this moment, but I pray that may Allah bless you, and I hope you will be happy to hear this. I am deeply grateful to all my family members, especially my brothers, sisters, and others, for always being by my side and praying for my success.

Finally, I would like to express my heartfelt gratitude and love for my wife and kids, who were always a source of motivation for me. They suffered a lot during my stay in Germany, but my wife did not only take care of the kids but also managed to complete all the responsibilities that I was supposed to do. I could not have accomplished this without your support and understanding. I love you all.

Contents

Abbreviations

AD_Walk	alzheimer's disease walking 86, 88
$F1_{Macro}$	macro F-1 score 72, 75, 99, 101
$F1_{Micro}$	micro F-1 score 72, 75, 99, 101
$Injury\_Walk$	injury walking 86, 88
$N2AI$ **dataset**	N2AI 93, 97
$Norm\_Walk$	normal walking 86, 88
NPH_Walk	NPH walking 86, 88
$PREC_{WAvg}$	weighted average precision 72, 75, 99, 101
r	Pearson's correlation coefficients 103
REC_{WAvg}	weighted average recall 72, 75, 99, 101
SI_{med}	median of SI time series 123, 130
SI_{mod}	mode of SI time series 123, 130
ACC	accuracy . . . . . 71 ff., 75 f., 79 f., 97 ff., 101
AD	Alzheimer's disease . . . . . 21, 23, 83, 85 f.
AI	artificial intelligence 25
ALS	Amyotrophic Lateral Sclerosis . . . . 2, 22 f., 116, 125 f., 128–131
ApEn	approximate entropy . . . . . 114 f.
ASI	average stride interval . . . . . 123, 130 f., 134
AvACC	mean accuracy 101
BA plot	Bland-Altman plot 103
CCA	canonical correlation analysis 92
CL	convolutional layer 62, 67
CNN	convolutional neural network . i, iii, 5, 25, 29 ff., 33, 40, 42, 50, 53, 61 ff., 65, 67, 71, 74, 79 ff., 86, 88, 90 f., 140, 142 f.
CSF	cerebrospinal fluid 23
DBN	deep belief network 29
DN	DenseNet 91
DN-201	densenet201 . . . . . ii, iv, 90, 93, 97 f., 101 f., 104 f., 140

DSN deep stacking network 29
F1-Score F-1 score . . . . . 71 ff., 75 f., 80, 98 f., 101
FC fully connected layer 62, 64
FPA foot progression angle 17
GC gait cycle . . . . . 9–12, 14 f., 93 f., 114, 141
GRU gated recurrent unit 29
GSR Gait Stability Ratio 16
HD Huntington disease . . . . . 21 ff., 116, 125–131, 134
HS heel strike . . . . . 3, 9, 51, 81, 94 f.
IC initial contact 10, 84
IRNV-2 inceptionresnetv2 . . . . . ii, iv, 90, 93, 97–102, 140
LCC Lin's concordance correlation coefficient 103, 106
LCDN local contrast divisive normalization 64
LED light emitting diodes 85
LSTM long short term memory network i–iv, 29, 38 f., 50, 67 ff., 74, 79 ff., 93, 140, 142
MCI mild cognitive impairment 83, 85
ML machine learning 5, 25
MRI magnetic resonance imaging 23
MsEn multiscale entropy . . . . . 114 ff., 119, 124–128, 133, 135, 141
NN neural network i f., iv, 5, 25–28, 51, 56, 58, 62, 69, 83, 92 f., 102 f., 106 f., 140, 145, 151
NPH Normal Pressure Hydrocephalus . . . 23, 53, 55, 58, 68, 71, 79, 86
NPV Negative predictive value 45
OFS optimal features subsets 66, 92
PCA principal component analysis 66, 92
PD Parkinson disease 20, 23, 116, 125–131, 134
PL pooling layer . . . . . 62, 67 f.
PPV positive predictive value 44, 53
PREC precision . . . . . 71 ff., 75 f., 80, 98 f., 101
Pred_Params predicted parameters . . . . . 102 f.
ReLU rectified linear unit 62
RN ResNet 91
RNN recurrent neural network . . . . . 29, 37 f., 50, 53, 66 f.
ROS random oversampling 58
RR RR 133

SampEn sample entropy . . . . . 114 f., 118 f., 125
SEn Shannon entropy . . . . . 42, 121 f.
SENS sensitivity . . . . . 53, 71 ff., 75 f., 80, 98–101
SGD stochastic gradient descent 69
SI stride interval . . . . . 9, 50, 93, 96, 116, 120, 123 f.
SPEC specificity . . . . . 72 f., 75 f., 80, 98 f., 101
SyEn symbolic entropy . . . . . 114 ff., 122 f., 130, 134
TL transfer learning . . . . . 40 f.
TNCSE threshold dependent normalized corrected Shannon entropy . . 115, 120 ff., 129 ff., 134 f.
TO toe off 3, 17, 51, 81
TPR true positive rate 44
True_Params true parameters . . . . . 102 f.
VSI variation in stride interval . . . . . 123, 130 f., 134 f., 141

Pre-publication of this Thesis

This thesis contains portions that have already been published, verbatim or in part, before submission. A reference in the table below corresponds to the relevant chapter of the thesis.

Chapter	Publication
Chapter 5	Abbasi AQ, Abbasi MM, Nahm W. On Selection of Threshold Values for Symbolic Entropy Analysis of Human Gait. *Annu Int Conf IEEE Eng Med Biol Soc*. 2022;2022:3640-3644. doi:10.1109/EMBC48229.2022.9871100
Chapter 5	A.Q. Abbasi, W. Nahm, M. Manshad, Gait analysis using stride interval of left and right foot in single time series under normal and neurodegenerative diseased conditions, Gait Posture,Volume 81, Supplement 1, 2020,

Chapter 1

Introduction

1.1 Motivation

Deep learning methods for computer vision have proven to be an exciting field of research in various biomedical fields. In recent years, deep learning has become increasingly popular for its ability to design, implement, and deploy complex and diverse applications, which are used in modern fields, including biomedical engineering. As technological advancements and the availability of high computational power and modern graphic processing units have increased drastically in the recent past, deep learning has been shown to provide excellent results when analyzing and recognizing medical images and events. Gait is one of the important biomarkers which can provide meaningful information predicting various disorders, including stroke, dementia, Parkinson's disease, arthritis, etc, [1–3]. Using deep learning methods for video-based gait analysis is beneficial for monitoring the progression of various such disorders.

Aging is associated with many of these disorders. There has been an increase in the number of elderly adults falling on flat surfaces due to the emergence of such diseases due to an aging population. A significant amount of treatment expenditures was incurred for gait-related diseases. An estimated 32% of community-dwelling elderly adults over 75 years of age will fall due to a gait disorder at least once in a calendar year, and 24% of these individuals will sustain serious injuries [4, 5]. Medical costs associated with falls are substantial in the United Kingdom (UK); fall-related injuries in adults over 60 are estimated to cost more than 981 million pounds annually [6]. The healthcare expenditures for elderly fall victims ranged from $48 million in Alaska to $4.4 billion in California. A study of fall-related injuries found that lifetime medical costs ranged from $68 million in Vermont

to $2.8 billion in Florida [6]. Because of this, falling has become a significant issue for the growing number of elderly people [4, 5, 7].

Another gait-related disease Amyotrophic Lateral Sclerosis (ALS) is a progressive neurodegenerative disease that results in the loss of motor neurons in the upper and lower limbs. In Germany, the prevalence-based total burden of illness per patient per year was 519,776,352 euros; the lifetime cost per patient was estimated at 246,184 euros [8]. As a result, gait-related problems and fall risk have been increasingly detected and recognized through the implementation of safety measures in high-risk workplaces, hospitals, and nursing homes [9]. A gait analysis can provide insight into a person's pattern of walking. A person's gait may be a marker of physical changes and the possibility of developing gait-related diseases and falling [10]. The human body's gait refers to the lower limbs' behavior during upright walking. For a normal gait cycle to be achieved, naturalness, coordination between the legs, labor saving, and periodicity must be achieved. An abnormal gait may occur before a gait disease is developed or during any gait disorder. There are many reasons why an abnormal gait might occur. Identifying and evaluating abnormal gait patterns are instrumental in guiding lower limb training regimens and flat-ground fall prevention strategies in medical rehabilitation. Gait patterns can be monitored in elderly patients so that proper preventive measures can be recommended to reduce the risk of gait disorders. Many syndromes associated with different gait disorders can be diagnosed through gait analysis, parameter quantification, and its interpretation [11]. Due to the inadequacy of human vision to recognize and quantify gait patterns, computer vision has become a hot topic in biomechanics and healthcare research in recent years [10, 12, 13].

In the literature, significant number of articles has discussed the use of computer vision for gait analysis. Nevertheless, most of them are devoted to gait biometrics for the identification of humans, and only a few have focused gait analysis for the detection of abnormal gait patterns[10, 12, 13].

As the state-of-the-art has investigated that the gait parameters are usually recorded under controlled conditions in a gait laboratory, using either wearable or non-wearable systems equipped with floor sensors or multiple cameras. On the other hand, it is known that patients move very consciously and, therefore, unnaturally under a strict laboratory environment. Therefore, there is a risk that the data collected in this way are subject to bias. Furthermore, there is also a need for flexible gait analysis methods that can be used, for example, in day-to-day hospital care. The overall goal of this study was to provide the basis for the development of an inexpensive and easy-to-implement method of obtaining spatiotemporal gait parameters that will then be used to distinguish between normal and abnormal gait patterns.

1.2 Aim of the Thesis

This research aims to provide a basis for the development of a simple solution for detecting gait events directly through videos captured in a home environment by means of classifying the feet' position, and then extracting spatiotemporal parameters based on those events. The objective is to propose a technological solution that facilitates specialists with objective gait measurements, helping to improve the objectivity of the gait analysis performed by specialists.

This thesis proposes recording a subject's gait with a smartphone camera and then using computer vision algorithms to analyze the captured sequences to extract spatiotemporal gait parameters. The parameters are used to develop time series based on their values at specific times and then applying time series-based analysis methods to identify the presence or absence of abnormalities. The overall schematic illustration of the work is presented in Figure 1.1.

In this thesis, the following research questions are analyzed and investigated:

- Using deep learning techniques, how can we accurately classify the feet' position based on low-resolution marker-less walking videos captured via the frontal and postern views, outside the laboratory environment?
- How can we detect the basic gait events like heel strike (HS) and toe off (TO) from the feet's position and how to quantify spatiotemporal parameters of gait from those events?
- To what extent are the spatiotemporal parameters, quantified from the gait events detected from videos, reliable and can be used for gait classification?
- What are the effects of the refinements to the existing time series-based mathematical methods for gait time series analysis?

The overall goal of this study is to provide a basis for the development of an inexpensive and easy-to-implement method of obtaining spatiotemporal gait parameters that will then be used to distinguish between normal and abnormal gait patterns.

1.3 Structure of the Thesis

Part I introduces the relevant physiological and technical fundamentals for understanding the proposed approaches and results:

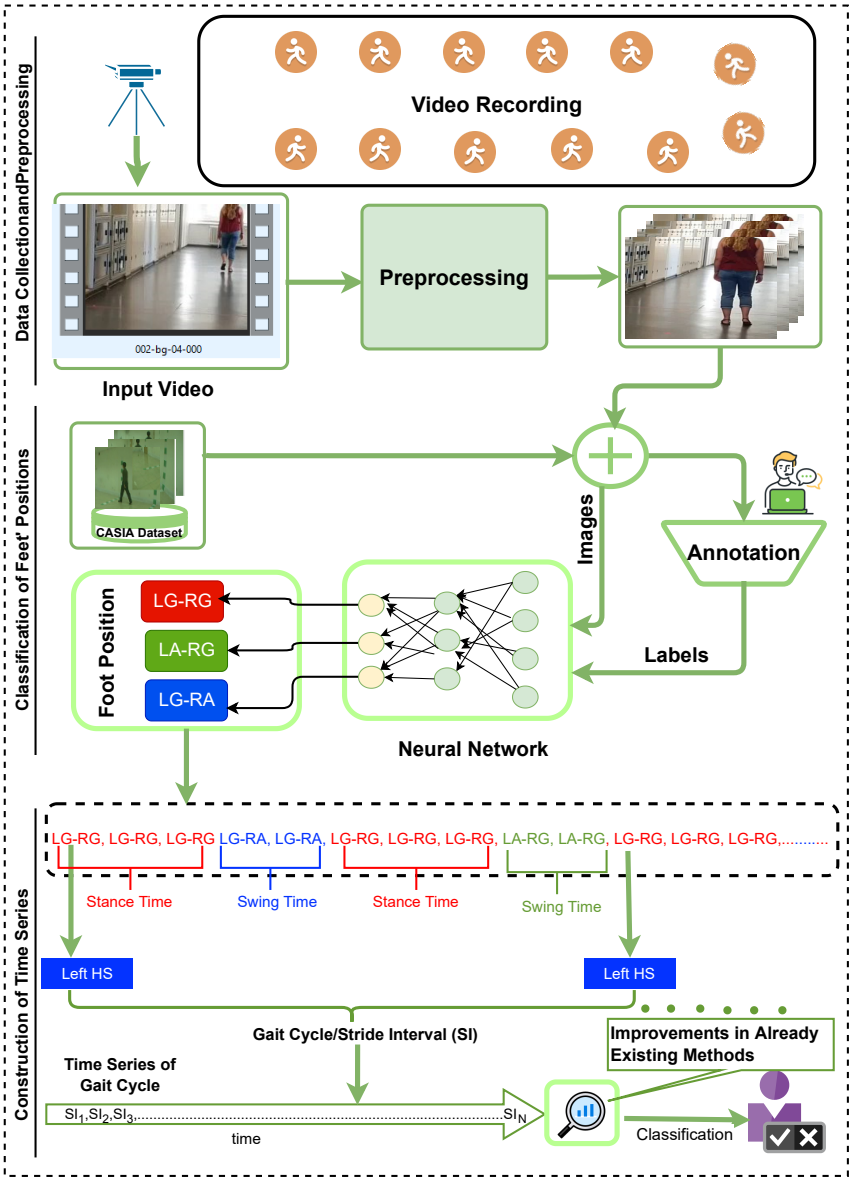


Figure 1.1: Illustration of the overall project, it has three parts, first is the data collection in form of videos using a smartphone camera and preprocessing; second is the classification of feet positions, and third is the spatiotemporal parameters extraction from the gait events and the time series based gait analysis

Chapter 2 provides the physiological fundamentals that are important for this thesis. First, there is a brief introduction to the anatomy and functionality of the gait and the gait cycle. Then, different terminologies and concepts related to gait analysis are summarized. Finally, various quantifiable parameters, which can be used for the analysis of gait symmetry, variability, and quality, are discussed.

Chapter 3 discusses the basic concepts about machine learning (ML) and its types. Then the basic principles and different types of artificial neural networks (NNs) are explained with more focus on deep learning and deep convolutional neural network (CNN). A brief explanation of the basic entropy measure i.e., Shannon entropy is also discussed, followed by a short description of statistical methods. In the end, different performance measuring metrics are briefly explained.

Part II presents the studies related to video-based gait analysis and the extraction of spatiotemporal parameters extraction from the gait events.

Chapter 4 discusses our basic study in which deep learning methods were used to classify the feet's position with respect to the ground from the videos recorded outside laboratory conditions, and without any marker on the body of the subject. The motivation and objective of the study are described in the introduction section, then related work is presented. Afterwards, the data collection setup and the data preparation steps are discussed. In the next section, different deep learning models and their training strategies are discussed. Finally, the results in terms of the performance of the different classifiers in predicting the gait events, are presented with the discussion and conclusion.

Chapter 5 presents the extraction of important gait parameters (which are further used for gait analysis) from the gait events detected by NNs using transfer learning methods. It is divided into two parts, the first part is the use of transfer learning methods to improve the performance of our models which were developed and presented in 4 of this thesis. In the second part, the methods to extract the spatiotemporal parameters from the gait events are presented and the same are compared with the ground truth parameters calculated from videos via observations. In the end, a discussion and conclusion are given.

Part III describes the time series-based gait analysis and refinements to entropy-based methods.

Chapter 6 presents improvements to the entropy-based methods that have been used for time series analysis of human gait dynamics, and have proven to be highly effective for a wide range of biological time series analyses as well. It is discussed in the first section that a double-foot stride interval should be used rather than a single-foot stride interval. The second section focuses on optimizing threshold values for symbolic entropy analysis methods for gait analysis. There is a discussion and conclusion in the last section of the chapter.

Part IV presents the summary of the overall work and gives an overall conclusion with certain limitations. Furthermore, possible future research topics and directions are identified.

Chapter 7 provides the general conclusion with a summary of the works presented in this thesis. It also presents some limitations of the work.

Chapter 8 gives some recommendations regarding possible future work related to this research, based on our results.

PART I

FUNDAMENTALS

Chapter 2

Physiological Fundamentals

This chapter describes the physiological fundamentals that are important for this thesis. First, there is a brief introduction of the anatomy and functionality of the gait and the gait cycle. Different terminologies and concepts related to the gait analysis are summarised. Various quantifiable parameters, which can be used for the analysis of gait symmetry, variability, and quality, are discussed. Finally, different pathologies for which gait disturbance is one of the early symptoms are discussed with their prevalence in the world.

2.1 Gait and Gait Cycle

Gait refers to the way we walk or describes how we move. An individual's pattern of walking is referred to as their gait. During walking, muscles are coordinated and balanced so that the body is propelled forward in a rhythm known as the stride [14]. In order to move forward, the left and right feet perform repetitive movements, and this repetitive pattern is referred to as the gait cycle (GC). There are two main stages in the GC: the stance stage and the swing stage. The stance phase is the duration in which the foot is placed on the ground. The duration when the foot is raised from the ground until the heel is placed on the ground is the swing phase.

The interval between two consecutive same gait events of the same foot is referred to as the gait cycle duration or stride interval (SI) [15, 16]. Typically, it is considered as the time between two consecutive heel strikes of the same foot as in normal walking, the heel is the first part of the foot to contact the ground. One complete gait cycle is shown in Figure 2.1, starting with the heel of the right foot touching the ground and ending with the same heel touching the ground again. There are several sub-phases that occur between these two heel strikes (HSs), which are described in the following sections.

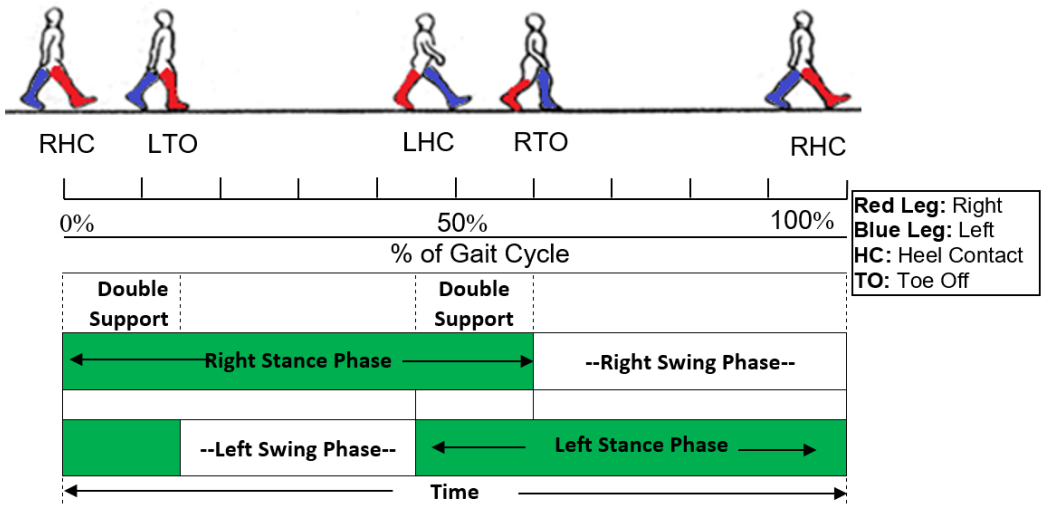


Figure 2.1: A complete GC starting from the heel strike of the right foot and ending at the next heel strike of the same foot. Stance phase is almost 60% of the gait cycle while swing phase is almost 40% of the GC [17].

2.2 Gait Cycle-Sub Phases

The stance and the swing phases of the stride may be further divided into specific sub-phases related to the normal function. These sub-phases are initial contact, loading response, mid stance, terminal stance, pre-swing, initial swing, mid swing, and terminal swing [16]. During the stance phase, weight is accepted on the foot and during the swing phase limbs are advanced to move forward. This terminology is very useful for referring to specific portions of the GC when describing pathological gait. The division of stance and swing into sub-phases is shown in Figure 2.2 and described in the following sections.

2.2.1 Initial Contact

The first part of the foot that strikes the ground is the heel. The right foot is touching the ground (Figure 2.3 (a)). This is the initial contact (IC) of the right leg. At this moment, the knee is stretched, the hip is flexed, and the ankle is dorsiflexed to neutral. In this phase, the left leg is in the completion of the terminal stance phase [18].

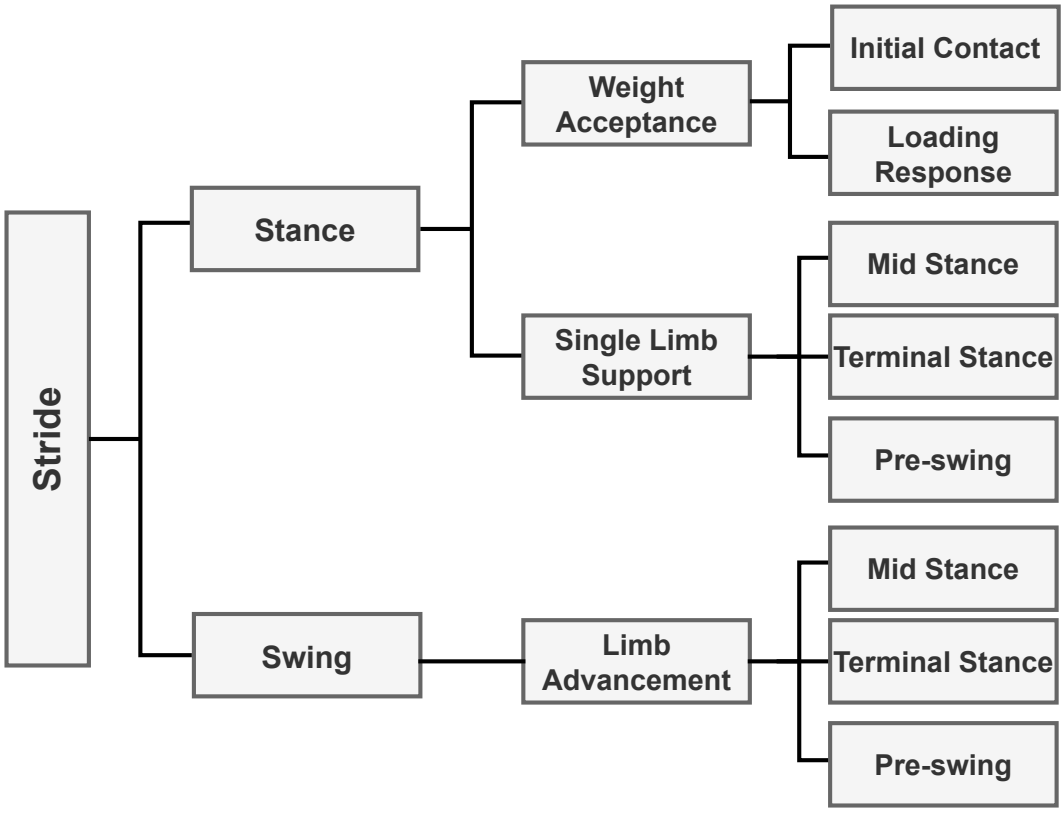


Figure 2.2: Division of stride and swing phases of the GC into sub-phases

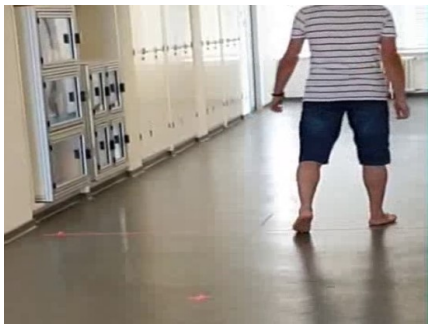
2.2.2 Loading Response

While one foot touches the floor and continues touching till the second foot is raised for swing, the right leg got the whole-body weight (Figure 2.3 (b)). This phase is essential for jolt absorption, load bearing, and ahead movement. The left foot is in the pre-swing phase. The next job of the cycle (GC) is one limb support. In this phase, a single limb essentially supports the whole-body weight and provides truncal constancy whereas forward movement remains continued [18].

2.2.3 Mid Stance

This is the first half of the one-limb support interval. It starts with lifting the left leg and lasts till the weight is aligned to the other foot (Figure 2.6 (a)). The right leg progresses over the right foot with ankle dorsiflexion, meanwhile, the knee and hip are extended. The left leg is progressing in the loading response stage [18].

(a) Right leg initial contact

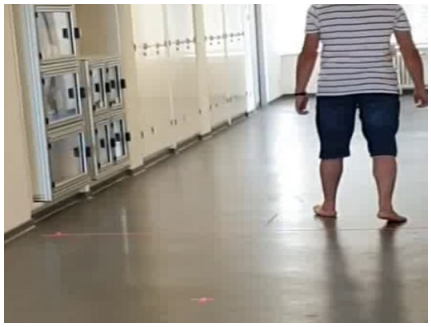


(b) Right leg loading response

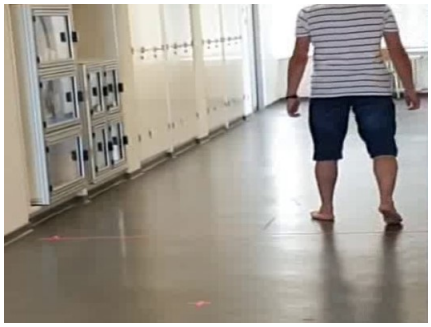
Figure 2.3: The two sub-phases of the GC. One is initial contact (a), and the other is loading response (b) (images from the gait data collection).

2.2.4 Terminal Stance

Starts when the right heel is upswing and remains until the left foot heel touches the ground (Figure 2.6 (b)). The weight of the body progresses away from the right foot because the increased extension in the hip places the leg in a more trailing situation [18].



(a) Right leg mid stance



(b) Right leg terminal stance

Figure 2.4: The Two sub-phases of the GC. (a) shows the mid stance and the (b) the terminal stance (images from the gait data collection).

2.2.5 Pre Swing

Pre swing is the second double stance interval during one GC. It starts with the initial contact of the left foot and finishes when the right toe-off (Figure 2.5 (a)). When the left leg

contracts, the right foot makes growth in ankle plantar flexion, reduce hip extension, and rise knee flexion. The weight is transferred to the opposite limb from the ipsilateral [18].

2.2.6 Initial Swing

It begins when a foot is raised from the ground and finishes when the swinging leg is opposite to the stance leg. The right leg is forwarded with an increase in knee and hip flexion (Figure 2.5 (b)). The ankle is dorsiflexed to some extent, to ensure floor clearance. A footdrop gait is apparent during this phase. Here the other (left) leg is in the mid stance phase [18].

(a) Right leg pre swing **(b)** Right leg initial swing

Figure 2.5: The two sub-phases of the GC. One is pre swing (a), and the other is initial swing (b) (images from the gait data collection).

2.2.7 Mid Swing

This phase starts with the end of the initial swing and remains until the tibia is vertical and the swinging-limb is at front of the body. Progression of the right leg is made by more hip flexion (Figure 2.6 (a)). In response to gravity, the knee is permitted to extend. It is the late mid stance phase of left leg [18].

2.2.8 Terminal Swing

Starts when the tibia is upright and finishes when the foot strikes the floor. Knee extension causes limb advancement. The ankle remains dorsiflexed to neutral, and the hip keeps its

flexion (Figure 2.6 (b)). The other foot is in the terminal stance phase [18]. When human

(a) Right leg mid swing

(b) Right leg terminal swing

Figure 2.6: The two sub-phases of the GC. One is Mid Swing (a), and the other is Terminal Swing (b) (images from the gait data collection).

walk walk, different steps are successively performed in a periodic way. Figure 2.7 shows the transition from one gait event to the subsequent gait event inside a GC.

2.3 Spatiotemporal Gait Parameters

There are several parameters that can be used to quantify gait. Spatial-temporal gait parameters include spatial (distance-based) as well as temporal (time-based) parameters. Quantifying aspects of the GC, such as time and spatial measures, allow to analyse the gait symmetry, variability and quality. In the following subsections, some important spatiotemporal gait parameters are introduced [19–24]:

- **Step Length**
  It is the distance between the point of contact of one foot with the ground and the following occurrence of the same point of contact with the other foot. When both feet are in contact with the ground, the right step length is the distance from the left heel to the right heel, expressed in meters (m) (Figure 2.8).
- **Step Time**
  The step time is measured by the time elapsed between a gait event of one foot and the following same event of the opposite foot. It is measured in seconds (s).
- **Stride Length**
  Traditionally, it is the distance between the initial contact of one foot and the following initial contact of the same foot (Figure 2.8). However, instead of the initial contact, it

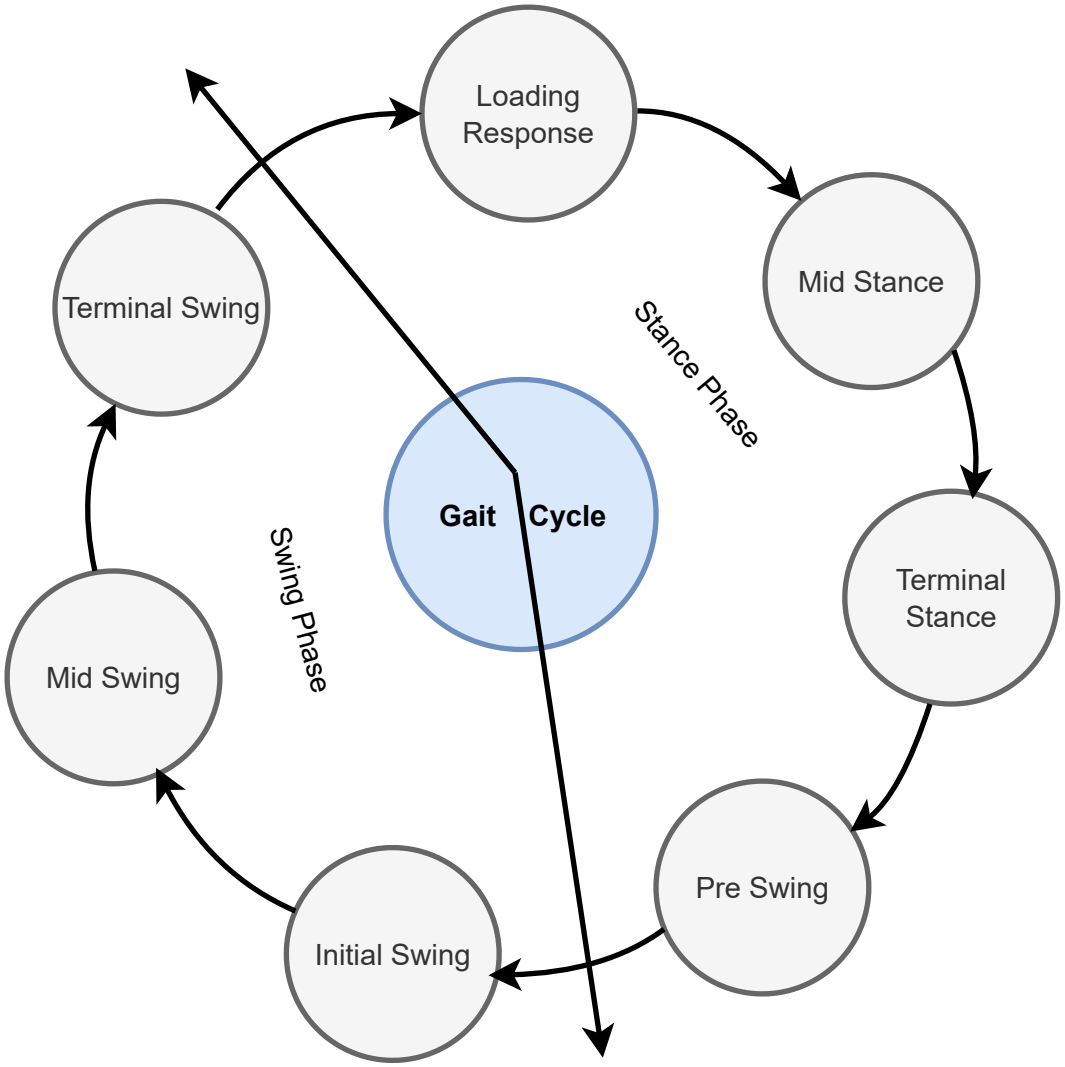


Figure 2.7: Gait is a cyclic or periodic process in which different sub-phases are performed after one another. Transition from one sub-phase to other sub-phase during the GC is shown.

could be any other gait event in the GC as well. It is sometimes referred to as cycle length and is measured in meters (m).

- **Stride Interval or Gait Cycle Time**

  It is the measure of the time elapsed between the initial contact of one foot and the next initial contact of the same foot. It is measured in seconds (s). This is the most important temporal parameter, and has been used in most gait related studies.

- **Gait Velocity**

  A person's gait velocity can be used as a possible measure of their functional status, because it can be directly observed [25]. It is defined as the distance traveled per second

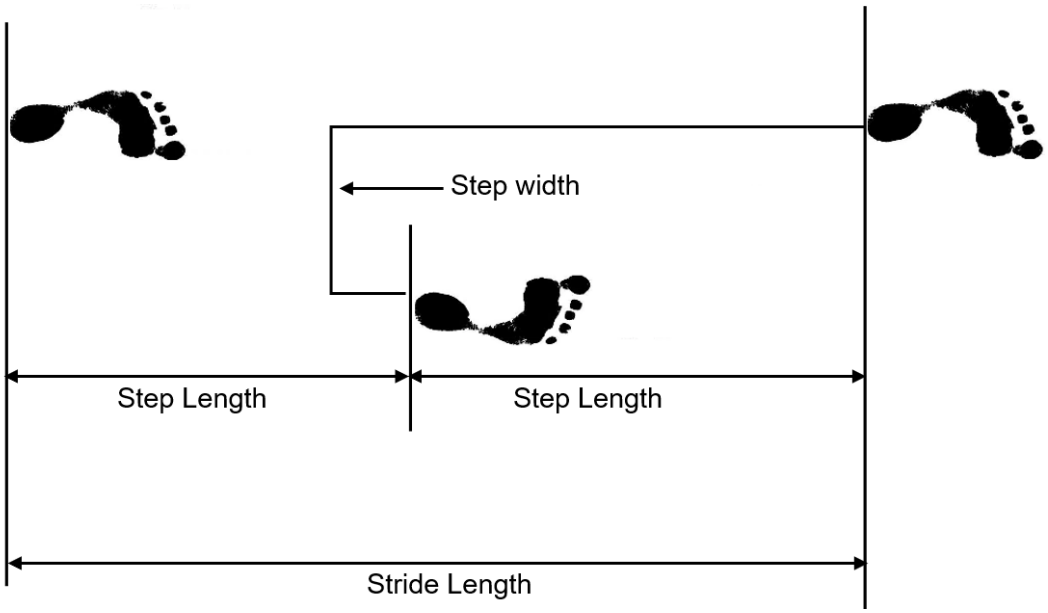


Figure 2.8: Illustration of step length, stride length, and step/stride width of a GC

(m/s). Due to its affordability and reliability, it is a suitable measure of the functional capacity for the clinical evaluation of gait speed. Additionally, each individual has a self-selected speed that is highly dependent on the environment, the floor type, the footwear, and their disability. Moreover, gait velocity is dependent on stride length and cadence, and an increase or decrease in either of the parameters (stride length or cadence) may affect not only adults, but also toddlers [26]. In addition, gait velocity can be used as a screening tool to identify health-related outcomes such as quality of life, and physical, cognitive, and functional decline[25].

- **Cadence**

  The number of steps taken per minute is referred to as cadence. Cadence can be measured and assessed through patients with any deviations [27] where an increase in cadence may indicate an improvement in walking abilities and vice versa. Also, cadence has a direct relation to intensity and can be used to identify different patterns of walking attitudes in free-living environments by following the amount of time we spend at a higher level. Additionally, cadence is considered an important objective measure of functional mobility and virtually affects all patients with gait abnormalities [28, 29].

- **Gait Stability Ratio**

  As expressed in steps/meters, the Gait Stability Ratio (GSR) is the ratio of cadence to gait velocity. Despite limited studies that have documented GSR, it is imperative to

identify the dynamic component of walking in older adults. It provides a measure of walking stability per unit of distance in a gait cycle [30].

- **Stance/Swing Ratio**

  The ratio of the time when the foot is on the ground to the time when the foot is in the air. For a normal person it is 60:40, but when there is some disturbance with the gait, normally the stance phases is increased as patient will stop more between taking the steps.

- **Stride Width**

  It is also referred to as Walking Base. The distance between the two feet is also very important parameter while gait analysis. Basically, it is a measurement of how far apart the two feet are from one another (side to side). This is usually measured from the ankle joint center of one foot to the same point on the other foot. For some of the gait related disease, it is increased; while for others, it is decreased.

- **Foot Progression Angle**

  It is important to determine the foot progression angle (FPA) when assessing patients' knee adduction moments. This angle is defined as the angle between the direction of the heading and the orientation of the foot.It is averaged from heel strike to toe off (TO) during the stance phase of walking for each step (toe-in angle is positive and toe-out angle is negative) [31]. For one step, it is simply called step angle.The foot progression angle of right and left feet is shown in Figure 2.9.

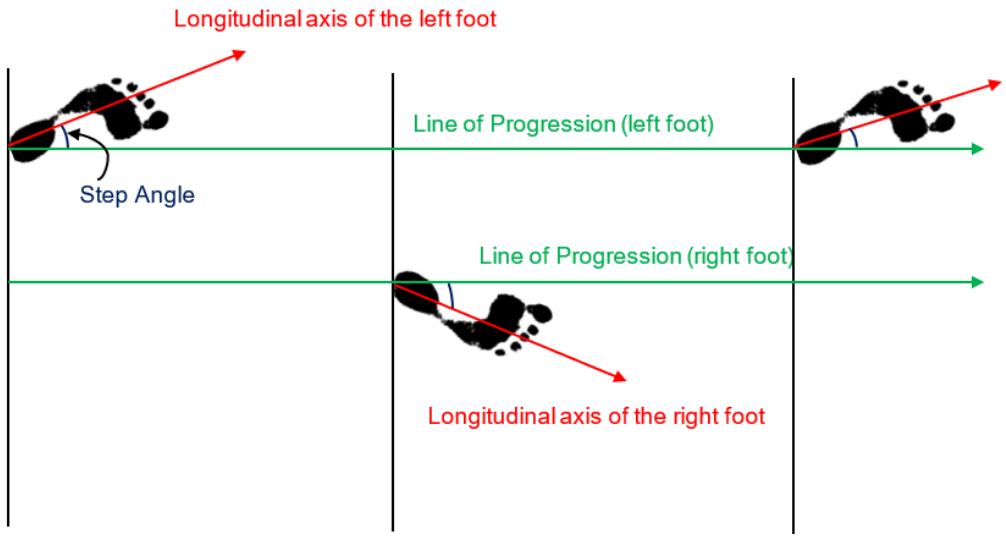


Figure 2.9: Step angle during walking, it is the angle between line of progression and longitudinal axis of the respective foot

2.4 Relevance of Parameters in Gait Analysis

There are various parameters (discussed in the preceding section) that contribute to the characteristics of human gait, which are evaluated qualitatively and quantitatively during research related to gait. These parameters may vary and depend upon the field of research and application. Table 2.1 demonstrates the summary of important gait parameters and their relevance with respect to gait analysis [32].

Table 2.1: Relevance of various spatiotemporal parameters of gait in different fields.[32]

No.	Gait Parameter	Relevance		
		Medical	Security/Recogonition	Sports
1.	Step Length	Yes	Yes	Yes
2.	Step Time	Yes		
3.	Step Width	Yes	Yes	Yes
4.	Stride Length	Yes	Yes	Yes
5.	Stride Time	Yes	Yes	Yes
6.	Cadence	Yes	Yes	Yes
7.	Step Angle	Yes	Yes	Yes
8.	Velocity	Yes	Yes	Yes
9.	Swing Time	Yes		
10.	Stance Time	Yes		
11.	Traversed distance	Yes		Yes
12.	Stop duration	Yes		
13.	Stop Events	Yes		Yes
14.	Fall	Yes		
15.	Joint Angles	Yes		Yes
16.	Body Segment Orientation	Yes		Yes
17.	Gait Autonomy	Yes		
18.	Body Posture	Yes	Yes	Yes
19.	Gait Cycle phases	Yes	Yes	Yes
20.	GSR	Yes	Yes	Yes

Some of the important spatiotemporal parameters, that are quantified from video-based gait analysis, are discussed in chapter 5 of this thesis.

2.5 Gait Analysis

Gait Analysis is the process of quantitatively or qualitatively analyzing different parameters of the gait individually and in groups to determine the state of the health/recovery. This

analysis is intended to determine the dynamic posture and coordination during movement, allowing for the evaluation, recording, and correction of any gait disorders [33]. There is a need to analyze the components of various phases of gait to diagnose various neurological disorders as well as to determine the progress of patients during rehabilitation and recovery from the effects of neurologic diseases, musculoskeletal injury, or amputation of a lower limb. Analyzing each phase of gait is an essential part of diagnosing a variety of neurologic conditions [19].

2.5.1 Normal Gait

The term "normal gait" refers to a dynamic sequence of movements of the trunk and the limbs that are rhythmical and alternating, and which results in the forward progression of the center of gravity [34]. It is usually established between the ages of 4-8 years old [35]. In the gait of a toddler, the trunk moves freely, the base of support is wide, the arms are in high guard positions, the foot is lifted during the swing, the feet are flat on the ground, and the toes are pointing outward during short, quick, rigid steps [36]. As a result of increased stability and limb length, mature gaits have reciprocal arm swings and heel strikes with higher velocity, cadence, step length, and single-limb stance times, as well as a ratio of pelvis span to ankle, and stance time of both feet on the ground [37]. Healthy adults between the ages of 59 and 80 have a preferred walking speed of approximately 1.4 m/s and an average stride length of between 150 and 170 cm. Young adults have an average cadence of 115 to 120 steps per minute [38].

2.5.2 Gait Deviations in the Elderly & Disease

In addition to having decreased muscle bulk, strength, and flexibility, elderly people suffer from some hearing and vision loss as well as a decrease in muscle bulk. Physicians must be able to distinguish between gait changes caused by disease or by aging. A reduction in velocity and a reduction in step/stride length are the major changes in gait. It has been found that when elderly people increase their velocity, they generally take more steps rather than increasing the length of their strides. Generally, the elderly have difficulty walking in situations where speed is required (such as crossing the street), or agility is required (such as walking on uneven surfaces or in crowds), or in the dark. In addition, the arm swing decreases, the pelvis rotates less, and the foot strikes and pushes off with a flat foot

approach. Neurodegenerative and other diseases cause a wide range of gait changes that will be discussed in the following sections.

2.6 Neurodegenerative Diseases

Essentially, neurodegenerative diseases are a set of conditions that primarily affect neurons in the human brain. Age-dependent disorders are becoming increasingly prevalent as a result of recent increases in the elderly population, a significant threat to human health. These disorders represent an important threat to human health [39]. One of the main contributing factors to the development of many neurodegenerative disorders is advancing age [40]. These disorders vary greatly in their pathophysiology. Some of the disorders where gait disturbance is a characteristic symptom are listed below.

2.6.1 Parkinson's Disease

It is a type of neurodegenerative disorder that affects the neurons producing dopamine ("dopaminergic"), which are located in a specific area of the brain called the substantia nigra, which plays a predominant role in the production of dopamine [41]. Parkinson disease (PD) is one of a variety of movement disorders and is characterized by muscle rigidity, tremor, a slowing of movement (bradykinesia), and, in extreme cases, a loss of movement (akinesia). As a result of the insufficient production and action of dopamine, which is produced in the dopaminergic neurons of the brain, the primary symptoms are the result of decreased stimulation of the motor cortex by the basal ganglia. There may also be secondary symptoms such as cognitive impairment and subtle language problems. In addition to being progressive, PD is also chronic in nature [42–48]. In the early stages of Parkinson's disease, gait disturbances are a common symptom. These effects are **i) shuffling:** PD results in short and barely-elevated steps, resulting in an audible shuffling sound as the feet barely leave the ground. Generally, small obstacles tend to trip the patient, **ii) reduced arm swing**, and **iii) festination:** this condition is characterized by short steps, stooped posture, and imbalance, leading to a faster and faster gait, often culminating in a fall, iv) Gait freezing: having difficulty moving the feet, especially in restricted or cluttered spaces or when initiating gait [49–55]

According to the Parkinson's disease foundation, nearly one million Americans are living with Parkinson's disease (an international organization devoted to treating Parkinson's disease)[56]. In 2030, the number is expected to rise to 1.2 million. Regarding prevalence,

Parkinson's disease follows Alzheimer's disease as the second-most common neurodegenerative disease, and over 10 million people worldwide suffer from the disease. It is estimated that 4% of individuals with Parkinson's disease are diagnosed prior to the age of 50. Parkinson's disease is a progressive condition that tends to worsen with age[56].

2.6.2 Alzheimer's Disease

Research and statistics show that the most common cause of dementia worldwide is Alzheimer's disease (AD) [57]. Dementia is a general term used to describe memory loss and other cognitive abilities necessary for everyday activities [58]. As a neurodegenerative disease, Alzheimer's Disease is mainly associated with aging; however, Alzheimer's Disease is not a normal part of aging. As Alzheimer's Disease progresses, symptoms will worsen. The current understanding of the pathophysiology of this disease is constantly evolving, as there is no cure for it, but there are treatments available to slow down the progression [58–61].

The symptoms of Alzheimer's disease include walking difficulties in patients[62]. In addition to walking slowly and irregularly, patients with these difficulties often find it challenging to negotiate turns, climb on a stepping stool, avoid obstacles in their path, or lie down and rise from the doctor's couch [63–66]. It has been reported that about 50% of Alzheimer's patients experience difficulty walking after being diagnosed about three years ago[62].

A report from the Alzheimer's Association for 2021 suggests that approximately 6.2 million Americans aged 65 and older suffer from Alzheimer's disease (AD). If medical breakthroughs do not develop to prevent, slow, or cure Alzheimer's disease by 2060, this number could increase to 13.8 million by then. There were 121,499 deaths reported from Alzheimer's disease in the latest available data, which makes it the sixth-leading cause of death in the United States, as well as the fifth-leading cause of death among Americans 65 and older based on official death certificates from 2019 [67]

2.6.3 Huntington Disease

Huntington disease (HD) is a progressive, incurable neurodegenerative disorder that presents with involuntary movements, dementia, and behavioral changes as a result of changes in the brain [68]. HD was named after George Huntington, the physician who first described it as hereditary chorea back in 1872 [69]. HD is characterized by gross atrophy of the caudate nucleus and putamen, along with selective neuronal loss and astrogliosis (an abnormal

increase in astrocytes), particularly in the neostriatum (part of the Basal Ganglia). In the cerebral cortex, there is also a marked loss of neurons [70]. The patients with HD often exhibit uncontrollable dancing movements (chloroform) and gait ataxia[71]. People with Huntington's disease are generally in their 30s and 40s and do not typically suffer from concomitant diseases or physiological changes associated with aging. Having impairments limited primarily to the central nervous system, HD provides a contrast to aging for studying the conditions necessary to correlate stride with interval [72]. According to research, HD prevalence varies based on ethnicity and geographical location, indicating that demographic differences shape the number and composition of individuals with HD. It is not surprising that populations of Northern European descent have the highest prevalence of HD in the world, as HD is believed to have originated in Northern Europe [73, 74]. Several studies examined the prevalence of HD throughout the world during the 1970s and 1980s. There were an estimated 4-7 people affected by the disease in Europe per 100,000 [75, 76]. According to the American national institute of neurological disorders and stroke (NIH), currently, more than 30,000 Americans have HD [77].

2.6.4 Amyotrophic Lateral Sclerosis

This disease is characterized by a progressive progression of symptoms that get worse over time. Amyotrophic Lateral Sclerosis (ALS) primarily affects the motoneurons in the brain stem, cerebral cortex, and spinal cord [78–80]. ALS has no cure at the moment, and there is no effective treatment that can slow or reverse its progression. ALS symptoms include muscle stiffness or weakness. As the disease progresses, individuals lose strength, speech, eating, movement, and even breathing capabilities. In most cases, people with ALS die due to respiratory failure, usually within three to five years of exhibiting the first signs of the disease. A small percentage of ALS patients survive for ten or more years [81]. ALS patients typically have abnormal gaits. The average walking velocity of the patients with this disease has been reduced and has been observed to be lower than that of the patients without it [82]. It is estimated that at least 16,000 people suffer from ALS in the United States at any given time, according to the ALS association [83]. There are an estimated 15 new cases each day. People with ALS are typically between the ages of 40 and 70, with an average age of 55. However, it is also common for individuals in their twenties and thirties to become afflicted with the disease. ALS affects men 20 percent more often than women. As people age, the incidence of ALS becomes more equal between men and women. On average, an ALS diagnosis takes about one year before a final diagnosis is made. Approximately 90

percent of ALS cases do not have a family history, and the remaining 10 percent occur as a result of a mutated gene.

2.6.5 Normal Pressure Hydrocephalus

Normal Pressure Hydrocephalus (NPH) is a disorder of the brain characterized by an accumulation of excessive cerebrospinal fluid (CSF) inside the brain's ventricles, which leads to impaired mental and physical functions, including difficulty walking, incontinence, and disorientation. Unlike the other diseases, this is not a neurodegenerative disease. The brain's ventricles, which are filled with fluid, accumulate excess cerebrospinal fluid when someone has NPH . The term "normal pressure hydrocephalus" refers to a condition in which the CSF pressure is normal despite the excess fluid in the brain. By enlarging brain ventricles, excess CSF disrupts and damages nearby brain tissue, leading to problems walking, thinking, and reasoning, and bladder control problems [84]. A patient with NPH may experience difficulty walking, which is sometimes compared to walking like a person on a boat, bent forward and with legs spread apart. Fortunately, it is one of the few causes of dementia that can be treated or reversed. A high-volume spinal tap is used to determine if an individual has the potential to benefit from surgically inserting a shunt if symptoms and results of an evaluation and magnetic resonance imaging (MRI) indicate normal pressure hydrocephalus. After removing a large amount of spinal fluid, doctors observe the individual for 30 to 60 minutes for signs of improvement in walking, thinking, and reasoning. The majority of people initially suspected of having normal pressure hydrocephalus do not improve after cerebrospinal fluid removal[84–86]. Hydrocephalus is primarily a disease of older adults, affecting nearly 700,000 adults in the USA. Neurological conditions, like Alzheimer's and Parkinson's, can be misdiagnosed as normal pressure hydrocephalus [87].

Relevance of the Diseases

The gait data in NPH, and AD are used in chapter 4, and chapter 5 of this study, while time-series based gait analysis in chapter 6 uses the data in PD,HD, and ALS.

Chapter 3

Technical Fundamentals

In this chapter, the basic concepts about machine learning (ML) and its types are given. Then the basic principles and different types of artificial neural networks (NNs) will be explained with more focus on deep learning and deep convolutional neural network (CNN). A brief explanation about the basic entropy measure Shannon entropy is also discussed as entropy based methods are used in this thesis for analysis of time series data of human gait, followed by a short description of statistical methods. At the end different performance measuring metrics are briefly explained.

3.1 Machine Learning

Machine Learning is a sub-field of artificial intelligence (AI) that provides computer systems the ability to automatically learn and adapt from experience by using algorithms and statistical models without its behavior explicitly being programmed [88–106]. There are three major types of machine learning: supervised learning, unsupervised learning, and reinforcement learning.

In supervised learning, the computer gives an example input given by the "teacher" and its expected result, to learn the universal instructions for mapping input to output. As a case, the input signal may be merely partly accessible to precise comments.

In unsupervised learning, tags are not given in a learning algorithm, leaving only the label to discover the structure in their input. Unsupervised learning can itself be a goal (a hidden pattern of finding data) or a means of achieving it (learning function).

As a behavior-driven approach to learning, reinforcement learning is the process of learning from mistakes. It is designed to learn from mistakes on the basis of reward and punishment. When the reinforcement learning model is put into any environment, it will

initially make a lot of mistakes, but will gradually learn to make fewer mistakes. Video games, resource management, industrial simulations, etc., are examples of systems where reinforcement learning is used. [93–98].

The utmost challenges and limitations in machine learning in biomedical engineering are to determine applicable data sets from the enormous amount of smart health data which are collected by data acquisition techniques. The gathered data is now to train to fulfill the essential needs of the application. For example, measuring (the movement of people, object features, and the improvement in health with the development) requires covering these dynamic features. Collecting and annotating large amounts of data has become a challenge.

Machine learning is additionally divided into diverse algorithms on the bases of varying problem perspectives. It includes clustering, instance base, and decision trees, which help in solving complex problems.[107]. There are many more algorithms of machine learning but our focus in this thesis is more on neural networks and especially on deep neural networks which are discussed in the following sections.

3.1.1 Artificial Neural Network

Artificial neural network is one of the most powerful predictive tools for analyzing data and making predictions from it. A neural network is made up of an interconnected group of nodes called neurons or perceptrons, which are connected through links. An artificial neuron is an information processor based on the behavior observed in biological neurons. It takes input, processes it, and then passes it through an activation function in order to generate activated output. The basic structure of an artificial neuron is presented in Figure 3.1. The Figure 3.1 shows that the neuron has inputs $x_1, x_2 \ldots\ldots, x_n$, each input has weight as $w_1, w_2 \ldots\ldots w_n$ which is used as the synaptic junction; $g(x,w)$ is a basically a summing function;$f(g)$ is an activation function that activates or deactivates based on the value of g; and y is the final output. The formula to calculate $g(x,w)$ is shown in Equation (3.1) and output y is calculated by applying activation function to $g(x,w)$ as Equation (3.2). The activation function is discussed in section 3.2.6

$$g(x,w) = b + x_1 w_1 + x_2 w_2 + \ldots\cdots + x_n w_n \tag{3.1}$$

$$y = f(g(x,w)) \tag{3.2}$$

A NN composed of many layers, each with a group of neurons. The term "fully connected layer" refers to a layer in which all the neurons from the previous layer are connected to all

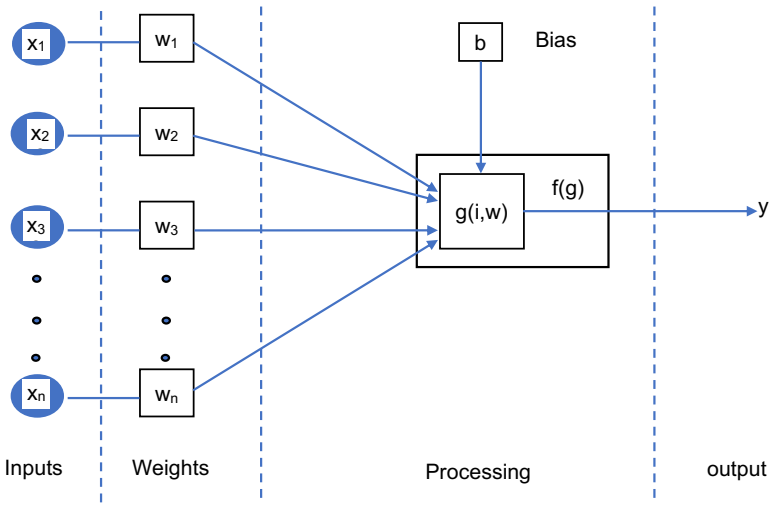


Figure 3.1: Representation of an artificial neuron; x are the input values; w are the weights for each input; b corresponds to the bias; g(x,w) is the summing input function; f(g) is the activation function; y is the output of the neuron

the neurons in the subsequent layer. It is not mandatory that all neurons within one layer must be connected to a neuron of the next layer. Commonly an artificial neural network has at least three layers: input, hidden (may be one or more) and output layer. [96, 108–114]. Figure 3.2 represents the basic structure of an artificial NN.

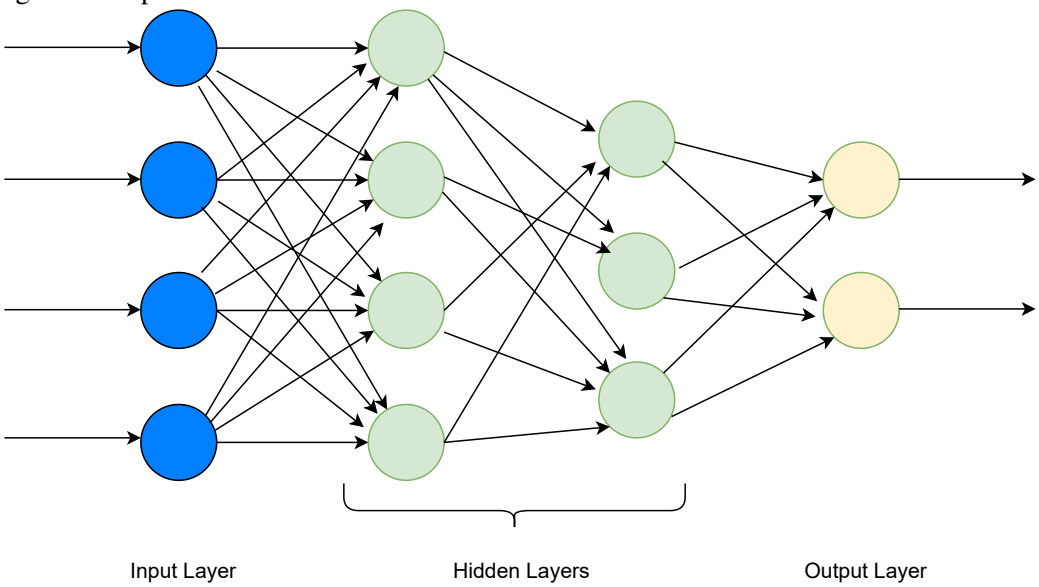


Figure 3.2: Representation of a 3 layers feed forward neural network with one input layer (blue), two hidden layers (green), an output layer (yellow). Input layer is fully connected to the next hidden layer.

Clustering and classification can both be accomplished using artificial neural networks. These networks work as classifiers when they are provided with labeled training data on which to train. If labeled data is not given to the network, they will perform clustering to group the data in different clusters depending upon similarities in the data. This three-layer network is also called a shallow network. When the number of hidden layers between the input and output layers are increased, then it is called a deep network.

3.1.2 Deep Neural Networks

A deep neural network is an artificial NN with many hidden layers between the input and output layers. It can be considered a stacked neural network in which multiple layers are stacked. There are single input and output layers and multiple hidden layers in the deep neural network. [115–119]. Figure 3.3 represents the structure of the deep network.

Deep learning architecture has existed for many years. However, new graphics and

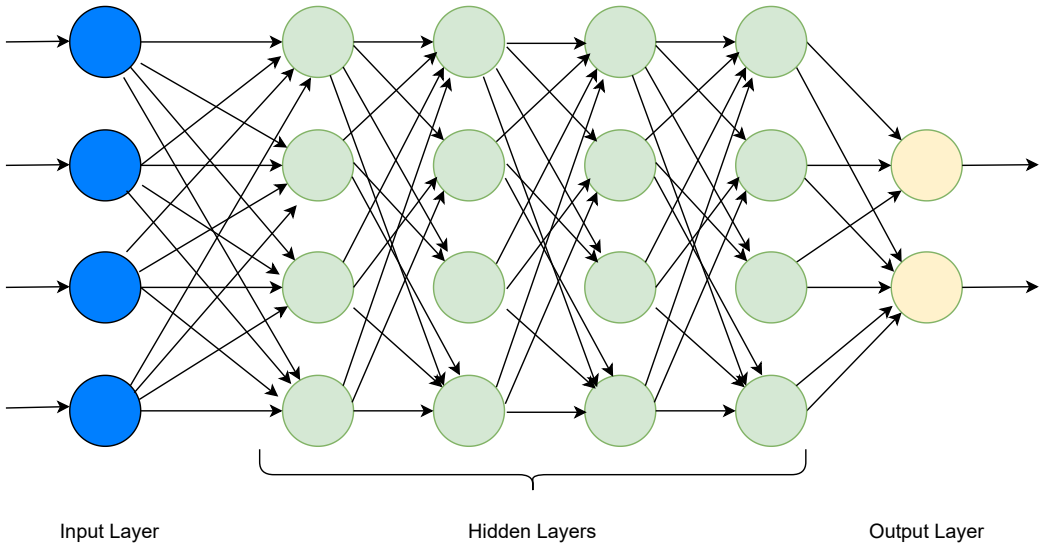


Figure 3.3: An example of a deep neural network having one input layer (blue), multiple hidden layers (green), and an output layer (yellow). Input layer is fully connected to the next hidden layer and some of the hidden layers are also fully connected. It is a classic example of a feed forward deep neural network.

architecture processing units put it at the front position of artificial intelligence. Deep Learning methods are largely used in applications such as image recognition, object tracking, language processing, and information retrieval [120–126].

3.1.3 Types of Deep Learning Architecture

Three different kinds of deep learning architectures are explained here. Amongst all the deep learning architectures, CNNs, recurrent neural network s (RNNs), deep belief networks (DBNs), long short term memory networks (LSTMs), and gated recurrent units (GRUs) are the best part of the history of deep learning. Deep learning is not a single approach, but it's a class or a group of algorithms and topologies that one can apply to a broad spectrum of machine learning and artificial intelligence problems. Moreover CNNs, RNNs, LSTMs and other deep learning architectures, including deep stacking networks (DSNs) are specially used for the classification and regression in the field of image recognition and natural language processing. [125–129]. Our focus in this thesis is mainly on CNNs, RNNs, and LSTMs, that are explained in the following sections.

3.2 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a type of deep neural networks that are designed explicitly for multi-array sequential or grid-like topological data and have been successfully implemented for numerous applications such as computer vision, natural language processing, time-series forecasting, speech recognition, and others [130, 131]. Deep CNNs are made up of cascade multi-level trainable layers, primarily corresponding to feature generation and detection stages. Deep CNNs, contrary to dense networks, have a width, height, and depth and replaced a single unit neuron with a convolution kernel. Deep CNNs are represented by a multi-layered hierarchical structure. The most common layers are convolutional, pooling, and fully connected layers. These layers are discussed in detail in the following sections. The terms CNN and Deep CNNs are used interchangeably in literature. The architecture and learning stages of CNN are shown in Figure 3.4.

3.2.1 CNN Architectural Representation and Training

Training of CNN consists of feed-forward and backward stages. The input is assigned to CNN. CNN exploits the learnable kernels with certain parameters on input and extracts a set of features known as feature maps. The feature maps are propagated in a feed-forward fashion by passing through a stack of layers, and at the end, output is estimated. The predicted output is compared with the ground truth, and the error is computed based on the difference. The backward stage computes the gradient via chain rule by backpropagating the error layer

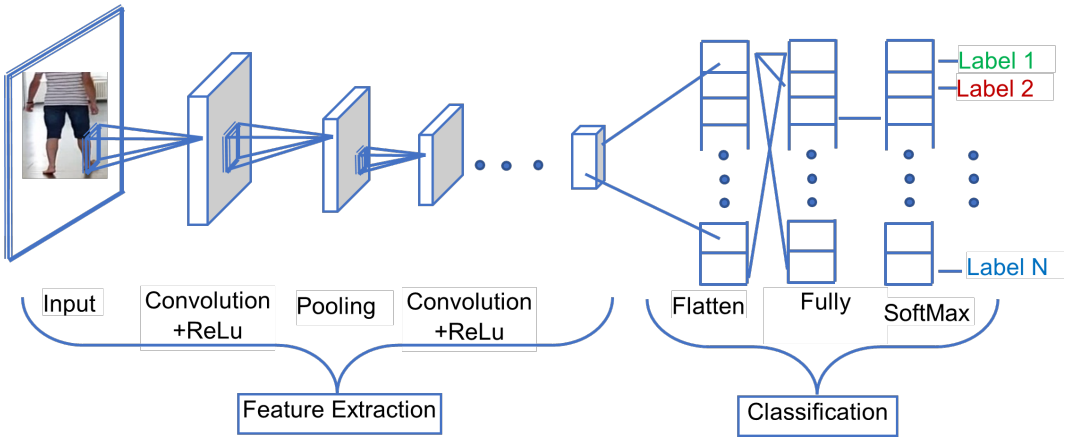


Figure 3.4: Illustrates the schematic procedure of a CNN from left to right. There is an input image, then multiple single regions from the image are extracted by convolutional layer which is passed to the pooling layer where sub-sampling is performed and this process is performed multiple times in the following layers. At the end a flatten layer with fully connected dense layer is used to normalize and then classification is performed.

by layer. The weights are updated based on a gradient and are used in the next phase of the feed-forward stage. The network is iterated multiple times through a learning phase until it is learned sufficiently, and loss is reduced to a certain amount [131]. The characteristic building units of CNN are discussed below.

3.2.2 Convolutional Layer

The convolutional layer is made up of multiple feature extraction modules known as convolutional kernels. The convolution operation explores the local area of the image known as receptive fields. It mines the patterns by multiplying the weights of the kernel with the receptive field and sliding across the image[132]. The convolution operation is represented in Equation (3.3)

$$F_{(c,u,v)} = \sum_{c,u,v}^{C,U,V} \sum_{i,j}^{I,J} I_{(c,u+i,v+j)} \cdot k_{(i,j)} \tag{3.3}$$

In Equation (3.3) $I_{(c,u+i,v+j)}$, is the element of input image (I) at location $(u+i, v+j)$, whereas $k_{(i,j)}$is the value of kernel that is centered at (u,v) index. Image shows local contextual relation, thus only neighboring values are taken into account, while it is not possible to randomly correlate distant pixels. These feature motifs are the basis of objects or patterns and can occur at different positions in the image. Thereby, keeping in mind this characteristic, the image is convolved with the same convolution kernel. The exploitation of

the local region instead of a complete image by the convolution kernel reduces the parameters and returns the position-invariant feature representation [133].

3.2.3 Pooling Layer

Pooling is also a local operation like convolution. It summarizes the semantic information of the local receptive field by a single value statistic. Pooling operation typically employs the average or max operation. As features are location invariant in motifs, therefore, exact location does not matter, and detection can be done by coarse-graining the position [134]. This makes the feature representation translational and positional invariant thus, prevents the neural network from memorizing. Additionally, it also regulates the complexity of the network by down-sampling the spatial dimension [135]. The pooling operation is discussed in section 4.5.1 of the thesis.

3.2.4 Fully Connected Layer

Fully connected layers in CNNs are used for the decision-making process and are placed after the convolutional and pooling layers (feature-extraction stage) [136]. Contrary to the local operation, a fully connected layer globally analyses the contribution of all the features that are learned from the feature extraction stage [137]. These layers help in the learning and extraction of class-specific discriminating features. A fully connected layer makes dense connections, and its mathematical expression is shown in Equation (3.4)

$$d_i = \sum_{n}^{N} \sum_{i}^{I} o_n \cdot w_i \tag{3.4}$$

In Equation (3.4), w_i is the i^{th} neuron of a fully connected layer that is linked with n^{th} input unit o_n , whereas d_i is the resultant connection.

3.2.5 Batch Normalization

Batch normalization describes a a post processing step which can be included after one or multiple layers inside the network. It acts as a regularizing factor that standardizes the output of each layer by performing batch based averaging. It addresses the problem of change in internal covariance of feature-map that occurs after multiplying the convolution kernel with feature-maps. Batch normalization improves the convergence rate of the deep neural network

[138]. Batch normalization is shown in Equation (3.5)

$$F_{BN} = \frac{F_c - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \tag{3.5}$$

In Equation (3.5), F_{BN} is the convolved feature-map, and μ_B and σ_B^2 are the batched based mean and variance of a feature-map, respectively. Likewise, ε is added to avoid the division by zero.

3.2.6 Activation Function

The activation function acts as a transformation function that helps to learn the non-linearity of the complex patterns. It maps the features within a specified range and decides which neuron to fire or not. The choice of activation function has a significant impact on the learning and training of neural network [139]. Initially, sigmoid (expressed in Equation (3.6)) is used as a non-linear mapping function; however, it suffers from a gradient saturation problem and is computationally expensive. Nowadays most commonly used activation functions are ReLU (Equation (3.7)) and LeakyReLU (Equation (3.8)).

$$y_{(u,v,c)} = \frac{1}{1 + e^{-F_{(u,v,c)}}} \tag{3.6}$$

$$y_{(u,v,c)} = max(F_{(u,v,c)}, 0) \tag{3.7}$$

$$y_{(u,v,c)} = max(F_{(u,v,c)}, 0) + \lambda min(F_{(u,v,c)}, 0) \tag{3.8}$$

Equation (3.6) shows that the sigmoid function maps the input in a range between 0-1 using the exponential function. Unlike sigmoid, ReLU is a piece wise linear function that is zero for negative value while retaining any positive values. LeakyReLU solves the dead neuron problem and averts the gradient from zero by assigning a small weight to negative values.

3.2.7 Dropout

Dropout is also used as a regularizing factor to control overfitting. It works by randomly dropping some connection with a certain probability which introduces noise into the neural network. Thus, it encourages each neuron to learn distinct transformation and reduces the chances of co-adaptation [140].

3.2.8 Hyperparameters

The network parameters are the weights and biases which are learned by the network. Hyper-parameters, therefore, describe the parameters to effectively adjust the model parameters. CNNs have a large number of hyperparameters, including several layers and filters, size of filters, stride, activation function, learning rate, optimization strategy, etc. The number of neurons and layers extract a diverse set of features at different levels of abstractions. Likewise, different sizes of filters extract features at a different level of granularity. Optimization and scheduling strategy have an effect on the convergence rate [141].

3.2.9 State-of-the-art CNNs

There are already many CNNs exist which are considered to be as state of the art networks for many problems. Some of the latest CNN architectures are discussed below.

3.2.9.1 InceptionNet

Inception Nets replaced the conventional convolutional layer with the block architecture in the CNNs (Figure 3.5). The introduced block is named as inception block and is based on the idea of split-transform-merge [142]. The inception block works by dividing the input into multiple paths and applying the convolution operation at different spatial scales (1x1, 3x3, and 5x5), whereby at the end, merge the multi-resolution transformation using depth-wise concatenation. The proposed block encapsulates both fine and coarse grain information. The extraction of features on different scales enables to deal with image on different resolutions. GoogleNet reduced the computation cost associated with deep networks by using a 1x1 convolutional operation before large-sized filters and for depth regulation. In advanced versions, Inception Net replaced the symmetric filters (5x5 and 7x7) with asymmetric convolution (5x1 and 1x5) to reduce the computation [143]. However, the heterogeneous topology of blocks makes it challenging to customize the output of multi-path before concatenation.

3.2.9.2 ResNet

ResNet proposed the idea of shortcut links to address the problem of vanishing gradient associated with deep neural networks. It used a shortcut link to assign the output of the previous block to the next block output by skipping some of the intermediate layers [144].

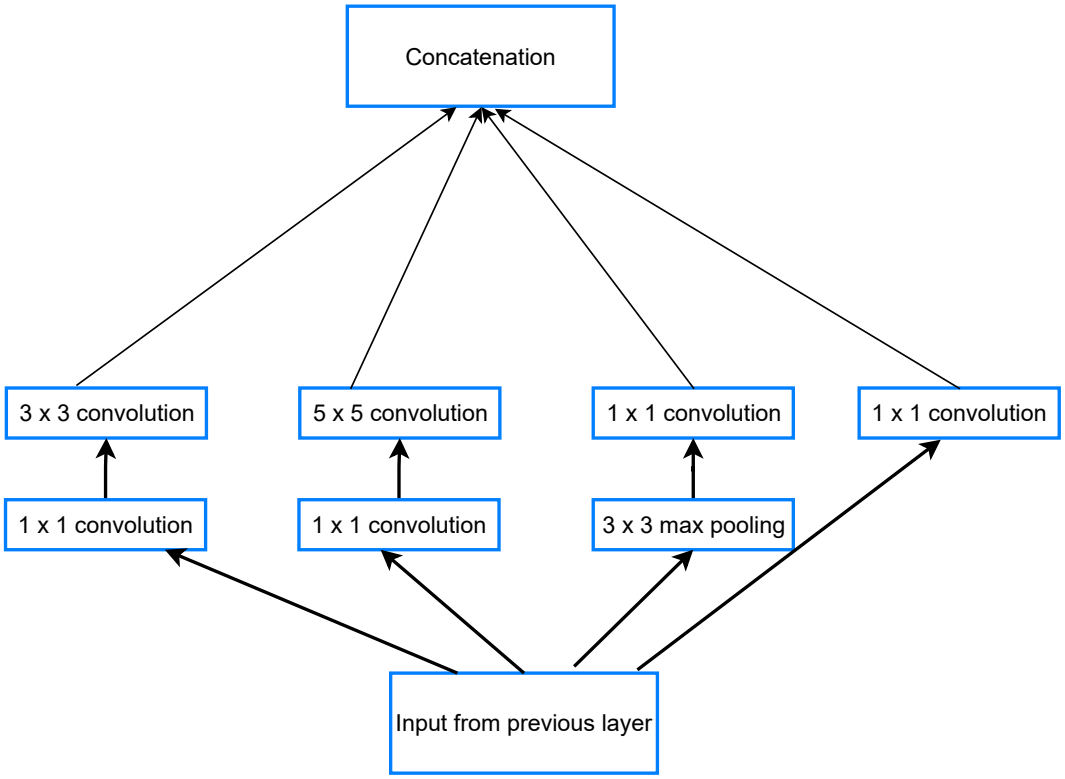


Figure 3.5: Illustrates the schematic diagram of Inception net. Figure inspired by [131].

Cross-layer connected blocks are known as residual blocks and perform reference-based optimization (Figure 3.6). Thus, residual learning encourages each block to learn new valuable features and improves the convergence rate.

$$f_m(F) = f_t(F,k) + F \tag{3.9}$$

$$f_t(F,k) = f_m(F) - F \tag{3.10}$$

In Equation (3.9),F is the input that is added to the transformed output f(F,k). Whereas, $f_t(F,k)$ in Equation (3.10) is a residual function that learns the new mapping.

3.2.9.3 WideResNet

WideResNet focuses on increasing the width of the network rather than the depth [145]. WideResNet solves the feature reuse problem of the deep Networks and emphasises that the main improvement in network performance is because of residual links, and depth has a supplementary effect [146]. They regulate the width of the network by introducing a new

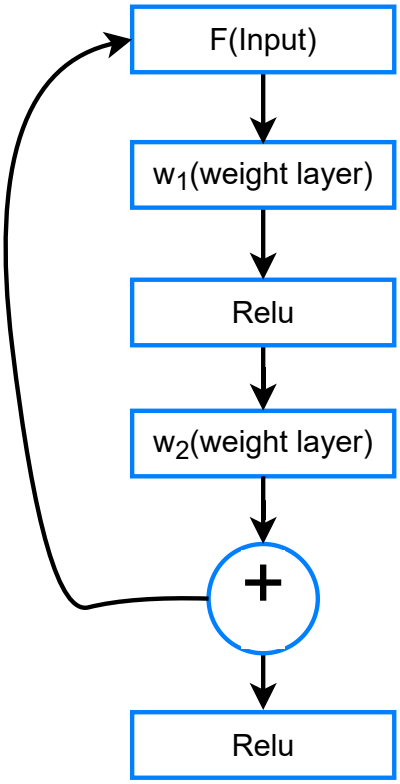


Figure 3.6: Illustrates the schematic diagram of ResNet. Figure inspired by [131].

hyper-parameter called cardinality. They also incorporated skip connects between the blocks like ResNet but with reduced network depth.

3.2.9.4 ResNext

ResNeXt also exploited the idea of split-transform-merge but in a simplified way [147]. ResNext modified the inception block by using convolution filters of the same spatial size at each path and merging the output of multiple paths through addition rather than concatenation operation (Figure 3.7). They introduced a new term, cardinality, that defines the number of transforming paths within each layer. In addition to this, they also incorporated residual links to smoothen the convergence and address the vanishing gradient problem.

3.2.9.5 DenseNet

DenseNet employs cross-layer connectivity to address the feature reuse problem. They ensure feed-forward feature propagation and multi-level information at each level by connecting

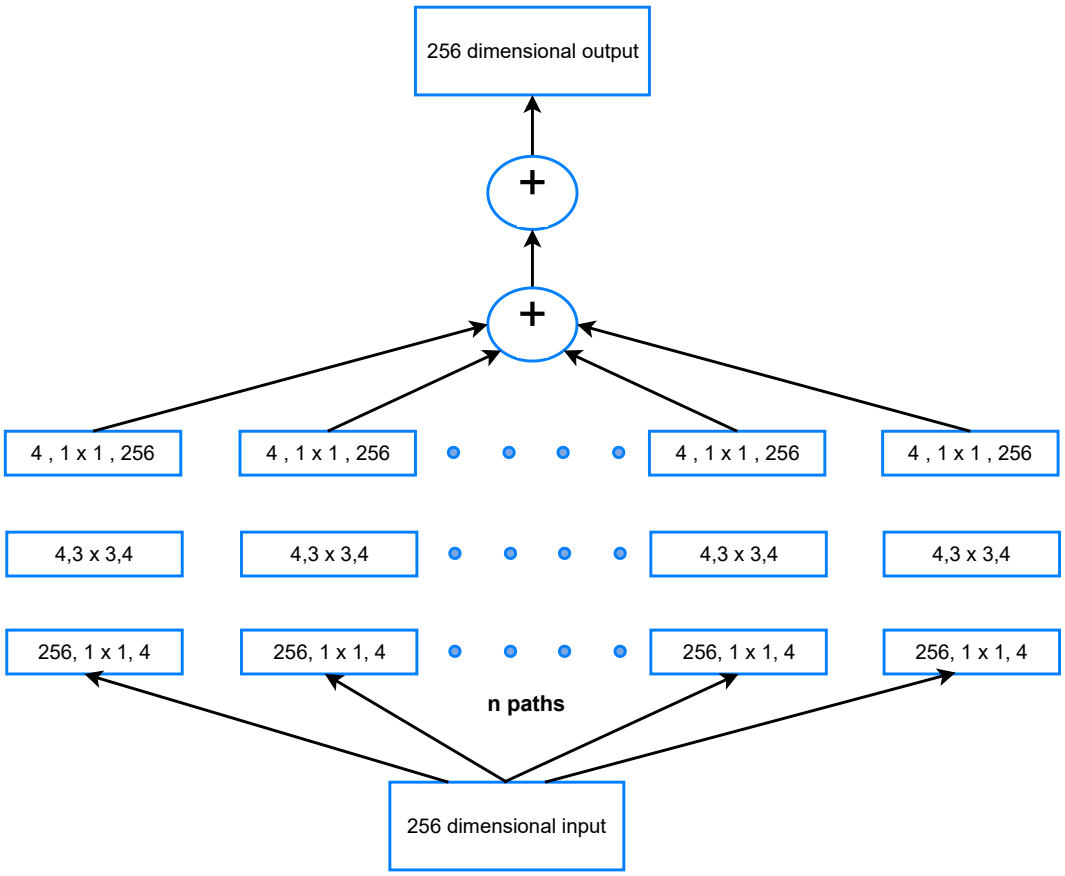


Figure 3.7: Illustrates the schematic diagram of ResNext. Figure inspired by [131].

each layer to the next layer (Figure 3.8). The concatenation operation explicitly preserves the information of the previous layer. DenseNet can become computationally expensive for deep architecture due to increased feature-map depth[148]. The shortcut connections make each layer accessible to the gradient that improves performance on a small size dataset.

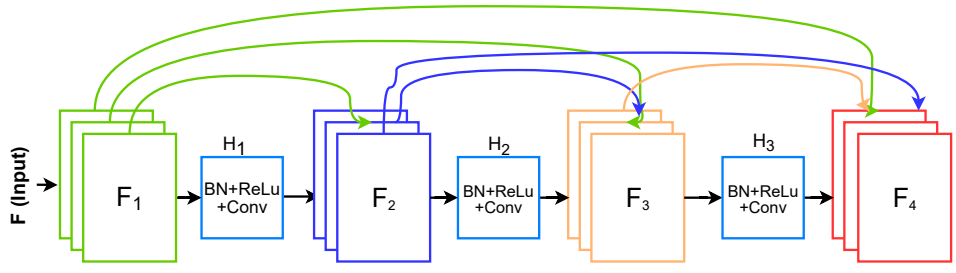


Figure 3.8: Illustrates the schematic diagram of DenseNet. Figure inspired by [131].

3.3 Recurrent Neural Networks

Normally neural networks work as taking input, processing it, and producing output. They lack the memory or storage element and cannot remember their previous states or output. Moreover, in the real world, there are many applications in which the system can't decide the correct next state until it doesn't have information about previous states or some context information. One of these application is the human gait, where the next steps are dependent upon the previous steps and there is a correlation among stride intervals. In such a cases time-agnostic networks may not perform good. Recurrent neural networks (RNN) are network with loops due to which they have persistence. Figure 3.9 shows the structure of RNN, while new state $h(t)$ is calculated as following.

$$h(t) = f_c(h(t-1), x(t)) \tag{3.11}$$

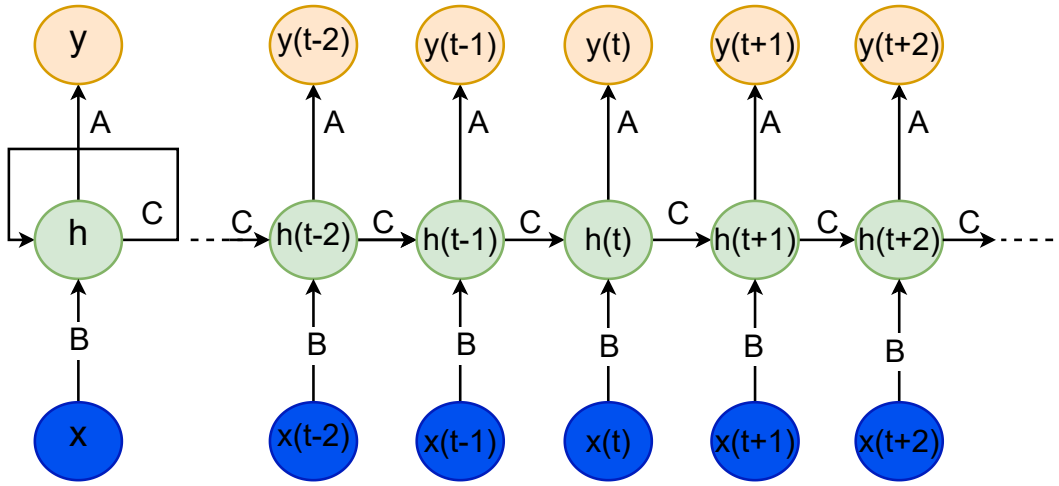


Figure 3.9: Illustrates the schematic diagram of RNN; $x(t)$ is the is the input vector at time t, $h(t)$ is the new state and $h(t-1)$ is the old state, $y(t)$ is the is the output of the network at time t.

The RNN can be viewed as multiple copies of the same network passing messages from one state to the next. Because these networks can transfer messages between states, they are considered to have memory elements. In RNNs, the temporally encoded information or features from previous inputs are also used along with the normal input to this layer. Reverse connections to the preceding layer or feedback loops to the current layer extend the neural network structure to make it recurrent. RNNs also allow for data extraction from the input history. However, it is pertinent to note that the time interval in which previously

seen information is considered is limited. With each step, the influence of the current input decreases, which means that information that has existed for a longer period of time is given less consideration [149].

3.3.1 Long Short-Term Memory

Recurrent neural networks can have memory elements for the current (hence "short term") or previous states. LSTM is a type of RNN that is used to handle long-term dependencies in ways that are not possible for normal neural networks automatically. RNN is an appropriate candidate for such applications in which the information is relatively close to the point where it is required. For situations where the gap is large, RNN is unable to handle such applications. For these situations, the RNN's variant LSTM has been developed, which has produced outstanding results [88, 96, 108–113, 115–118].

The LSTM has proven to be extremely effective in different fields like speech generation, image captioning, translations, etc. In addition to handling long-term dependency issues, it can recall information for an extended period of time and retrieve it when needed. The repeating network in LSTM consists of four layers linked together, as shown in Figure 3.10. The repeating modules are referred to as cells. Within a cell, the state, known as the cell state, plays an instrumental role. The Figure 3.10 illustrates that each cell takes some inputs x and generates an output h based on x. These functions will be discussed extensively in the following sections. A cell's state is linked with the state of other cells, and it is up to the cell

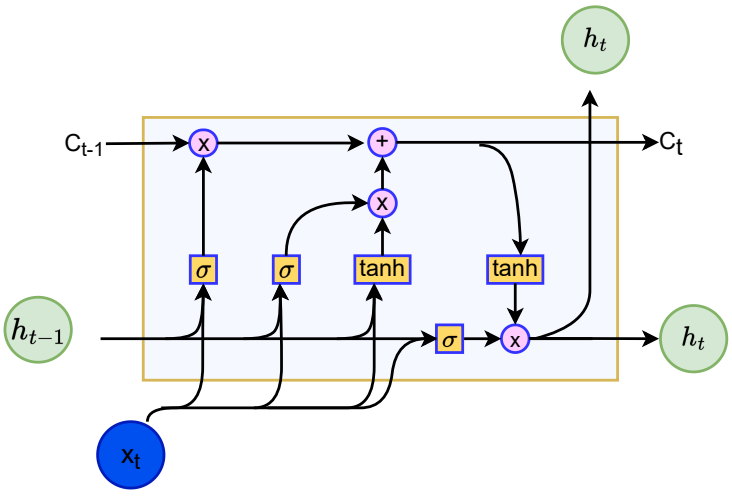


Figure 3.10: Illustrates the schematic diagram of LSTM. C_{t-1} is the previous cell which is connected to this cell, h_{t-1} is the output of the previous cell based on input x_{t-1}; x_t is current input; h_t is the output of this cell. These cells are stacked to make complete LSTM. Picture reproduced from [150]

whether it passes on the newly generated cell state to the next cell or the previous cell state. A yellow rectangle indicates the layers within the cell. The first layer is a sigmoid function layer, which allows the previous state of the cell to be passed through to the current state or stops it. It makes this decision based on the current input x_t and the previous cell output h_{t-1}. Based on the inputs, the sigmoid function will generate a value ranging from 0 to 1, which is then multiplied by the previous cell state to determine whether it is allowed or blocked. The functionality of this layer is represented in Equation (3.12), where f_t is the output of the layer, and b is the bias.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{3.12}$$

The next step is to determine what new information should be added to the state of the cell. This is determined by the two following layers, which work together to determine what information should be added to the current cell's state. The first layer uses the sigmoid function and the second layer uses the *tanh* function, as shown in Equation (3.13) and Equation (3.14), where i_t is the sigmoid output and $\hat{C}_t$ is the *tanh* output.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{3.13}$$

$$\hat{C}_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_C) \tag{3.14}$$

The second and third layers update the old cell state C_{t-1} to a new cell state C_t. The sigmoid function (second layer) determines how much to update the current state. As shown in Equation (3.15), the third layer (*tanh* function) generates the vector of candidate values for C_t.

$$C_t = f_t * C_{t-1} + i_t * \hat{C}_t \tag{3.15}$$

Finally, the fourth layer specifies what is to be output. In other words, the sigmoid function determines which part of the cell state will generate output, while the *tanh* function determines which results to output. Equation (3.16) and Equation (3.17) depict the functions of the fourth layer inside the cell.

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \tag{3.16}$$

$$h_t = o_t * tanh(C_t) \tag{3.17}$$

As all cells participate in decision-making processes by looking at previous messages stored in them, LSTM takes the previous cell state and decides whether to keep it or generate a new one. There are many applications for LSTMs, such as speech recognition, time series prediction, music composition, robot control, action recognition, and sign language translation.

3.4 Transfer Learning

A deep CNN is a powerful representational learning algorithm capable of dealing with complex medical images. However, deep CNNs require a large number of datasets for training due to their parametric and computational complexity. They get over-fitted on a small dataset and their performance suffers. A small number of medical images can be used to train highly parameterized deep CNNs using transfer learning (TL).[151, 152].

TL is a type of ML that allows transferring or reusing existing knowledge for some new tasks. It overcomes the limitation of conventional ML models that requires training from scratch for a new task. Additionally, in conventional ML models, the use of the trained model for some new data or problem requires it to belong to the same distribution as the training data [153]. Due to the limited number of data available in our study, there was a chance of overfitting, therefore, transfer learning approaches were also used to overcome the problem of over-fitting and make the model more generalized. In the following sections, some important concepts about the transfer learning are presented.

3.4.1 Pre-trained Models as Feature Extractors

Deep learning models are hierarchical in design and learn the different levels of features. It can be shown that the first layers of CNN learn to detect generic features such as lines, edges, and colors regardless of the specific types of images and across different domains. The intermediate layers learn the patterns or parts of objects that are made up of generic features. Higher layers in CNN extract the high-level abstractions that are task-specific. These feature extraction stages are connected to fully connected layers that are used for the decision-making process. The multi-layered structure of deep CNNs makes it a good feature extraction module that can be used to extract both low and high levels features [141]. For this reason, the final fully connected layers of pretrained deep CNN is removed, and the gradient is set to zero to use it as a generic feature extractor. Layers can be removed at different levels; thus, different type of features can be extracted. The learned representations are assigned to a shallow ML model for the decision-making process[154].

In TL, a pretrained model is fine-tuned to adapt to the new task. The last fully connected layer is modified, and a new task-specific layer is added. The newly added layer is trained on the target data to learn domain-specific features and adjusted to the number of classes to be distinguished between for a classification problem. In some of the cases, knowledge of the entire network is utilized by fine-tuning all the layers. Likewise, in some cases, selective

layers of the network are fixed, while other layers are updated during training. The pretrained models provide a good set of weights; thus, the fine tuned model converges to better local minima in a short time [155].

3.4.2 Advantages of the Transfer Learning

Transfer learning has showed a significant contribution in the performance boosting of neural networks. The advantages of the transfer learning in deep learning model are discussed below [131, 154].

- **Better Convergence**
  Provides a good initial set of weights, which improves the convergence and performance of the network overall.
- **Smaller Training Dataset**
  In medical field it is always the problem of sufficient data. Like in this thesis, where a limited number of data is available. TL addresses the problem of the large dataset required for the training of deep networks from scratch. The initialization of the model's weights from a pretrained network reduces the chances of overfitting in a deep network.
- **Reduced Training Time**
  When a new task is being performed, the use of pre-trained networks can often be a valuable tool in expediting the training of the model.
- **Improved Generalization**
  TL can improve the model performance on unseen data. They are adapted from one domain to another and learn a valuable set of features that can be applied to different contexts.

The transfer learning techniques are useful in studies like this thesis, where the dataset is relatively small. In this case, the pre-trained weights help to properly extract relevant features from the images. TL is used in chapter 5 of this thesis, where it helped to improve the accuracy of the classification of feet position.

3.5 Data Augmentation

In addition to transfer learning, data augmentation is another option that can be used to avoid overfitting on small datasets. A dataset is artificially extended by data augmentation.

Creating such data is relatively easy for some machine-learning tasks. It is possible to generate a "new" image for image classification using CNN, for example, by simply rotating, scaling, and adding noise to the image. As the local information is swapped for a CNN, the transformed image is entirely different, and so the input values differ as well. It is already seen in Section 3.2 that CNNs attempt to extract semantic information from the input images. Rotating an image shifts its semantic information in the image which causes other neurons to fire to extract this information. It has been demonstrated that dataset augmentation techniques can significantly reduce the generalization error of machine learning leading to a more robust classifier. [114].

Different data augmentation techniques are applied to gait data in this thesis and are discussed in chapter 4 of this thesis.

3.6 Shannon entropy

The basic entropy measure is Shannon entropy (SEn), which has its root in information theory. A variable's Shannon entropy represents how much uncertainty, or information, is inherent in the possible outcomes of the variable on average. Shannon entropy is given by the formula:

$$H = -\sum p_i log_b p_i \tag{3.18}$$

In Equation (3.18), p_i is the probability of value i to be in the stream of values of the time series (in our case it is stride interval duration), and b is the base of the logarithm used. Normally logarithm of base 2 is used when the corresponding unit of entropy is in the bits [156, 157]. Other entropy-based methods used in this thesis will be discussed in the respective chapter(s).

3.7 Statistics

The significance of observed differences among different categories of data can be determined using statistical tests. Statistics play a critical role in the planning of clinical trials and laboratory experiments in biomedical research. The purpose at this stage is to determine the design and size of an experiment that provides a reasonable chance of detecting clinical or scientific effects. Thereafter, statistics is again used to get relevant information from the results to make a decision about the clinical condition of the patient. In this work, simple

statistical measures like mean, standard deviations, variance, confidence intervals, P-values, standard errors, etc. are used[158].

3.8 Evaluation Measures

To measure and evaluate the performance of the different machine learning algorithms, multiple metrics were used in this thesis. Some typical procedures including accuracy, precision, F-measure, etc were selected. Here are the mathematical formulations for these factors.

3.8.1 Confusion Matrix

A confusion matrix is an appropriate tool for assessing the performance of a classifier. The results predicted by the classifiers are presented in a tabular form that splits the precise prediction of class from unfitting predictions [159]. It tells the correct and incorrect predictions. Other routine measures like accurateness, exactness, remembrance and F-measure can be calculated by means of this matrix. Confusion matrix is represented in below in Figure 3.11. The four cells of matrix show true positives (TP), true negatives (TN), false positives (FP)

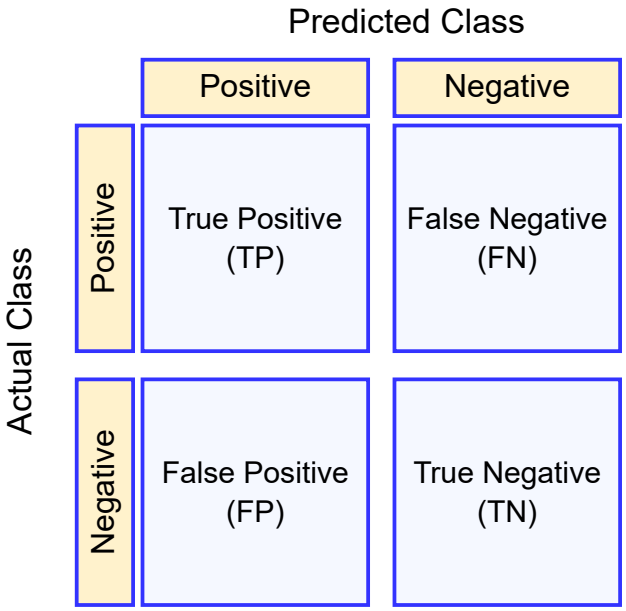


Figure 3.11: Confusion Matrix showing the relation between actual and predicted classes in different scenarios

and false negatives (FN).

- **TP** = Amount of predictions that are properly classified by the classifier as confident.
- **TN** = Amount of predictions that are properly classified by the classifier as undesirable.
- **FP** = Amount of predictions that are incorrectly classified by the classifier as confident.
- **FN** = Amount of predictions that are incorrectly classified by the classifier as undesirable.

Multiple metrics can be obtained from the confusion matrix, which can quantify how well the classification performed regarding different aspects.

- **Accuracy**

  Accuracy is the element of occurrences that are properly categorized separated by the aggregate amount of occurrences. It can be given as:

$$\text{Accuracy} = \frac{\text{Correctly classified instances (TP+TN)}}{\text{Total No. of instances (TP+TN+FP+FN)}} \tag{3.19}$$

- **Precision**

  Precision is the fraction of number of precise calculations by the complete predictions. It calculates the fraction of instances that are truly positive. In relations of likelihood, precision is the possibility that an instance is correctly classified. It is also called positive predictive value (PPV). In terms of confusion matrix, it can be measured as:

$$\text{Precision} = \frac{TP}{TP+FP} \tag{3.20}$$

- **Recall**

  Recall is the degree of the segment of optimistic occurrences that were properly categorized. It is also called Sensitivity or true positive rate (TPR). In terms of confusion matrix, it can be measured as:

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3.21}$$

- **F-Measure**

  F-Measure is the vocal mean of accuracy and recall. It delivers a equilibrium amongst precision and recall and uses mutually to compute a performance degree. Its formulation is specified as:

$$\text{F-Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision + Recall}} \tag{3.22}$$

- **Negative Predictive Value**

  Negative predictive value (NPV) is the ratio of true negatives to all the negatives in the predicted class.

$$NPV = \frac{TN}{TN + FN} \tag{3.23}$$

For measuring the classifier performance, correctness single-handily is not a suitable measure. Additional metrics can also be used to evaluate the performance of the classifier. For more information on these metrics, the reader is referred to the work by Singh et al. [160].

PART II

VIDEO BASED GAIT ANALYSIS

Chapter 4

Classification of Feet' Positions from Videos Using Healthy and Normal Pressure Hydrocephalus (NPH) Gait Data

This chapter presents the basic study in which deep learning methods were used for the classification of feet' positions with respect to the ground, from the videos recorded outside laboratory conditions, and without any marker on the body of the subject. The motivation and objective of the study are discussed in the introduction section, then related works are presented. After that, the data collection setup and the data preparation steps are discussed. In the next section, different deep learning models and their training strategies are discussed. Finally, the results in terms of the performance of the different classifiers in classifying the foot's position, are presented with the discussion and conclusion.

4.1 Introduction

For centuries, scientists and clinicians have studied the walking patterns of humans and other animals (A detailed history of gait analysis is given by Baker [161]). From Borelli's staggered poles used to study his own gait [162] to modern tools such as three-dimensional motion capture, gait mats with instrumentation, inertial sensors, and wearable devices, gait analysis has evolved into an exciting field of research. Despite technological advances, we continue to encounter limitations when it comes to measuring human walking in clinical and laboratory settings. It remains expensive, time-consuming, and difficult to access specialized

equipment and expertise when using current techniques [163]. Gait analysis is the process of quantitatively or qualitatively analyzing different parameters of the gait individually and in groups to determine the state of the health or recovery. Getting these parameters correctly requires a sophisticated laboratory setup and technical expertise. Once these parameters are correctly quantified, one can get insight into the walking patterns of human gait. Therefore, clinicians and researchers are trying to develop such methods for gait analysis that can not only give the correct quantification but are also simple in use and able to record long-term gait data. With the technological advancement and availability of high computational power and modern graphic processing units, deep learning has been shown to provide excellent results in the analysis and recognition of medical images and medical events in recent years. In the field of diagnostics and clinical everyday routines, deep learning is a promising method that can be applied to support physicians. At present, deep learning applications are already being used by doctors in a variety of ways, including diagnosis [164], image registration [165], multi-modal image analysis [166], and image segmentation [167].

In contrast to traditional machine learning methods, deep learning-based methods do not require additional feature extraction or generation in order to classify data, because the model generates features from data on its own and classifies them based on the generated features. Due to its versatility, it can also be applied to complex nonlinear classification problems or expanded to include multiple input channels, such as videos, for analysis[168].

There are several types of deep learning networks, but arguably the most common network in the field of image and video analysis is the convolutional neural network (CNN) [169]. Using multiple convolutional kernels, CNNs are designed to extract information from an image.

As gait parameters cannot be quantified from a single image because gait is a periodic process in which the left and right foot perform actions one after another in a cyclic way. Different gait phases inside this movement are important to deduct and one phase is dependent upon the others. If we take the most basic feature of human gait that is used in the majority of the studies i.e. stride interval (SI), then one stride interval is highly correlated to the previous stride interval(s). Hence CNN can only extract information from current input and it does not have information about the previous state, therefore alone CNN might not be suitable for gait analysis. Consequently, a combination of CNNs and recurrent neural network s (RNNs), such as long short term memory network (LSTM) network, can significantly improve recognition performance.

There are various pose estimation algorithms that can be used to track human movement and obtain body point information. These are based on a three-dimensional motion capture

system and require more than one camera for recording the movement, and some of them require specific markers on the human body.

As state-of-the-art, gait parameters are usually recorded under controlled conditions in a gait laboratory, using either wearable systems (explained in the following sections) or non-wearable systems using floor sensors or multiple cameras. On the other hand, it is known that patients move very consciously, and unnaturally under a strict laboratory environment. Therefore, there is a risk that the data collected in this way are subject to bias. Furthermore, there is also a need for flexible gait analysis methods that can be used, for example, in day-to-day hospital care.

Gait analysis methods require certain parameters and extraction of these parameters from frontal and postern views of the gait videos is a technical challenge that can be addressed by using the in-built power of deep neural networks (NNs). The deep NNs enable the feature extraction from the images of the walking. These features then can be further utilized for the classification of feet' position with respect to ground. As a result of this classification, we can find information about gait events such as heel strike (HS), and toe off (TO).

This leads us to the objective of this study, in which we aim for a marker-free video-based gait analysis under normal conditions using deep learning methods. We want to answer the question: are the deep learning methods able to classify the feet' positions directly from the walking videos that are captured using a low-resolution single smartphone camera in a free environment? And what is their performance in the classification of the foot's position?

4.2 Related Work and State-of-the-Art

The literature presents various wearable and non-wearable solutions for gait analysis and feature extraction in order to identify gait features. **Wearable Systems:** This method involves placing sensors at various body positions to measure gait. Some of the types of sensors that can be used are accelerometers, force sensors, gyroscopes, foot switches, extensometers, and inclinometers. **Non-wearable Systems:** These are non-invasive and they are not directly in contact with the body. Non-wearable systems are basically vision-based systems. The vision-based gait assessment can be conducted using either a video camera, thermal vision sensor, or depth camera. Alternatively, the environment-based gait assessment can be conducted using infrared and pressure sensors deployed on the floor[170]. Nevertheless, both such solutions require controlled research facilities for analysis, which limits their applicability to external/outdoor environments[171].

In the field of healthcare, as well as in the field of recognition, both gait analysis systems have been extensively used by many researchers for gait analysis. Camera-based systems are relatively new but accelerometers [172, 173], gyroscopes [174], Inertial Measurement Units (IMUs) [175] and force sensors [176] are widely used to measure gait characteristics [177]. For gait-based authentication, Derawi et al. [172] use a hip-worn accelerometer to measure cycle length. Similarly, [174] monitors trunk angle changes by attaching a gyroscope to the trunk for fall detection.

Wu et al. [178] attempt to accurately identify gait by observing participants walking 10 meters with a force plate at the foot. The goal of Chen et al. 's paper [179] is to achieve recognition of different human activities using sensors. The accelerometer used in this study is placed on a mobile phone and is used to collect data while subjects carry out eight different activities of daily life (ADL).

Using five inertial sensors attached to the lower body, Zebin et al. [180] aim to identify six common everyday activities. A variety of ADLs were identified in the study using CNN, including sitting, standing, walking, lying down, walking upstairs, and walking down. A study by Ordóñez et al. [181] examines the use of wearables to recognize the human activity. Through the use of this method, five different activities of daily living are recognized, in addition to sporadic right arm movements. A total of 7 IMU sensors and 12 accelerometers are placed on various parts of the human body to collect data.

In their paper, Camps et al. [182] explore the detection of freezing of gait (FoG) in Parkinson's Disease (PD) patients. A nine-channel signal is collected from subjects by an IMU during data acquisition. This paper by McGinnis et al. [183] aims to measure gait speed in participants suffering from neurological disorders based on data collected with accelerometers from two groups (sample and control).

Using the optimal combination of sensors, Steffan et al. [184] set out to identify stable and unstable body postures. A total of 34 possible sensor placements are tested during data acquisition. Additionally, a multi-marker motion capture system measures the normalized movement of different subjects. The vision-based solution of Hu et al. [185] provides a means for detecting FoG. In addition, [186, 187] and [188] are gait-based assessment solutions for Parkinson's Disease (PD), cerebral palsy, and a variety of chronic disease progression, respectively.

Recently Saleh et. al. [189] conducted a study entitled as analysis and best parameters selection for person recognition based on gait model using CNN and image augmentation. Shin[168] uses inertial sensor placed freely in human pocket for hemiplegic gait detection using deep learning methods .

Stenum [163] uses pose estimation for two-dimensional video-based analysis of human gait.

They use the publicly available data set recorded with multiple cameras and compare their results with three dimensional motion capture systems. In recent years, use of deep learning neural networks has increased for human action recognition problems. The top-performing models to solve this problem have been based on CNNs and RNNs[190–192]. By using public video datasets, these architectures can be used to learn representations based on raw video frames. However, these algorithms are complex and require large computational resources to train; furthermore, in some cases, other preprocessing steps, such as calculating optical flow between video frames, are necessary[190].

In an another recent study, authors proposed and validated a new marker-based gait event detection method in which they collected data from healthy young adults and assessed with other diseased and older adults during complex motor tasks. They achieved high sensitivity (SENS), positive predictive value (PPV), and F1 score[193].

A detailed systematic review of latest research trends in gait analysis using wearable sensors and machine learning is given by [170] recently. All of the reported studies either use wearable systems, floor sensors or multiple cameras and markers placed on the body with context awareness.

4.3 Data Collection Setup

Two rounds of data collection were performed to have a large enough dataset and in different conditions. In the first round of this study, gait data in normal and in Normal Pressure Hydrocephalus (NPH) were collected. (The second round of data collection is discussed in chapter 5 of the thesis). There are three classic symptoms of NPH: progressive gait apraxia, urinary incontinence, and dementia. The reader is referred to the section 2.6.5 of this thesis for further details about NPH. We selected this disease because we are expecting as the next step is the clinical application under clinical conditions. The first clinical use is already planned in collaboration with the Department of Neurosurgery of the Cologne Medical Center. In this study, a walking pathway suggested by [194] was used, inside the Institute of Biomedical Engineering (IBT), Karlsruhe Institute of Technology (KIT). It was predefined because we wanted to get some gait parameters by observation also. The volunteers were asked to walk with a normal, self-selected speed as well as by mimicking an abnormal gait pattern of typical patients of NPH. Performed movements were recorded by two cameras to use the video streams therein as input data for the algorithms.

A group of 20 subjects (14 males, 6 females, age: 28 ± 4 years, height: 175 ± 13 cm, and weight: 70 ± 12 kg) participated voluntarily in this research and reported no gait

abnormalities. Participants with any condition that might interfere with normal walking patterns, typically the falls within the preceding six months, were excluded from participation. Statistical information was also collected, including gender, height, weight, and age.

The gait characteristics of the subjects were assessed while walking on a 10-meter (m) walking pathway. The study consisted of filming the patient walking on a 20 m path (10 m in one direction, turning, and then 10 m back to the starting point) during four trials, totaling 80 m of the walking (Figure 4.1) . Subjects were asked to walk and were also given these instructions: "start from this point", "follow this path", "go to the location indicated", "stop suddenly", "turn around", and then "return to the starting point".

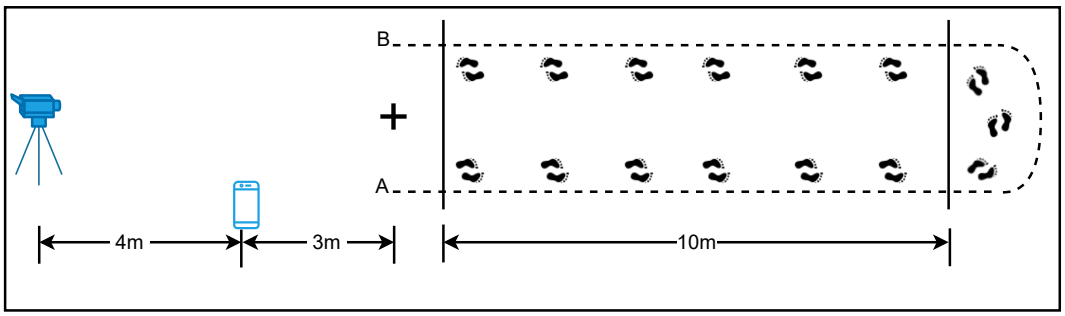


Figure 4.1: Illustration of the experiment setup. A and B are the start and end points respectively, and subject was standing at + sign before the start of walking. The mobile phone camera had been recording from the side of the corridor and 3m away from the starting area, whereas the high-quality camera had been in the middle and 7m away. Walking pathway inspired by [194]

Before the experiments, all participants were briefed regarding the purpose and method of the experiment. Recordings were taken by the experienced staff in the Institute for Biomedical Engineering at Karlsruhe Institute for Technology. All methods were performed in accordance with guidelines and regulations by the KIT Ethics' commission. Informed consent was also obtained from the participants to take part in this study.

The subjects were not asked to wear special clothes or equipment, but to attend with their casual clothing and shoes. Before the measurements, the subjects had trained with the slow walking task until they felt comfortable performing the desired gait pattern. During the measurements, subjects were instructed to perform two walking tasks. The first task was normal walking with a self-selected speed on a 10 m path, making a 180-degree turn, and returning back to the starting point with the same speed (Figure 4.1). The second task was to complete the same parkour with the pre-given gait pattern at a slow pace. Each subject performed each task twice, once with shoes and once with socks. The slow walking pattern was required to show large step width (>one-foot length), low step length (<one-foot length)

and toe-out position of the feet for NPH patients, as described by [194]. As this is a pilot study no other human or animal was allowed to be in the walking pathway while recording. For some of the subjects who were not able to emulate the desired gait in one attempt, the recording process was repeated.

The movements tasks performed by the subjects were recorded with a high-quality camera (Basler AG, Ahrensburg/Germany) with a sampling rate of 25 fps from the front view as well as with a mobile phone camera (Samsung Electronics GmbH, Schwalbach/Germany) from the front-sagittal view as illustrated schematically in Figure 4.1. The high-quality camera was used together with an objective (Kowa, Düsseldorf/Germany) having a focal length and a minimum working distance of 60 mm and 100 mm, respectively. The camera was fixed by means of a tripod, whereas the mobile phone was held as still as possible by a person standing on the side of the corridor Figure 4.1. The camera was connected to a laptop (Samsung Electronics GmbH, Schwalbach/Germany) by a proper USB cable to be able to use the software interface.

4.3.1 CASIA Dataset

Besides the collected data set, a publicly available gait database called CASIA-Dataset B was also used [195, 196]. This dataset is basically collected for gait recognition by the Institute of Automation, Chinese Academy of Sciences (CASIA) and they provide it to gait recognition and related researchers in order to promote the research. It is a large multi-view gait database (collected with multiple cameras). There are 124 subjects, and the gait data was captured from 11 views. Three variations, namely view angle, clothing and carrying condition changes, are separately considered in their dataset. We selected only 10 subjects because we wanted to avoid any overfitting of the model for the CASIA data set.

4.4 Data Preparation

Data preparation is an essential step in machine vision tasks and is widely used to support later processing stages for accurate predictions. The purpose of data preparation is to refine the information for an accurate match. Usually, these refinements are implemented as a component of data preparation and include data augmentation, data profiling, data preprocessing & filtering, and data cleansing. Therefore, three components of data preparation are opted here: 1) data preprocessing and filtering, 2) data annotation & labelling, and 3) data augmentation. Data annotation and labelling is the process of assigning labels/ground truth

to the data. Data augmentation is being utilized for equal distribution of data in different classes and data preprocessing is used to enhance the visual quality, conversion of videos to images, and resizing of images. The description of these steps is given in subsections.

4.4.1 Preprocessing and Filtering

After collecting all the videos, they were preprocessed for filtering because we wanted to keep the strict 10 meter walk (going in one direction and coming back) including taking the turn. It was required because some basic information was also collected by observation in order to validate the outputs of the neural network. Therefore walking outside the desired area was trimmed from videos. Similarly, the videos were also cropped to the size of 640x420, because there was some extra area (floor, ceiling, side walls) that was not relevant to the region of interest. Some basic statistical information was derived from the videos for each subject in all of the 4 cases, like time to complete the task, speed, and stop events. Detection of these parameters was accomplished by visual examination. Subsequently, the videos were converted into images at the frame rate of 30 fps to make them ready for input to a NN.

4.4.2 Data Annotation and Labelling

The aim of this research is to develop a gait analysis system without wearable systems or sensors and by only using the camera. Therefore, the determination of the ground truth for a particular image is very important before training a neural network. Images are annotated by assigning labels to them. A human operator analyzes a set of images, identifies relevant objects within each image, and annotates the image to indicate, for example, its shape and label. In order to train computer vision models, these annotations can be used. Human annotations serve as ground truth and the model uses them to learn to detect objects or label images on its own. In image classification, object recognition, and segmentation, this process can be applied.

Manual Annotation

The dataset originally contains a total of 86,232 images which are converted from videos of walking volunteers in two different walking conditions with two different protocols. To determine the ground truth or true label of these images, manual image annotation is performed. During walking human's feet position could be in one of three positions: both feet are on the ground, the left foot is in the air and the right foot is on the ground, and the left foot is on the ground and the right foot is in the air. Therefore, each image is assigned one of

the following labels.

- **LA-RG**: Left foot in Air-Right foot on the ground
- **LG-RA**: Left foot on the ground -Right foot in the air
- **LG-RG**: Left foot on ground-Right foot on the ground

Some of the examples of each class are shown in Figure 4.2. In the next step, each class is assigned a corresponding number because neural networks can only understand/process numbers. Numbers 0,1,2 are assigned to LA-RG, LG-RA, and LG-RG classes respectively.

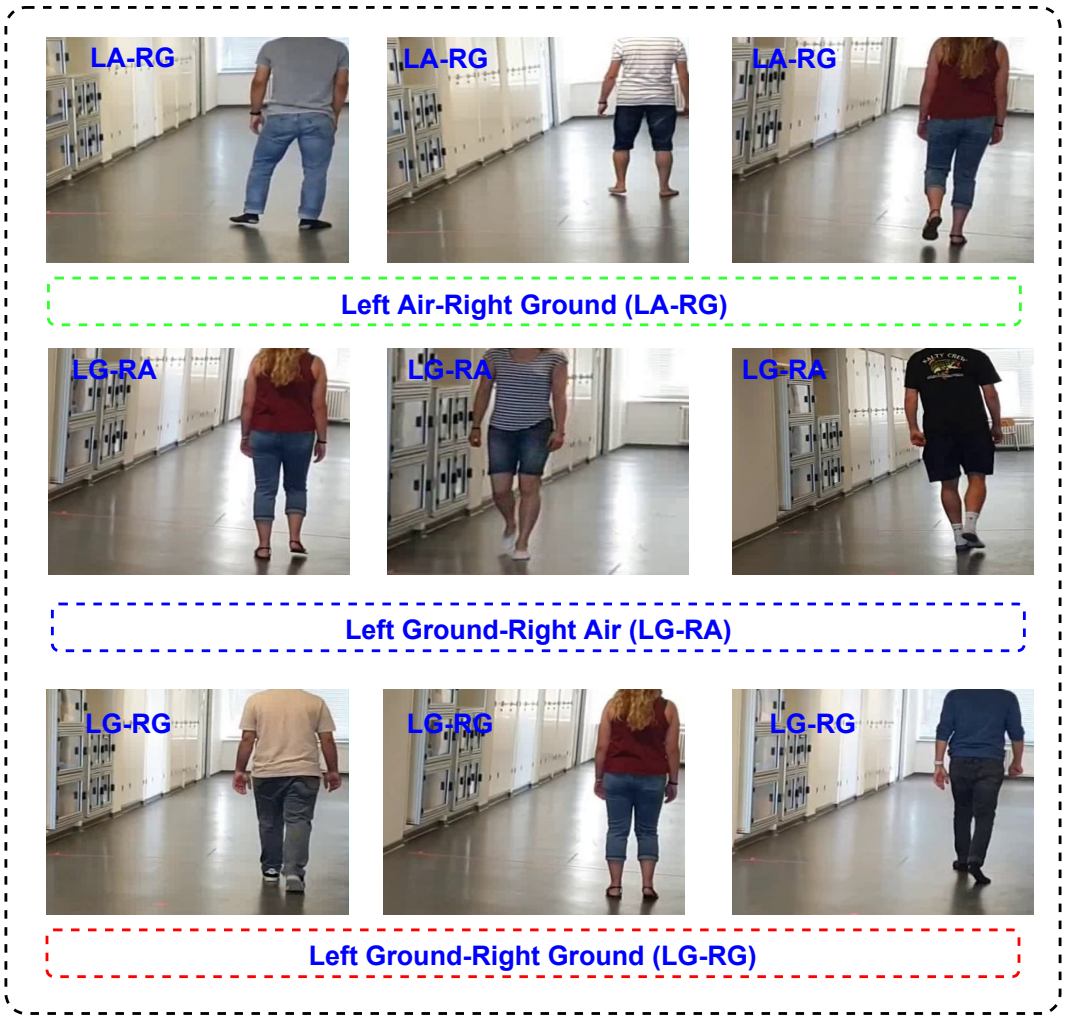


Figure 4.2: Data annotation examples from each class. The first three images are labelled as LA-RG, the next three images are labelled as LG-RA, and the last three images are labelled as LG-RG.

General Considerations in Manual Annotation

The position of the feet remains the same for several frames during walking. Therefore, marking the change in the feet' position is only necessary. This phenomenon can be explained with the help of an example; if the label of the first frame is LA-RG, then it will remain the same for several frames because switching from one position to another takes time. The next label will be LG-RG for a certain number of frames, and then the label will be changed to LG-RG. After some frames, it will be changed to LG-RA. Therefore it is optional to look at each frame one by one. It reduces the effort of labelling the frames. Nevertheless, there is a chance of error in labelling, which should be considered because it will impact the accuracy of classification of feet position by deep NN. This error is difficult to quantify systematically; however, a bias can be inducted to remove the effects of terror. In this thesis, we did not include any such bias while analyzing the performance of the different deep NNs in classifying the feet' position.

4.4.3 Data Augmentation

As discussed in the above section, the data set was the walking videos from 20 subjects and the total no of images was 86,232. There were 3 classes of data namely **LG-RG**: When both feet are ground, **LA-RG**: when the left foot is in the air and the right foot is on the ground, and **LG-RA**: when the left foot is on the ground and right foot is in the air. Gait is a cyclic process and 60% of the gait cycle is the stance phase when both feet are on the ground, and rest 40% of the gait cycle is the swing phase when one foot is in the air. This proportion of the stance and swing creates an imbalance in the data set. For example for subject 2, during walking in NPH condition, the total number of images, for class LG-RG was 773, for class LA-RG was 515, and for class LG-RA was 498. This is almost in line with the general gait cycle stance//swing ratio. This class-wise variation in samples becomes imbalanced classification problem that caused poor predictive performance, specifically for the class with less number of samples. If the imbalanced classification problem is severe then it is more challenging to suggest a robust approach. Besides, datasets with a small number of sample spaces are another problem for researchers while the training of a model is carried out. To handle these problems, the data augmentation process is selected to enhance class-wise data by considering existing data. Data augmentation is a technique in machine learning that can be used to avoid overfitting in small datasets. A dataset is artificially extended by data augmentation. Hence, this process was applied to the collected dataset with a random oversampling (ROS) technique to generate synthetic data on existing gait images as reported by [197]. The reason to choose ROS in data augmentation is two-fold: 1) to handle the

class-imbalanced problem by providing a suitable way to equally distribute class-wise data, and 2) to increase the class-wise total number of samples up to the appropriate size of data. To form the balanced data, first, the required number of samples is randomly chosen for one set as described in Table 4.1.

Then, four different operations are applied for data augmentation. The selected operations, for instance, image rotation, horizontal flipping, vertical flipping, and adjusting brightness are used to equalize the number of samples in different classes. The randomly selected gait images are resized to 280×240 and then data augmentation operations are performed on the selected samples to handle class-imbalanced problems. The description of these operations is given below.

- **Image Rotation:** The rotate operation is applied to generate a randomly rotated image R_{img} from -30 to 30 degrees.
- **Horizontal Flipping:** This operation is applied to generate a randomly flipped image HF_{img} horizontally. For example, IM is assumed as a matrix from an image, then the flip operation reverses the elements of each row under horizontal flipping.
- **Vertical Flipping:** This operation is applied to generate a randomly flipped image VF_{img} vertically.
- **Adjust Brightness:** This operation is applied to adjust image intensity values by stating contrast limits with low and high values in the range of [0, 1] to produce a new image BR_{img}.

Later, rotation, horizontal flipping, vertical flipping, and adjust brightness operations are referred to as rotate, hor_flip,ver_flip, and brightness, respectively. While performing these operations, 1vs1, 1vs4, and both strategies are utilized to generate synthetic data. Resultantly, three different augmented datasets are generated based on these strategies. To realize the 1vs1 strategy, the first four sets are prepared from the LG-RG class of the gait dataset, where 70 non-overlapped images are randomly chosen (from NPH walking of one subject). Then a single operation (e.g., rotate) is performed by selecting an image from a given set of LG-RG images. It means that against one image, only one operation will be performed instead of all operations, as shown in Figure 4.3. In addition, images of the selected set will not be considered for further operations. As an output, 280 augmented images are generated for the LG-RG class. The detailed description of a class-wise selected number of samples for each set, the number of augmented images with original images, and the total images are shown in Table 4.1.

Similarly, different images are collected from the LA-RG and LG-RA classes of the gait dataset to prepare four different sets of each class. These two classes contain a smaller

Table 4.1: An example of data augmentation statistics for imbalanced and small sample Gait datasets of one subject (subject number 2) in NPH walking, a class-wise selected number of samples for data augmentation, and resultantly, total augmented images and total images are presented

Images and set Details	Strategies for data augmentation operations								
	1vs1			1vs4			Mixed		
	LG-RG	LA-RG	LG-RA	LG-RG	LA-RG	LG-RA	LG-RG	LA-RG	LG-RA
Class-wise selected images	70	80	80	70	80	80	-	30: 1vs 1 30: 1vs 4	30: 1vs 1 30: 1vs 4
Augmented Images	280	320	320	280	320	320	-	240	240

Original Images	LG-RG	LA-RG	LG-RA
	773	515	498

Total Images	LG-RG	LA-RG	LG-RA
	1333	1395	1368

number of images as compared to LG-RG. Thus ,there is no option to create four sets from each class for data augmentation. Therefore, images of each of the classes are divided into two sets to collect 80 images from each class. In this scenario, there are two different sets for data augmentation, so two operations (e.g., rotate and hor_flip) are applied against a single set instead of one operation as applied on selected samples in the case of LG-RG class. Hence, two operations of rotate and hor_flip are performed on the first set of 40 images of each class, and the other two operations of ver_flip and brightness on the second set of 40 images. Consequently, 320 augmented images are acquired for LA-RG and LG-RA classes, as shown in Table 3.5. To apply the 1vs4 strategy, a single set of 70 images from LG-RG and a single set of 80 images from the other two classes of the gait dataset are randomly selected. Then, four operations (rotate, hor_flip, ver_flip, and brightness) are done on each image chosen from the given set of all classes instead of a single operation. It means that four images will be generated from a single image, as shown in Figure 4.3.

As 1vs1 and 1vs4 strategies are applied to generate augmented data from all three classes of the data, furthermore a mixed (1vs1 + 1vs4) strategy is also adopted for data augmentation. This strategy was applied to classes with smaller samples, i.e. LA-RG and LG-RA.

Therefore, in the second step, images from LA-RG and LG-RA classes are randomly selected to create different sets, as described in Table 4.1. Then, all selected operations are executed. For example, using the LA-RG class of the gait dataset, the 1vs1 strategy on 30 images and 1vs4 strategy on the other 30 images are applied. Based on these strategies, augmentation is performed for the balanced data distribution in all three classes to handle the imbalanced classification problem for gait event prediction.

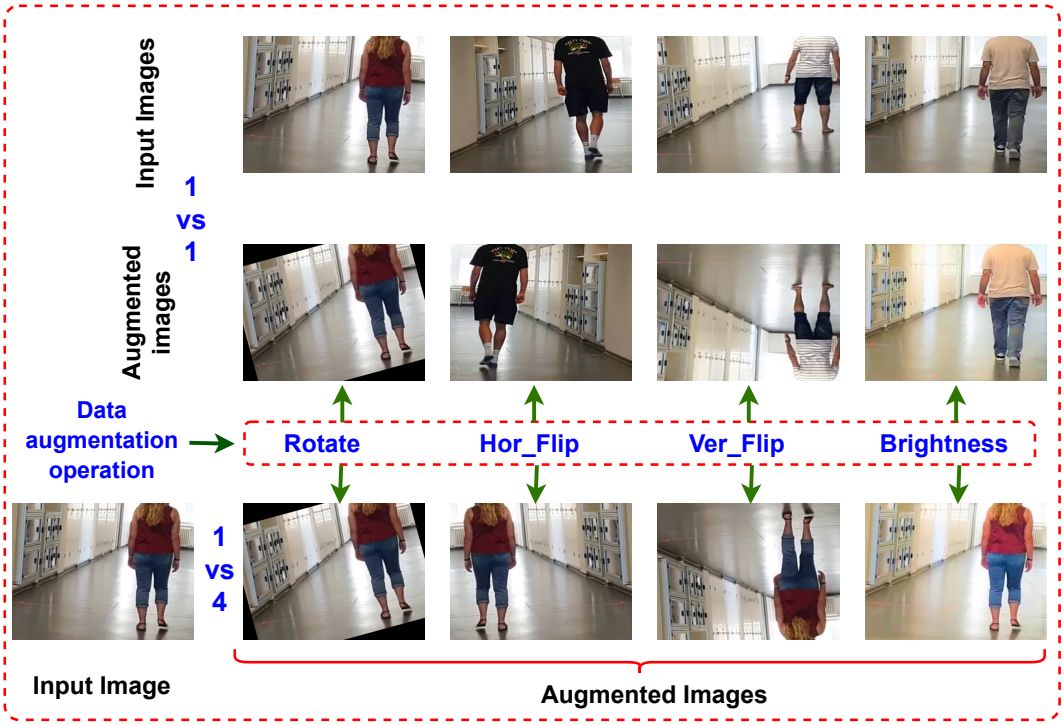


Figure 4.3: An example of data augmentation statistics for imbalanced and small sample Gait datasets of one subject (subject number 2) in NPH walking, a class-wise selected number of samples for data augmentation, and resultantly, total augmented images and total images are presented.

After image augmentation of the whole dataset, the total number of images are 144240 out of which 49472 belong to class LG-RG, 47399 belong to class LA-RG, and 47,099 belong to class LG-RA.

4.5 Methods and Implementation

Deep learning methods that are used for machine vision problems normally use deep CNNs for recognizing or classifying the input data to certain labels. These methods require a large amount of labelled data for training. Once these models are trained with adequate accuracy, can be used for testing over unseen data. The proposed approach in this study is also based on CNN where we have videos as raw input which are collected during walking in two different conditions (and two different protocols). These videos are preprocessed to convert them into images and also make size adjustments. Then manual data annotation is performed in which each image is assigned a label out of three class labels. To enlarge the data set size and to reduce the imbalance among classes, data augmentation is performed using basic augmentation methods. Besides this dataset, a publicly available gait dataset

CASIA [195, 196] is also used to make the dataset adequate enough for the training of a deep neural network. After that this data along with the labels is given to a neural network to classify this into three classes which are basically foot positions during walking. This whole scenario is represented in Figure 4.4. Data collection, preprocessing, data annotation, and data augmentation steps are already discussed in the previous sections. We used different NN architectures in this work to find a better classifier. In the following sections, the architectures of the neural networks used for the classification of gait events are discussed.

4.5.1 CNN Based on AlexNet

The research has shown that a variety of pre-trained deep learning models are available to solve classification problems, object detection problems, and recognition problems. To avoid the in-built complexity and high computation time, we want to train the model based on a simple and foundational CNN, AlexNet introduced by [103]. It can be easily trained and optimized as compared to complex CNN architectures such as VGGNet, RestNet, and GoogleNet [197].

This CNN architecture consists of multiple layers such as five convolutional layers (CLs) including pooling layers (PLs) and fully connected layers (FCs) layers and this model is used to learn discriminative features from an input image. It consists of 5 CLs, 3 PLs, and 3 FCs layers. Each CL comes before the PL. This proposed deep CNN architecture is presented in Figure 4.5 and the parameters setting of the network is depicted in Figure 4.6. The standard parameters of the AlexNet model are used to learn deep features without any other optimization as they are already well-tested. In this work, an RGB image is resized into 227×227×3 dimensions and then a bi-cubic interpolation algorithm is applied for the equalization of image details.

Thus, an image with a size of 227×227×3 is input into the network. The convolutional operation is applied by using the following equation.

$$Z_x^l = b_x^l + \sum y f_{xy}^l \otimes z_y^{l-1} \tag{4.1}$$

where Z_x^l represents output channel values up to point x at layer, Z_y^{l-1}shows input channel values up to point y at layer $l-1$, and f_{xy}^l represents convolutional filter between x^{th} and y feature maps. The bias b_x^l is added to move the activation function towards successful learning. For activation of neurons, the rectified linear unit (ReLU) is applied through Equation (4.2).

$$ReLU(z) = max(0,z) \tag{4.2}$$

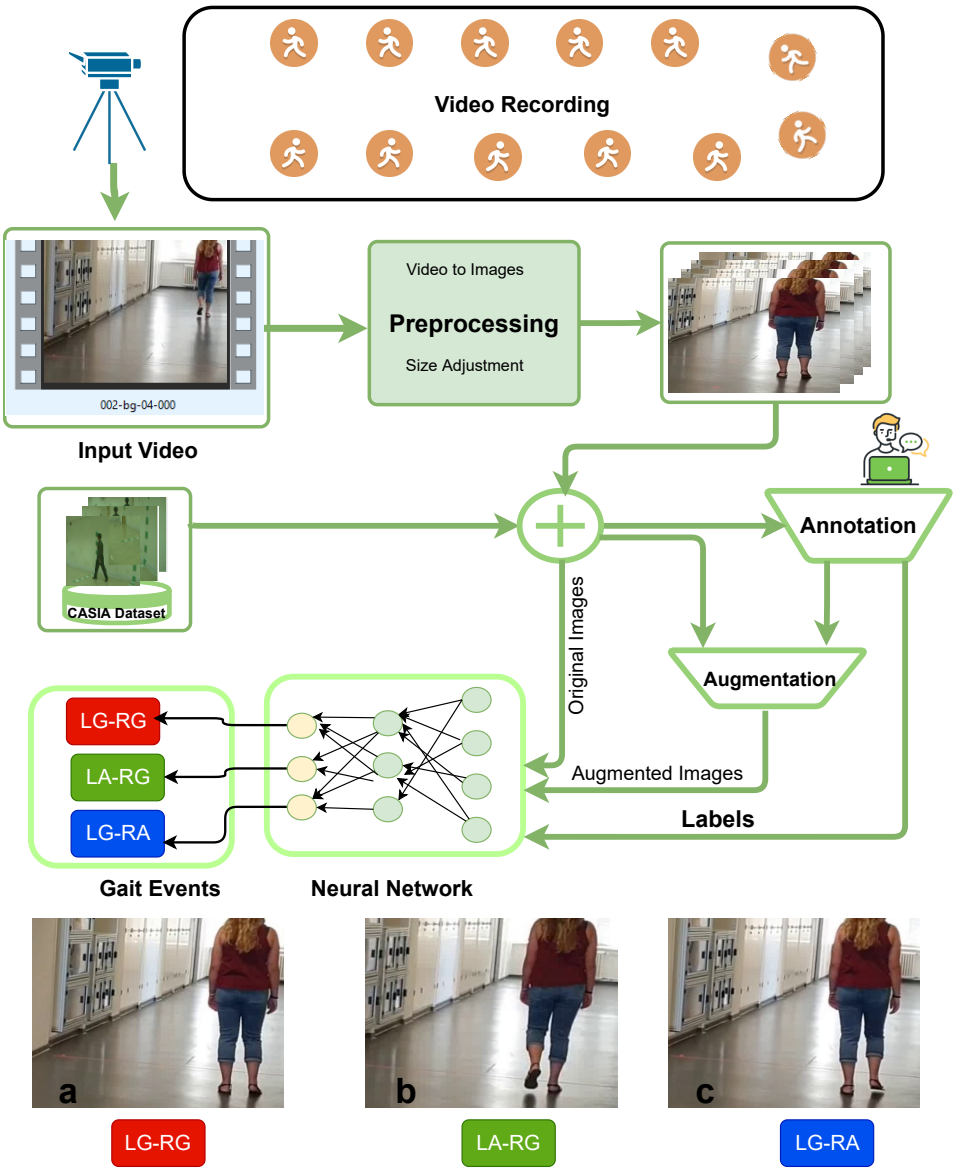


Figure 4.4: A schematic overview of the proposed approach; Videos are collected while people walk; Then these videos are prepossessed and converted into image at frame rate of 30fps. Secondly, We also used publicly available CASIA gait data set [195, 196] to enlarge the dataset. After that true labels are determined with the help of Annotation and then images and labels are given to NN for Training to make the model able to classify into one of three feet' positions (LG-RG LA-RG, LG-RA)

Downsampling in CNN architectures is also carried out using max pooling, which is a relatively simple process without a learning curve. In this network, max pooling is applied after the first, second, and fifth layers. The Figure 4.7 illustrates the maximum pooling operation performed on a given sample feature region. Basically, it applies a $k \times k$ sized filter and selects the maximum value using the function Equation (4.3)

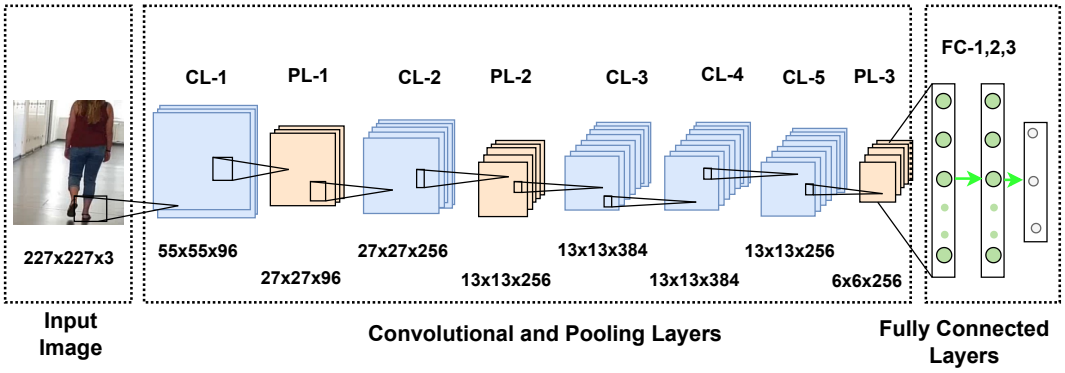


Figure 4.5: A schematic overview of CNN model having 5 convolutional layers, 3 pooling layers and 3 fully connected layers to extract features from input image.

$$Z_{pqc} = \max_{(i,j)\in M_{p,q}} U_{ijc} \tag{4.3}$$

where $M_{p,q}$ shows pooling region having indices i, j, U_{ijc} region of the feature map, color space channel c and then a pixel value Z_{pqc} is obtained as an output of max pooling operation. After the first two pooling layers, local contrast divisive normalization (LCDN) is applied by considering the interaction between channels C (multi-channel images), where the variance of local area M_{ab} is computed by applying Equation (4.4).

$$\sigma_{ab}^2 = \frac{1}{c}\sum_{c=0}^{C-1}\sum_{(i,j)\in M_{ab}} w_{ijc}(x_{a+i,b+j,c} - \bar{x}_{ab})^2 \tag{4.4}$$

The divisive normalization is then calculated according to Equation (4.5).

$$Z_{abc} = \frac{x_{abc} - \bar{x_{ab}}}{max(\hat{c}, \sigma_{ab})} \tag{4.5}$$

If $\sigma_{ab} < \hat{c}$ then divide with $\hat{c}$. In divisive normalization, continuous change in the denominator depends upon two values (1) constant value $\hat{c}$ and (2) variance value, as expressed in Equation (4.6). The application of the normalization process is robust to variations in illumination and contrast.

$$Z_{abk} = \frac{x_{ab} - \bar{x}_{abk}}{\sqrt{\hat{c} + \sigma_{ab}^2}} \tag{4.6}$$

The remaining part of the architecture consists of three FC layers. The first two layers deal with extracted features of previous layers and decrease the dimensionality of such features from 9216 to 4096. It is a deeply applied architecture because, in gait images, a given pattern may appear differently due to various factors, including distance from the camera, clutter in the background, and variations in lighting and pose. An adequate deep

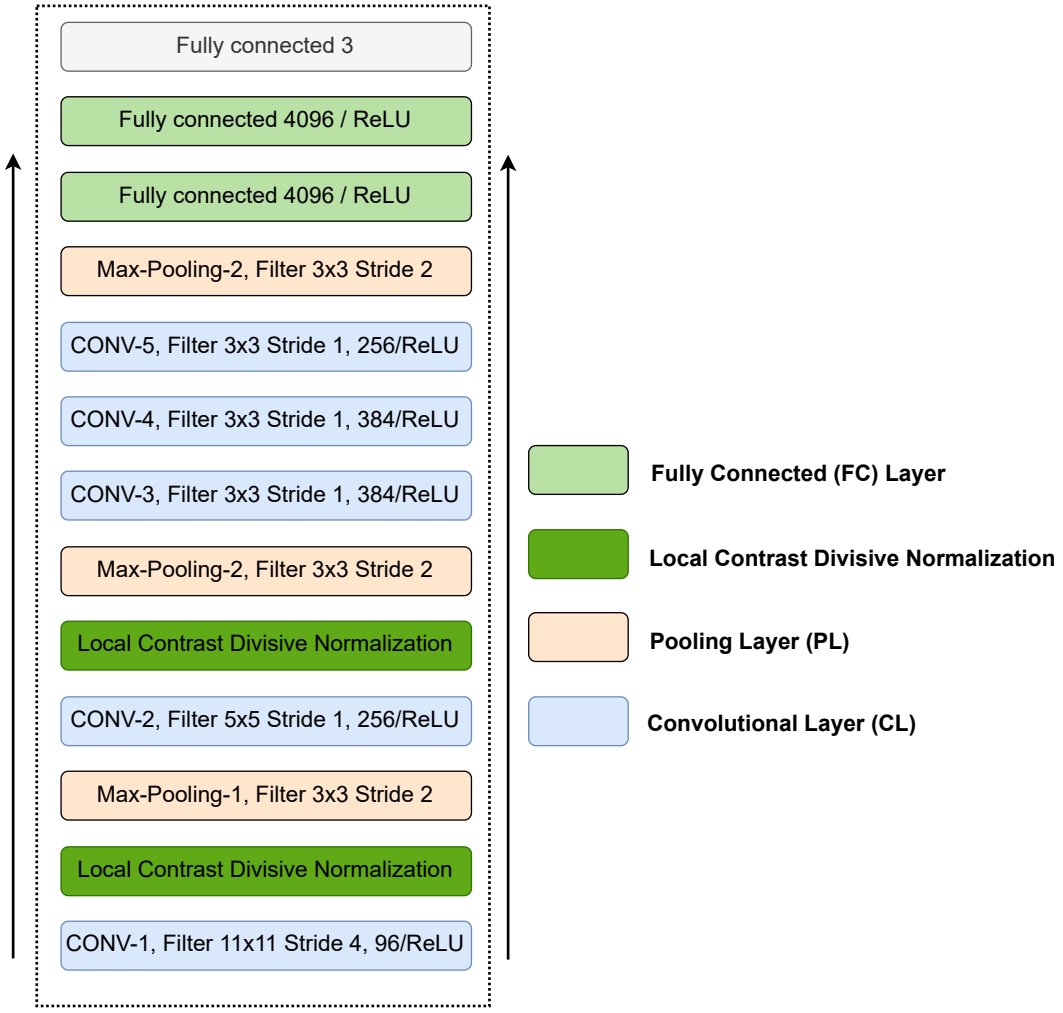


Figure 4.6: Parameter setting for each layer of the model presented in Figure 4.5. Number of filters, activation function and other parameters are given.

architecture is necessary to address these problems of appearance and inherent ambiguity. A stochastic gradient learning algorithm is subsequently used to train the CNN model. As shown in Equation (4.7), all k consensus clusters and their associated sample images are considered for deep feature extraction.

$$C_k = \left\{S_{(k,s_1)}, \ldots, S_{(k,s_2)} \ldots, S_{(k,s_N)}\right\} \tag{4.7}$$

where C_k represents consensus cluster, $S_{(k,s1)}$ is considered as sample image, k denotes k^{th} cluster, N depicts total sample images in C_k, and S_N represents total number of sample images in one consensus cluster.

Finally, a deep CNN model is applied and feature extraction on each sample image of consensus clusters g_{cc} is carried out using Equation (4.8).

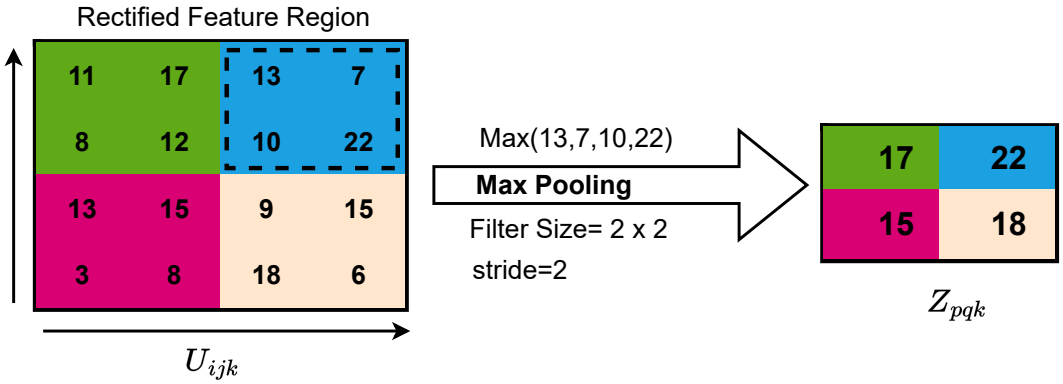


Figure 4.7: An Illustration of Max Pooling operation, maximum value from each of the 2 x 2 matrix is selected in max pooling.

$$DF_{1\times d} = \gamma_{cc}(k, s_i) \tag{4.8}$$

where $DF_{1\times d}$ denotes deep features having dimension d and S_i depicts i^{th} sample image of cluster k, extracted by applying γ_{cc} operation. Similarly, this procedure is applied across all the consensus clusters to extract deep features of each sample image. The reduced deep features are then used to filter out the noise by discarding unnecessary information and preserving discriminative information by employing principal component analysis (PCA). The empirically selected 1000 deep features are used as optimal features subsets (OFS) to build a feature vector for all experiments.

4.5.2 CNN-LSTM Hybrid Model

In fully connected networks, the layers are fully connected and the nodes between layers are connection-less and process only one input. They lack the memory or storage element and cannot remember their previous states or output. Moreover, in the real world, there are many applications in which the system can't decide the correct next state until it doesn't have information about previous states or some context information. One of these applications is the human gait, where the next steps are dependent upon the previous steps and there is a correlation among stride intervals. If the one stride interval is small, the next will also respectively small. The abrupt change in stride interval duration is rarely seen. Therefore, these parameters are correlated and information about the current state depends upon previous states also. Therefore a RNN is required in such a case. The critical distinction between feed-forward neural networks and RNN is their ability to operate on the input space in addition to performing on the internal state space [198]. It is important to highlight that the

RNN model has two downsides: gradient disappearance and gradient explosion. Therefore the LSTM architecture was developed to overcome those issues by introducing the notion of input and output gates [199]. The working principle of the LSTM model is similar to the RNN. They differ, however, in the employment of gates for the recurrent neurons called the forget gate f, update gate, and output gate in addition to the internal processing unit called the cell [200]. Each of the gates is in charge of a specific task in the cell. The forget gate's role is to get rid of undesired information from the former state and output of the upper hidden layer h_{t1}. The update gate refreshes the state with new elements, while the cell is in charge of filtering the current state and finding the desirable and undesirable information to ensure that the output gate chooses the crucial information produced by it. The term x_t incorporates valuable features of gait events and is utilized as an input for the memory cell. The output layer selects the data that will be the outcome and is handled by filtrated input o_t and cell state c_t [201]. A detailed mathematical description of the LSTM model is given in section 3.3.1 of this thesis.

The hybrid models combine deep learning architectures and are being used in many applications. CNN-LSTM model combines the CNN and LSTM networks to automatically detect the gait events from gait images. In the proposed system, CNN is used for feature extraction from input images. Then the feature vectors are constructed into the sequence form, which is transmitted to the LSTM network. The LSTM layer learned the changing rules of gait events in the data to predict the class. The LSTM network has an internal memory that is capable of learning from imperative experiences with long-term states.

Therefore, a hybrid approach of CNN-LSTM is beneficial in a dataset like this study. The nodes in LSTM are connected from a directed graph along a temporal sequence that is considered an input with a specific sequence of events [202]. Hence, the 2-D CNN and LSTM layout feature combination can improve classification greatly.

Implementation Details: In CNN-LSTM hybrid models, there are three blocks, first block is the combination of CL and PL layers. After the pooling layer, the distilled feature maps are flattened into a single long vector which is used as input to the next block. The second block is the LSTM layers block which receives the feature vector from the previous block. Each layer represents an element vector that represents the features captured by the decoder model after reading the input sequence. In the first step, the internal representation of the input sequence is replicated many times, once for each time step in the output sequence. The sequence of vectors will be provided to the LSTM decoder. A fully connected layer is used before the final output layer to interpret each time step in the output sequence. This implies that each step would also contain the same layers. A fully connected layer and an output layer are used for each time step provided by the decoder. With this approach, the LSTM decoder

can determine the context for each step of the output sequence, and the wrapped dense layer can interpret each step independently while still applying the same weights. The LSTM is made of hidden layers with 40 units as 40 images in a sequence were given as input. This number was selected because all three feet' positions are covered in 40 consecutive frames. Increasing this number will increase the computational cost, and reducing this number may result in missing any of the feet' positions. The third block is of fully connected layer and the output layer, which produces the prediction of the gait events. The same CNN as described in the previous section (4.5.1) was used to keep things simple and make the comparison better. So it has 5 PLs, 3 PLs and a flatten layer. The feature vector from this flatten layer is given to the LSTM block, which takes input at different time steps and produces output according to time. The CNN-LSTM hybrid model is presented in Figure 4.8. The layers configuration of this hybrid model is presented in Table 4.2.

4.5.3 Training of the Neural Networks

Both models (CNN based on AlexNet, and CNN-LSTM hybrid) of the neural network were trained with the same training strategies to evaluate both networks' performance for predicting feet' position. The same strategies were used to avoid any effect of the training method on the classifier's performance. Furthermore, the images were normalized over the complete dataset. For the training, the total dataset was divided as 60% for training, 20% for validation, and 20% for testing. The test data was never used in any case by the neural network during training.

There were two types of datasets: the gait dataset collected from volunteers in normal and NPH walking (referred to as gait dataset), and the CASIA dataset collected for gait recognition[195, 196]. In the first attempt, only the gait dataset was used. It has 144240 images (original + augmented) converted from videos belonging to three classes. Therefore, training images were 86544, while validation and test images were 28848 each. As a second strategy and to enlarge the dataset size, both datasets were used to train the neural networks. To avoid overfitting, fewer images were selected from CASIA data set as compared to the original gait dataset. Moreover, all the images were used for training and validation. Images from the CASIA dataset were not used for testing because we only wanted to determine the network performance on the data collected using the proposed methodology, which is simple. Therefore, from the CASIA dataset, 53608 images were used for training, and 17869 images were used for validation. The selection of images from each class was also according to the identical percentage. Table 4.3 shows the schematic distribution of the train, validation and test split of both data sets.

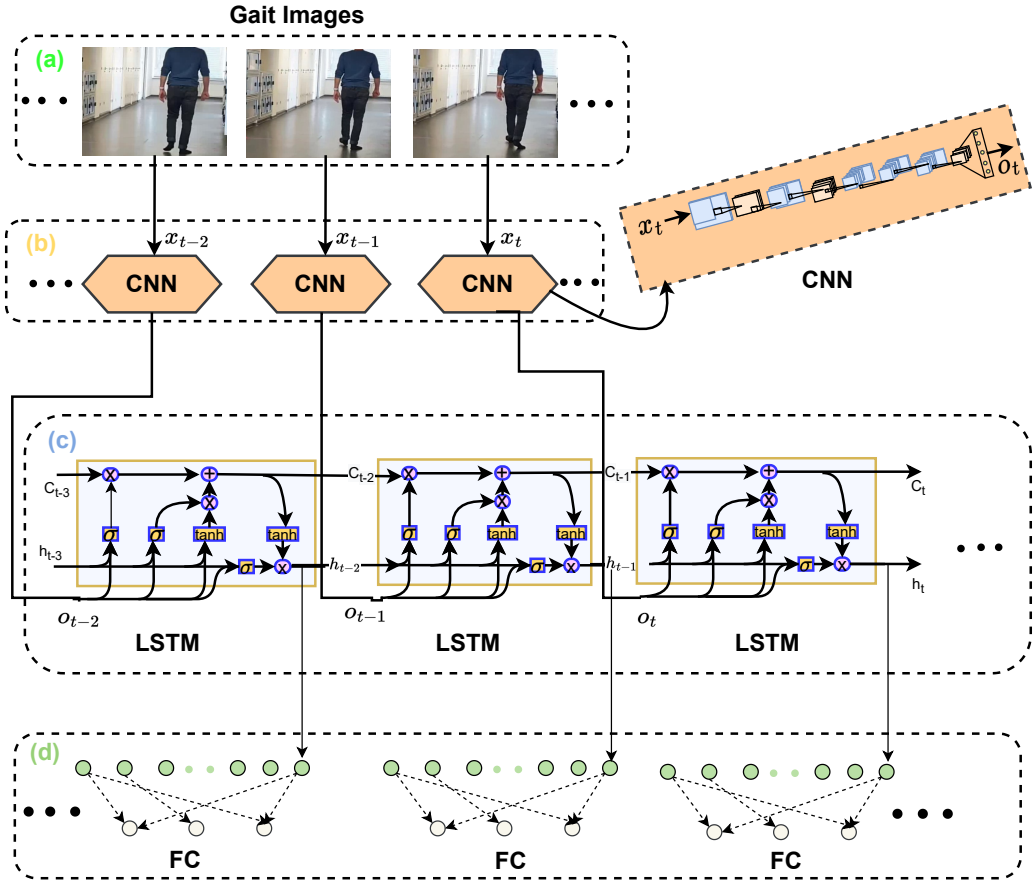


Figure 4.8: A representation of CNN-LSTM hybrid model for feet' position classification. (a) The input stream of images, (b) CNN based on the approach discussed in section 4.5.1. (c) LSTM networks with input from CNN, (d) fully connected layer and output layer.

We used adam optimizer, which is an extension to stochastic gradient descent (SGD) that has recently seen broader adoption for deep learning applications in computer vision and natural language processing, in all training sessions to optimize the NN. We used a learning rate of $lr = 0.01$, $momentum = 0.9$, and a linearly decreasing learning rate of 0.1 every ten iterations during the training phase.

Furthermore, we experimented with adding dropouts and early stoppings to reduce overfitting in the model. A dropout rate of 0.5 was added between each layer of the fully connected model. A patience = 5 was applied to the model's training to minimize validation loss. For training purposes, we used NVIDIA RTX 2080, 6x2.8GHz, 32 GB RAM, with the operating system Windows 10. The batch size for CNN was 128, while for CNN-LSTM hybrid model, it was 40 because we used the 40 nodes LSTM. The frame sequence was

Table 4.2: Layers configuration in CNN-LSTM hybrid model with 5 convolutional layers, 2 pooling layers, and a LSTM layer

Layer	Configurations	
CL-1	filters	96
	kernel size	11x11
	activation	ReLu
PL-1	pool size	3x3
CL-2	filters	256
	kernel size	5x5
	activation	ReLu
PL-2	pool size	3x3
CL-3	filters	384
	kernel size	3x3
	activation	ReLu
CL-4	filters	384
	kernel size	3x3
	activation	ReLu
CL-5	filters	256
	kernel size	3x3
	activation	ReLu
PL-3	pool size	3x3
Flatten	-	-
LSTM	hidden node	40
	activation	relu
	return sequence	TRUE
Time Distributed	Dense	100
	activation	relu
ouput		3

maintained to keep the temporal information. The images were randomly selected from both datasets. As this is a multi-class prediction, we used the cross-entropy as loss function. The maximum number of training epochs we used was 100, but, in most cases, it converged earlier.

Table 4.3: Distribution of data sets (gait dataset and mixed dataset) in training, validation, and test blocks

Data set	Total Images	Training	Validation	Testing
Gait Dataset (From Volunteers)	144240	86544	28848	28848
Mixed Dataset (Gait Dataset+CASIA)	231787	139073	46357	46357

4.6 Results

Two experiments were conducted in which the methods of detecting gait events were evaluated in terms of performance. In the first experiment, the dataset collected by recording the videos of volunteers in normal and NPH conditions (referred to as the "gait dataset") was used to train the CNN based on AlexNet (referred to as A-CNN), and the CNN-LSTM hybrid model (referred to as CNN-LSTM). In the second experiment, the gait dataset and CASIA data set were combined (referred to as mixed dataset) and used to train the same A-CNN and CNN-LSTM.

4.6.1 Feet' Positions Classifications on Gait Dataset

In this experiment, only the gait dataset was used. It has a total of 144240 images (original + augmented). From this dataset, 86544 (60%) images were used for training, and 28848 images were used for each validation and testing. We trained the A-CNN classifier without temporal information and the CNN-LSTM classifier with temporal information using 40 images as a sequence. Figure A.1 (a) shows the training and validation accuracy of the A-CNN model when we used the gait data. The corresponding training and validation loss is shown in Figure A.1 (b). With both dropout and early stopping, the training stopped on epoch 71 of the maximum of 100 epochs, with the best result appearing at epoch 67 with training accuracy (ACC) of 80%, validation accuracy of 75%, and validation loss of 0.26.

In the case of the CNN-LSTM model, the minimum validation loss was: 0.17, and the accuracy was 88%. The training stopped after 63 epochs of a maximum of 100 epochs, and the best accuracy was found at epoch 54, and the minimum validation loss was at epoch 56. The results are shown graphically in Figure A.2(a,b).

To measure and evaluate the performance of the discussed models on test data, multiple metrics were used. We selected some measures of the overall metrics for multi-class datasets including confusion matrix, ACC, precision (PREC), SENS, F-1 score (F1-Score), specificity

(SPEC), weighted average precision ($PREC_{WAvg}$),weighted average recall (REC_{WAvg}),micro F-1 score ($F1_{Micro}$),and macro F-1 score ($F1_{Macro}$).

A-CNN: The overall ACC of the A-CNN model was 74%. The $PREC_{WAvg}$ was 74.29% and REC_{WAvg} was 73.58%. The classifier predicted the classes with $F1_{Micro}$ of 73.5% and with $F1_{Macro}$ of 73%.

In the Table 4.4, the binary analysis of the classification with measures like ACC. PREC, SENS, SPEC, and F1-Score are shown. A comparison was made between the respective class and all others aggregated into one class. It means the table shows the classifier's ability to recognize the presence or absence of a specific gait event class among all classes of gait events. The A-CNN model predicted the individual classes with an accuracy of just over

Table 4.4: Results of the classification performance of A-CNN and CNN-LSTM on the test data from gait dataset with the measures of ACC, PREC, SENS, and F1-Score, rounded on 3 digits

	A-CNN Model					CNN-LSTM Model				
	ACC	**PREC**	**SENS**	**SPEC**	**F1-Score**	**ACC**	**PREC**	**SENS**	**SPEC**	**F1-Score**
LA-RG	0.828	0.708	0.750	0.863	0.729	0.884	0.809	0.831	0.908	0.820
LG-RA	0.837	0.679	0.776	0.860	0.724	0.888	0.795	0.843	0.908	0.818
LG-RG	0.807	0.811	0.699	0.884	0.751	0.874	0.861	0.801	0.920	0.830

80%, while the highest accuracy was of class LG-RA, and the lowest was class LG-RG with 80.67%. In contrast, the precision of the class LG-RG with a value of 81.06% was significantly high compared to the other classes with values of 70.8% and 67.9%. The sensitivity value of class LG-RA was also high compared to other classes, with a value of 77.5%. Regarding specificity and F1 score, class LG-RG was predicted better than the other two classes. The confusion matrix model showing the heatmap for A-CNN model is presented in Figure 4.9.

The variations in prediction among classes was about 2-3 frames and it was specifically seen while changing the foot position from one phase to another. For example, if the previous position of both feet was on ground (LG-RG), then some of the initial frames of the next phase (either LA-RG, or LG-RA) are predicted as LG-RG. The same is while changing from the other two positions to the LG-RG position. These variations are shown in Figure 4.10.

CNN-LSTM: The CNN-LSTM model showed a better accuracy and other measures compared to A-CNN model. This model predicted the true classes correctly with an over overall ACC of 82.3% and an of 88.2%. The $PREC_{WAvg}$ was 82.48% ,REC_{WAvg} was 82.32%, $F1_{Micro}$ was 82.32% , and $F1_{Macro}$ was 81.8%.

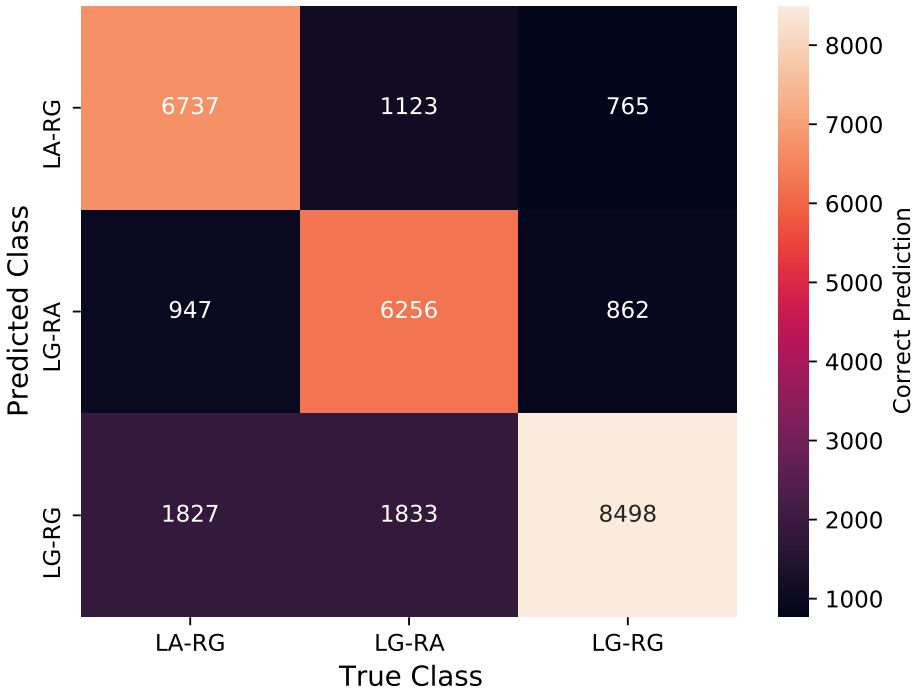


Figure 4.9: Confusion Matrix also showing heat map of A-CNN classifier on test data of gait dataset.

Further we also evaluated the binary classification for CNN-LSTM model. The results are shown in Table 4.4 where binary analysis of the classification with measures like ACC. PREC, SENS, SPEC, and F1-Score are shown. A comparison was made between the respective class and all others aggregated into one class.

The accuracy in case of individual classes was also improved compared to the other model.The highest accuracy was for classes LG-RA, and LA-RG with a value slightly above 88.8%.In case of precision,the class LG-RG with a value of 86.1% was significantly high compared to the other classes with values of 80.9% and 79.5%. Same was the trend with specificity and F1-score with values 92 %, and 83% respectively for LG-RG class. However, The sensitivity value of class LG-RA was also high compared to other classes, with a value of 84.3%. F1-Score for classes LA-RG, and LG-RA was 82% and 81% respectively.The confusion matrix for CNN-LSTM model showing the heatmap is presented in Figure 4.11. The variations in prediction among classes was also reduced as compared to the A-CNN classifier. At some points there is a variation of about 3 frames but mostly it is, up-to two frames. Like the previous model ,it was also specifically seen while changing the foot position from one phase to another. For example, if the previous position of both feet was

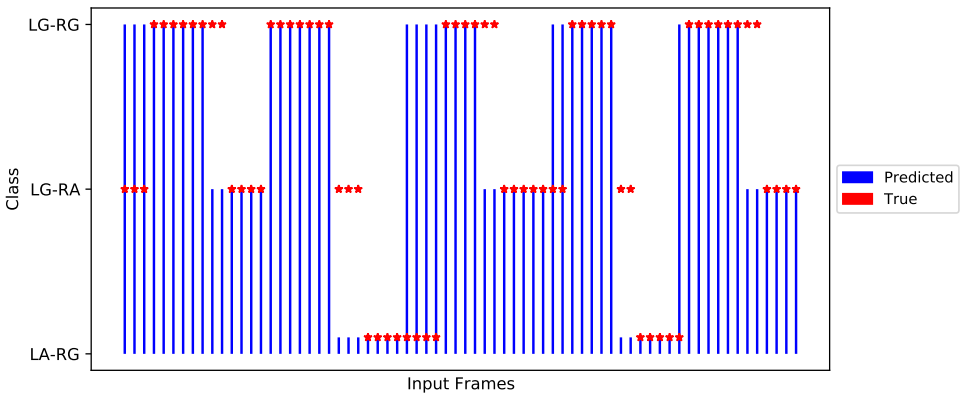


Figure 4.10: An illustration of the variations while predicting the labels for each frame of video using A-CNN model, there is a variation while changing the feet position.

on ground (LG-RG), then some of the initial frames of the next phase (either LA-RG, or LG-RA) are predicted as LG-RG. The same is while changing from the other two positions to the LG-RG position. These variations are shown in Figure 4.12. Based on the first result, it can be concluded that the CNN-LSTM hybrid model substantially improved accuracy. However, the trend remains the same; the individual accuracy of class LG-RA was highest in both cases. However, in the case of specificity, the LG-RG class was better classified compared to the other two classes. The ability of the CNN model individually to detect gait events was not enough, and it improved when we added the temporal information by using the LSTM model combined with CNN. The small difference may be because of the same CNN architecture used in both models. Although the misclassification is within 2-3 frames (shown in Figure 4.12) and especially at the end of each gait event (for example, from LA-RG to LG-RG, or from LG-RA to LG-RG); still accuracy is not well enough. As we already concluded that CNN-LSTM is a better model than the A-CNN model; therefore, in the next step, we increased the amount of data by adding the CASIA dataset to retrain the network and see the performance. This experiment is discussed in the next sections.

4.6.2 Feet' Positions Classification on Mixed Dataset

This experiment used the CASIA dataset to expand the dataset. The gait dataset has a total of 86332 images (original without augmentation) and we selected the same number of images from the CASIA dataset as well to prevent overfitting. Furthermore, we included the augmented dataset as well. A total of 231787 images were used in this experiment.

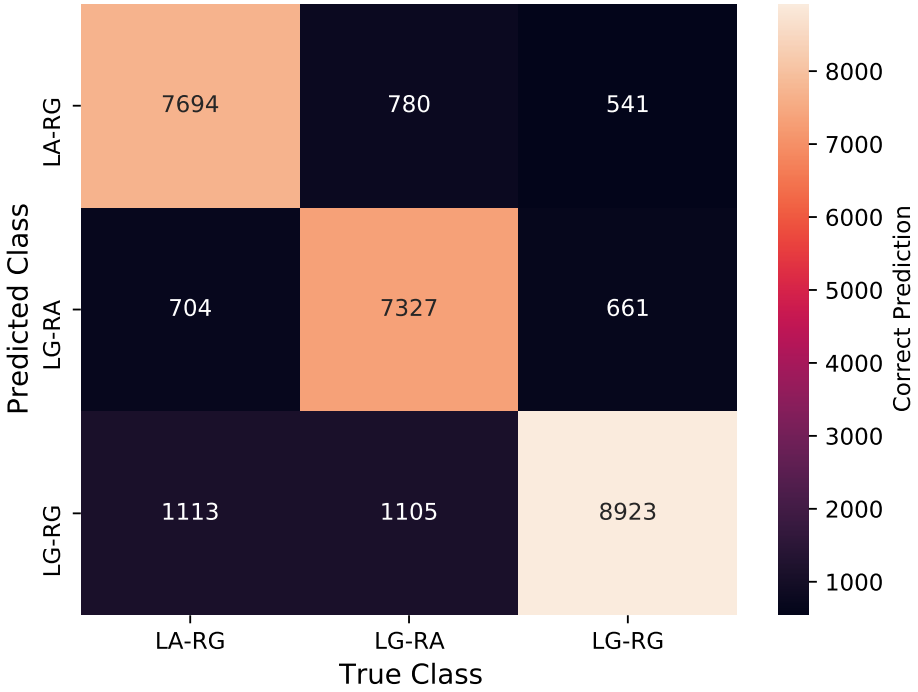


Figure 4.11: Confusion Matrix also showing heat map of CNN-LSTM classifier on test data of gait dataset.

139073 images (60%) were used for training, while 46357 images were used for testing and validation. It is to note that for testing, we only included the images from the gait dataset. As the CNN-LSTM model previously performed very well compared to the other model, we selected only the CNN-LSTM model for retraining with temporal information using the same amount of images as a sequence. Figure A.3 (a) shows the training and validation accuracy of the CNN-LSTM model on mixed dataset. The corresponding training and validation loss is shown in Figure A.3 (b). With both dropout and early stopping, the training stopped on epoch 71 of the maximum of 100 epochs, with the best result appearing at epoch 65 with training ACC of 93%, validation accuracy of 92%, and validation loss of 0.13.

To measure and evaluate the performance of the CNN-LSTM model on test data, multiple metrics were used. We measured overall metrics for multi-class datasets including confusion matrix, ACC, PREC, SENS, F1-Score, SPEC, $PREC_{WAvg}$,REC_{WAvg},$F1_{Micro}$,and $F1_{Macro}$. The overall ACC of the CNN-LSTM model was 91.4%. The $PREC_{WAvg}$ was 91.6% and REC_{WAvg} was 91.3%. The classifier predicted the classes with $F1_{Micro}$ of 91.3% and with $F1_{Macro}$ of 91.5%.

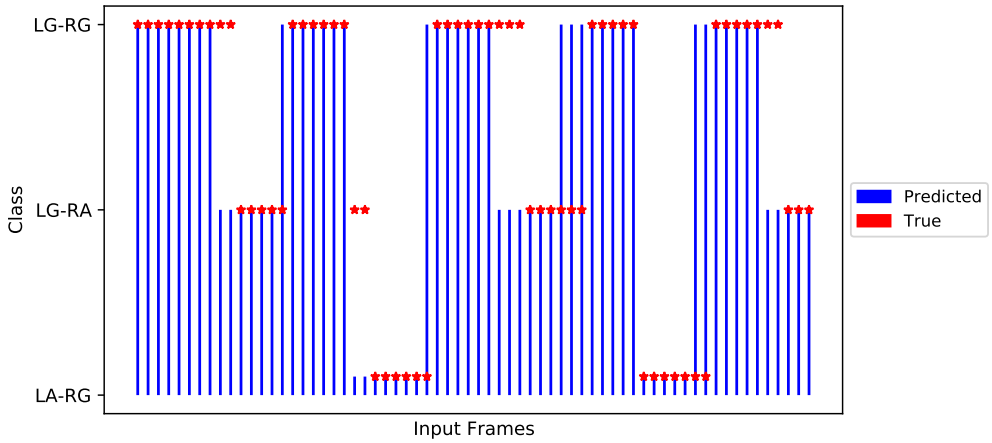


Figure 4.12: An illustration of the variation while predicting the labels for each frame of video using CNN-LSTM model, there is a variation while changing the feet position.

In the Table 4.5, the binary analysis of the classification with measures like ACC. PREC, SENS, SPEC, and F1-Score are shown. A comparison was made between the respective class and all others aggregated into one class. It means the table shows the classifier's ability to recognize the presence or absence of a specific gait event class among all classes of gait events. For comparison with small dataset i.e. gait dataset, we are also including the results of CNN-LSTM on gait dataset also in the table. The results show that, the overall accuracy

Table 4.5: Results of the classification performance of CNN-LSTM on the test data from gait dataset and mixed dataset with the measures of ACC, PREC, SENS, and F1-Score, rounded on 3 digits.

	Mixed Dataset					Gait Dataset				
	ACC	**PREC**	**SENS**	**SPEC**	**F1-Score**	**ACC**	**PREC**	**SENS**	**SPEC**	**F1-Score**
LA-RG	0.952	0.897	0.952	0.951	0.923	0.884	0.809	0.831	0.908	0.820
LG-RA	0.950	0.888	0.954	0.948	0.920	0.888	0.795	0.843	0.908	0.818
LG-RG	0.936	0.954	0.853	0.975	0.900	0.874	0.861	0.801	0.920	0.830

was improved by adding more data into the dataset. The accuracy in case of individual classes was also improved compared to the previous experiment.The highest accuracy was for classes LA-RG, and LG-RA with a value slightly above 95.0%. The class LG-RG was significantly better classified as per precision measure, with a value of 95.4% as compared to the other classes with values of 89.7% and 88.8% for LA-RG, and LG-RA respectively.

In terms of the , LG-RG showed good results with value of 97.3%, which is expected also because it is easy for classifier to deduct both feet on ground as compare to the other two cases. While the sensitivity was significantly low in this case with a value of 85.3%, where other two classes have above 95% sensitivity. F1-Score for classes LA-RG, and LG-RA was slightly high with value above 92% as compared to the LG-RG with a value 90%.The confusion matrix for CNN-LSTM model showing the heatmap is presented in Figure 4.13.

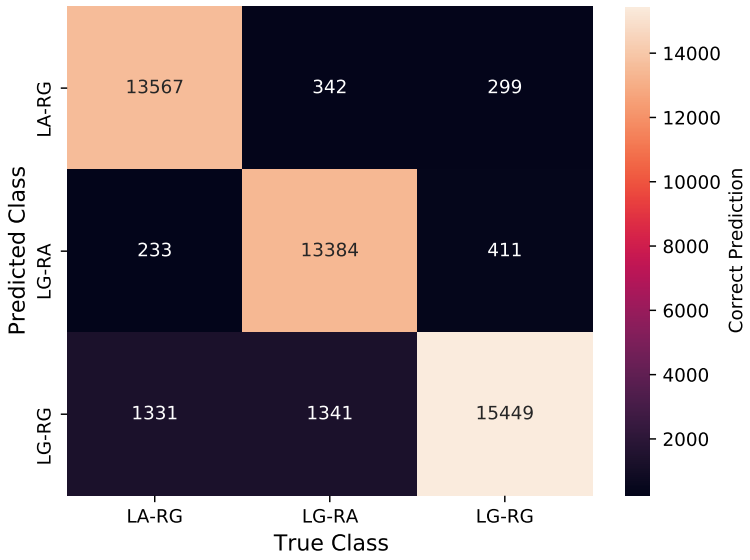


Figure 4.13: Confusion Matrix also showing heat map of CNN-LSTM classifier on test data (selected only from gait dataset).

The variations in prediction among classes was also reduced as compared to the previous approaches. There is no variation of 3 frames, the maximum variation is two frames. Like the previous models ,it was also specifically seen while changing the foot position from one phase to another. For example, if the previous position of both feet was on ground (LG-RG), then some of the initial frames of the next phase (either LA-RG, or LG-RA) are predicted as LG-RG. The same is while changing from the other two positions to the LG-RG position. These variations are shown in Figure 4.14.

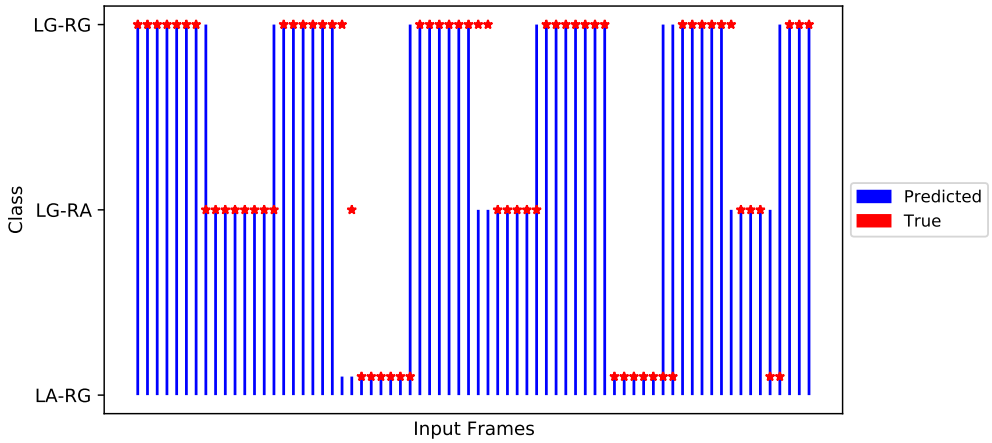


Figure 4.14: An illustration of the variations while predicting the labels for each frame of video using CNN-LSTM model on mixed dataset, there is a variation while changing the feet position.

4.7 Discussion

This study was aimed to propose a method for maker-free feet' position detection from marker-less video-based gait data outside laboratory conditions because, despite technological advances, we continue to encounter limitations when it comes to measuring human walking in clinical and laboratory settings. Using current gait analysis techniques remains expensive and time-consuming and makes it difficult to access specialized equipment and expertise. Therefore, it is imperative to have such methods that could give long-term data about the patient's health without any dual cognitive tasks or discomfort while using wearable sensors. Patients with neurological diseases occasionally have their gait movies recorded in the clinical setting. However, the lack of specific gait analysis equipment makes it difficult to analyze these movies quantitatively.

Therefore, one goal of this study was to have some primary video-based data without any markers or multiple cameras that could be utilized for gait analysis. As collecting video-based data requires certain regulations and approvals, getting gait data from hospitals at the start of the first study without proof of good results is difficult. Therefore, we collected the gait data from volunteers. One could argue that it may not be the original representation of patient health data because healthy individuals were used during the data collection. The counterargument is that the main objective in this study is to get the foot position respect to the ground, which are simple gait events. Once these events are captured, they can be utilized

for further analysis. Therefore, having video-based walking data is important regardless of the health condition. However, we asked the volunteers to walk in NPH condition so that we have some data other than normal walking. Having both types of data increase the generalization of the results. We also collected the data with shoes and without shoes while walking, which makes versatility in the dataset.

The other and main goal of the approach was to determine how computer vision methods could be utilized to detect feet's position from marker-less gait videos. In recent years these methods have shown very good performance in the field of human action recognition [190–192]. It has been proposed to use convolutional neural networks to study spatiotemporal human recognition [203], including operations to capture both spatial and temporal components [204]. A method for understanding action scenes has also been suggested [205]. Several studies have shown that two-stream convolutional networks are effective for recognizing spatiotemporal actions [190, 203].

Therefore, the purpose of this study was to employ these deep-learning approaches for feet's position classification with respect to the ground. This goal with an accuracy of 91% is also achieved; although there is a variation of 1-2 frames in the final model, we can argue that it could be ignored because of the high frame rate of 30fps while converting videos to images. Another important aspect is the chances of error in ground truth determination. If some systematic bias is inducted to overcome that error, then results may be more improved.

Let's look at the different experiments in detail. The first experiment in section 4.6.1 shows that the selected A-CNN (a CNN based on AlexNet) can separate and classify the individual classes of the gait dataset, but not with great accuracy. It is worth noting that we only used the architecture of the AlexNet and initialized all the weights randomly. So, no previous weights were transferred to this model. This model predicted the overall accuracy with 74%, but individual classes' accuracy was better, with the highest accuracy of 83.7% for the class LG-RA. The classes LA-RG and LG-RA were many times misclassified as class LG-RG. It is quite understandable because, after every swing phase of the foot, there is a stance phase in which both feet are on the ground. So stance phase contributes more to the overall gait cycle; hence the misclassification of the LA-RG and LG-RA classes to LG-RG class is quite natural. Further, we can see from the variations in Figure 4.10 that once the foot is stable in one position, the misclassification is not present, but switching from one event to the other event results in misclassification. The specificity of the class LG-RG was high as the most correct number of predictions were from this class, and we can observe this from the confusion matrix (Figure 4.9) also.

In the same experiment, adding the LSTM block to the CNN and making a CNN-LSTM hybrid model on the same dataset substantially improved the results. The ACCs of the

binary analysis for all classes was above 87 %. Also, the scores for the PRECs, SPECs, and SENSs, and F1-Scores are not unambiguous. Accordingly, there was a reduction in the variation in predictions among classes at the last stage of each gait event compared with the A-CNN classifier. In some instances, there was a variation of about three frames, but in most cases, it was only two frames (Figure 4.12). The results also indicate that the CNN based architectures are very good and powerful models in image classification, recognition, and prediction tasks. However, they are not as useful in this type of problem where temporal information is important. CNN can extract the features from the images very well, but once there is a temporal connection and correlation among the images, only CNN may not be a good approach.

In machine learning problems, the more data we have, the more predictions will be accurate; therefore, to test this, in the next experiment (section 4.6.2), the gait data was combined with the CASIA dataset, and the total dataset was significantly increased. The CNN-LSTM was trained with the large dataset but was tested only with the primarily collected data. We only selected the test data from the gait dataset because the CASIA dataset was captured using multiple cameras from different angles. It might be easy for the classifier to predict the foot position in that case. At the same time, in this study, the data was collected from just one camera from frontal and postern views (because we used the videos in the study collected with a smartphone only). The results were in line with the general assumption. We did not change the architecture of CNN-LSTM, and the hyper parameters were the same also so that we could compare. The ACCs of the binary analysis for LA-RG and LG-RA classes was above 95 %, while for LG-RG class, it was a little low. The SPEC value reached above 97% in case both feet are on the ground. Moreover, other performance measuring metrics also increased, which indicates the better classification ability of the classifier. Accordingly, there was a reduction in the variation in predictions among classes at the last stage of each gait event. There was no variation of three frames.(Figure 4.14).

Having demonstrated these performances of the proposed models on gait videos, it is pertinent that the temporal information encoded between the gait events may not be accurately detected by only CNN based models; therefore, combining the CNN based models with LSTM based model could potentially increase the classification of feet' position.
Although we did not perform any gait analysis in this study, we demonstrated how accurately we could detect the foot position from low-resolution marker-less videos. However, while comparing with some prior studies of gait analysis using camera-based technologies where the authors used OpenPose to investigate features of walking or other human movement patterns [163, 206–211], we anticipate that the methodology used in these studies may be capable of producing more accurate results, and the accuracy of the final model is a

little low. However, it is pertinent to note that marker-less video-based gait event detection (without pose estimation) is particularly applicable to clinic- and home-based gait analysis and requires no equipment other than a smartphone camera for data collection. In contrast, many other methods require expensive, inaccessible, and less portable equipment.

4.8 Conclusion

The purpose of this study was to develop a method to recognize the feet' positions based on gait videos in a simple and effective way in home-based conditions, using deep learning methods. We observed that deep learning methods could be used to estimate the feet' positions from marker-less videos captured under home conditions. Further, the gait events like HS and TO can be detected and used for a gait analysis system. We also conclude that, in the problems like in this study, where the sequence and timing of images are important, only CNN may not an efficient solution all the time. Thus, merging two or more models is an efficient alternative for problems like this.

Hybrid models are characterized by the integration of different architectures to improve efficiency by considering the advantages of each architecture. It is pertinent to note that when comparing the accuracy improvements obtained with the processing time of the LSTM and CNN-LSTM models, the time difference is not that significant, thus making the suggested architecture suitable for the detection of feet' positions in gait. It is imperative to note, however, that the time difference is dependent on a wide variety of factors, including the size of the data, the predicted time window, and the hardware being used. Based on the results of the experiments, the model (CNN-LSTM) was a better choice than the other models when a sufficient amount of data was available. According to the preliminary results, the proposed approach has a high potential for detecting gait events, thereby improving gait analysis methods outside of a laboratory setting. This method will be used further, by adding other regularization techniques for machine learning with more data in the next chapter.

Chapter 5

Quantifying Spatiotemporal Gait Parameters from Events Detected by Neural Networks Using Transfer Learning

This chapter explains the extraction of important gait parameters (which are further used for gait analysis) form the gait events detected by neural networks (NNs) applying transfer learning methods to feet's position classification. This chapter is divided into two halves, in the first part, the use of transfer learning methods is outlined, to improve the performance of the models which were developed and presented in chapter 4 of this thesis. In the second part, the methods to extract the spatiotemporal parameters from the gait events are presented and compared with the ground truth parameters calculated from videos, via observations. In the end, the chapter's summary is discussed and concluded.

5.1 Introduction

Measurements of locomotion patterns and variability can be obtained objectively and reliably through instrumented gait analysis. A targeted rehabilitation program can be developed based on these measures to investigate gait pathologies [212, 213]. Further, incipient neurodegenerative diseases can be detected or monitored using gait analysis [214, 215]. Gait disturbances have been demonstrated to be an early indicator of mild cognitive impairment (MCI) as well as a predictor of Alzheimer's disease (AD) progression [216]. Moreover,

gait performance can also be used to predict the likelihood of falling [217, 218], as well as mortality and morbidity [219, 220].

There is a tendency in elderly and diseased populations to walk more cautiously, as indicated by a reduction in walking speed and stride length, as well as an increase in step width [221]. Healthy adults' spatiotemporal parameters are commonly used to determine whether certain populations deviate from the norm. Furthermore, spatiotemporal parameters can be used to determine how aging and the disease affect walking patterns.

Generally, spatiotemporal parameters are considered to be key metrics while describing gait. It is possible to determine the level of impairment and to characterize functional gait performance using objective measures of the temporal and spatial parameters of gait [215, 217, 222]. To determine spatiotemporal parameters for each gait cycle, specific gait events must be identified. In terms of temporal gait parameters; stride duration, step length, and cadence are the most commonly used. The distance covered between two consecutive initial contacts (ICs) can be used to define spatial gait parameters (step and stride length).

Various sensing technologies, including foot switches, inertial sensors, pressure mats, and stereophotogrammetric systems, can enable the estimation of gait spatiotemporal parameters. With most of these technologies, there is a need for a controlled and dedicated environment, besides a lengthy set-up process, and a post-processing process. Although the integration of inertial sensors in handheld devices like mobiles and smartwatches has improved the situation, still, a limited number of parameters could be counted, and there is a chance of error. Further, these sensors take measurements based on body movement, and if a person is moving his/her hand or the place where the sensor is attached, the steps will be counted, which is not the correct representation of gait.

Gait parameters can also be assessed using optical motion analysis systems [223]. Some marker-based approaches have also been used for gait parameter detection from videos. In general, the process of marker-based gait analysis is not easily accessible to the general public due to the extensive laboratory setup and environment required. Therefore, this research was aimed at to quantify the spatiotemporal parameters of the gait from the gait events that were detected from marker-less videos using deep learning methods previously described in chapter 4 of this thesis. As compared to marker-based gait analysis, the marker-less gait analysis system is more user-friendly, portable, and simple.

Furthermore, as the accuracy and other performance metrics in the prediction of classes in the proposed final model for test data is almost 91%, therefore, benefits of transfer learning methods are leveraged to further improve the predictions of various gait events. Consequently, the new, improved models are utilized to detect the gait events to quantify the spatiotemporal parameters.

5.2 Related Work

We can define the level of impairment and characterize the functional gait performance by using objective measures of the temporal and spatial parameters of gait. Identifying specific gait events is essential for each gait cycle in order to calculate the spatiotemporal parameters. There are a number of various state of the art studies where diverse methods have been proposed for the estimation of these parameters during gait analysis. We have already presented some related work in section 4.2, and others are presented in this section.

Various spatiotemporal parameters as well as an automatic method for detecting gait events are presented in [224] and tested in real-life settings in the successive work by[225]. In another study, quantification of the spatiotemporal parameters of gait from magneto-inertial measurement units for healthy and diseased subjects was presented [23] In another research [216], the authors used multiple sensor-based systems for objective measurement of gait parameters using a dual-tasks paradigm. The gait parameters of 14 MCI patients, 6 AD patients, and 14 healthy control subjects were measured by using two actigraphs attached to each participant's waist in a study by [226]. Using three-dimensional motion analysis, the authors examined the spatiotemporal characteristics and margins of stability of one hundred and five healthy adults between the ages of 20 and 89 [227].

Sensor technologies have been used to monitor and assess motor behavior in elderly people in significant quantity [228]. In most studies of gait in individuals with MCI and AD, pressure-point systems [229–234] or passive infrared sensors [235] are utilized, which are not always affordable for all clinical sites [216].

It is also possible to extract spatiotemporal parameters using vision-based systems, which can be divided into two categories [236–238]. In the first method, markers or sensors are placed on the body of the subject in order to perform a model-based gait analysis. There are two types of markers: active and passive. The most accurate measurements can be obtained using active markers, which consist of visible light emitting diodess (LEDs). However, it is important to note that this approach has several disadvantages, including the need for a large and costly experimental setup and the fact that the markers placed on the subject's body impede his free movement. As a result, the user is unable to demonstrate a natural walking pattern [239].

In contrast to marker-based analysis, the second approach is holistic-based analysis, which is marker-less. It is executed in a marker-free environment. The video recordings of the subjects are captured using a video camera recorder. A silhouette image is then extracted using image processing techniques. The subject's parameters can then be analyzed

based on the model, the appearance, or the hybrid approach. Typically, these methods are used in the field of human activity recognition [240]. Active markers are being used in some clinical environments [241]. Although the results obtained through active markers are precise, but this technique is very time-consuming and complex. In [242], authors recently presented methods for recognizing human action in data streams that utilized the knowledge of seven pre-trained convolutional neural network (CNN) models to extract deep features for frame-level spatial information.

A marker-less video-based system capable of extracting the gait's spatiotemporal parameters is imperative in the light of the limitations of sensor-based and marker-based systems in terms of affordability, ease of use, and long-term recordings. Based on the gait events detected from videos using deep learning methods, this study presents a method for determining the spatiotemporal parameters of the gait.

5.3 Methods

The transfer learning methods are used to improve the performance of the classifiers presented in chapter 4 for better prediction of feet' positions. Further, the gait events detected from this classification are used to obtain the gait parameters like gait cycle, step length, cadence, etc. In this section, the proposed methodology for data collection, transfer learning, and consequently the parameters extraction is discussed.

5.3.1 Data Collection

To extract the spatiotemporal parameters from gait events and subsequent use of those parameters for gait analysis require primary data. As in the first study (chapter 4), data was collected in normal and Normal Pressure Hydrocephalus (NPH) walking conditions; another round of gait data collection from volunteers in 4 different conditions was performed. Different walking conditions in this round were used to make the dataset more versatile and large. These 4 conditions include normal walking, and 3 simulated abnormal conditions, representing walking in AD, walking in NPH, and walking with some injury. For simplicity, we will refer these walking conditions in this thesis as **normal walking (**$Norm\_Walk$**),alzheimer's disease walking (**AD_Walk**), NPH walking (**NPH_Walk**), and injury walking (**$Injury\_Walk$**)** respectively. In the $Injury\_Walk$, the subjects were dragging one foot while the other foot was performing normally. For further details about different

diseases for which gait disturbance is one of the first symptoms, the reader is referred to the section 2.6 of this thesis.

In this study, a similar approach was followed, as in the previous study, but the location and background were changed. It was deliberately done to have data with different home conditions and scenarios. The walking pathway of 10 meters suggested by [194] was used. The 10-m walk test is a clinical test that measures the length of time it takes for a patient to walk 10 meters[243]. Despite the need for clear and standardized instructions to ensure accurate data during walking and appropriate clinical decisions, it has been validated in a variety of conditions and is generalizable to clinical practice [243].

The total length of the observational area was known as we wanted to get some gait parameters by observation also so that we could have ground truth and compare the results with those extracted from the machine learning approach.

A group of 25 subjects (18 males, 7 females, age: 29 ± 5 years, height: 175 ± 13 cm, and weight: 70 ± 15 kg) participated voluntarily in this research and reported no gait abnormalities. The participants were not asked to participate in the study if they had any health conditions that might affect their ability to walk normally, typically a history of falls within the past six months prior to enrolling in the study. Some statistical information such as gender, height, weight, and age was also captured to allow further analyses.

The study consisted of filming the patient walking on a 20 m path (10 m in one direction, turning, and then 10 m back to the starting point) during four trials, totalling 80m of the walking (Figure 5.1). The subjects were asked to walk and were also reminded with instructions during the experiment, these instructions were: “start from this point,” “follow this path,” “go to the location indicated,” “turn around,” and then “return to the starting point.”

The purpose and method of the experiment were explained to all participants prior to the experiment. All methods were carried out in compliance with guidelines and regulations issued by the KIT Ethics Commission, and the KIT Ethics Commission granted ethics approval. The participants were also required to provide their informed consent to participate in this study.

The subjects were not required to wear special clothing or equipment but were encouraged to wear casual clothing and shoes to achieve a more natural appearance. Prior to the measurements, the subjects had trained with the diseased walking tasks until they felt comfortable performing the desired gait pattern. The subjects were instructed to perform four walking tasks during the measurements, as discussed above. In the case of some subjects who were unable to emulate the desired gait in one attempt, the recording process was repeated.

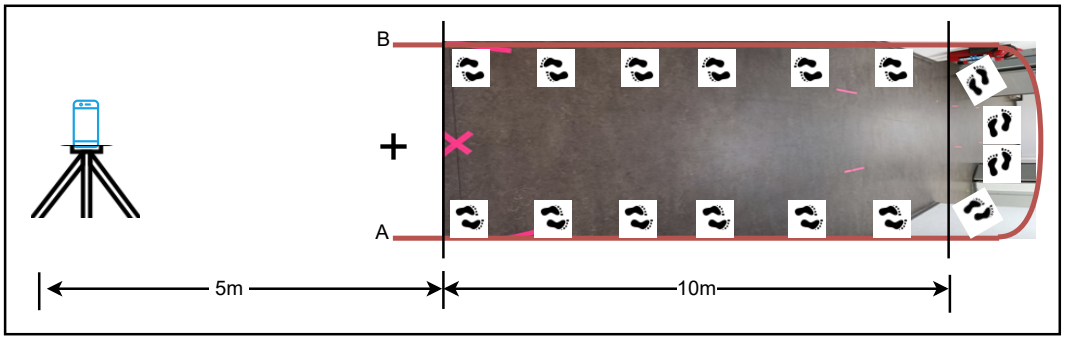


Figure 5.1: The walking pathway where the data was collected from volunteers in *Norm_Walk*, *AD_Walk*, *NPH_Walk*, and *Injury_Walk* conditions. A and B are the start and end points respectively, and the subject was standing at (+) sign before the start of walking. The mobile phone camera was used to record the walking from the center of the corridor and 5 m away from the starting area, the symbols in the pathway were used to have ground truth data regarding different parameters.

An android-based smartphone was used to record the subjects' movements from the front-sagittal viewpoint, as shown schematically in Figure 5.1. A tripod was used to fix the smartphone at a fixed location at a standstill position during recording.

At the end, we had a total of 100 videos of walking in 4 different conditions. Additionally, we also used the previously collected data to test the benefits of transfer learning for gait event detection and extraction of spatiotemporal parameters from those events. We will refer to this dataset as *N2AI* dataset which stands for *Norm_Walk*, *NPH_Walk*, *AD_Walk*, *Injury_Walk* dataset. Subsequently the videos were processed to enhance the visual quality, and to convert into images at a frame rate of 30fps along with resizing of images. Data annotation and labelling was done to determine the ground truth as discussed in the previous chapter 4.4.2.The walking time required by each subject in every task is reported in Table 5.1.

5.3.2 Transfer Learning and use of Pre-trained Deep CNN

The performance of deep CNN is usually superior to that of large databases when compared to small databases. Due to the availability of the learned kernels and weights that are available for the public to use, the already-trained CNN models are easily accessible for public use for a variety of computer vision tasks, such as object recognition [244], classification [245], segmentation [246], and detection [247]. Typically, CNN models are trained using large datasets to classify one thousand different types of objects. However, these models are not stable with small datasets due to the lack of large-scale data for model learning; therefore, model learning is a key issue when investigating relatively small datasets like in this study.

Table 5.1: Walking time (in seconds) required by each subject in 4 different conditions Norm_Walk, AD_Walk, NPH_Walk, and Injury_Walk, to complete 20 meter walk

Subject#	Walking Time (s)			
	Norm_Walk	AD_Walk	NPH_Walk	Injury_Walk
1	18	22	25	21
2	15	20	19	23
3	16	44	58	64
4	18	39	41	47
5	18	27	30	32
6	21	34	51	42
7	15	33	28	27
8	18	43	37	36
9	18	31	28	40
10	20	26	33	31
11	19	26	56	49
12	19	30	45	47
13	16	33	38	36
14	19	31	52	45
15	18	36	39	32
16	18	22	42	32
17	16	34	48	42
18	20	37	42	38
19	15	36	47	43
20	17	30	23	25
21	15	23	32	35
22	23	33	40	38
23	19	32	33	36
24	20	30	37	40
25	19	31	32	41

In this scenario, transfer learning is utilized with already learned weights, and re-training on the new dataset to extract the features and predictions is obtained using the fully connected layer. This method is outlined in Figure 5.2. It is common to apply transfer learning to small-scale dataset problems where a model has already been trained on a very large dataset, for example, ImageNet [103, 248]. Re-training an already trained model on new data is a much faster and more appropriate procedure than training a model from scratch. In order to acquire deeply learned information, a single round of deep feature extraction is required on

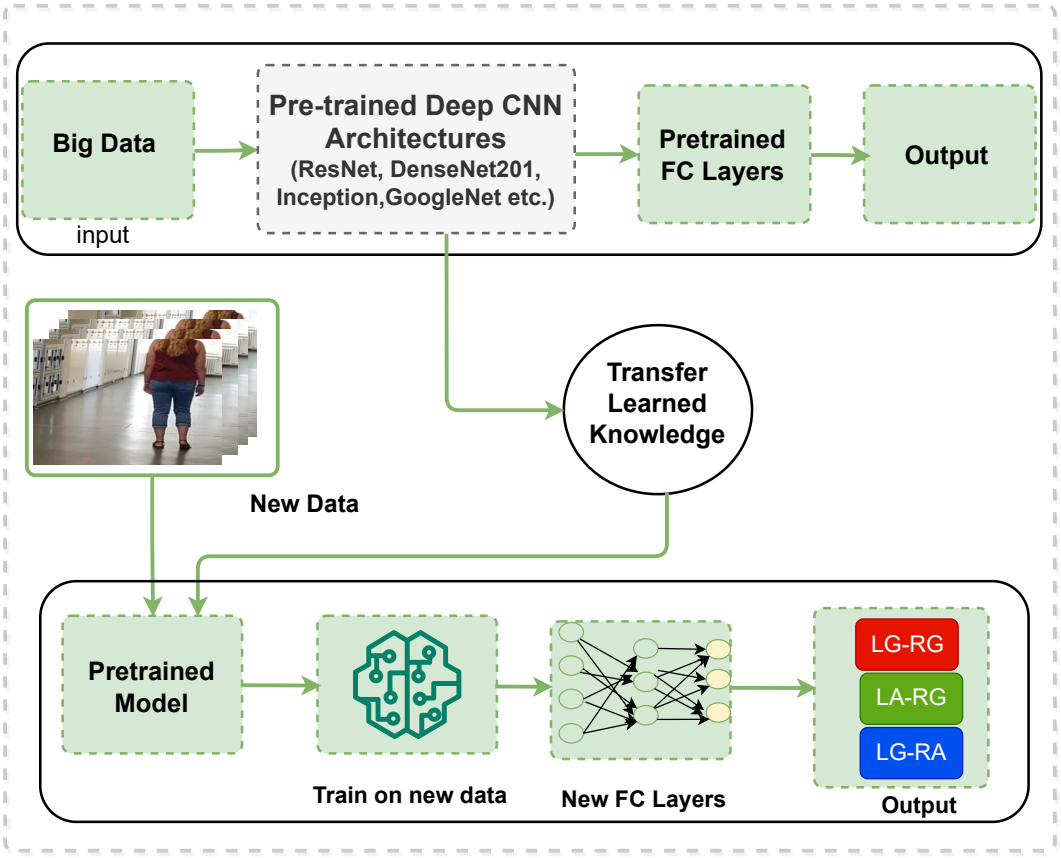


Figure 5.2: A schematic view of transfer learning where pre-trained NN are utilized to get learned weights and retrain on small scale dataset to get predictions

the training samples. In this study, small datasets are investigated for gait event classification and these datasets are more challenging due to the complex and diverse appearances, different backgrounds, and distinct clothes and shoes in the images. Thus, two already-trained CNN deep models, inceptionresnetv2 (IRNV-2) and densenet201 (DN-201) are selected and utilized as deep feature extractors. There are two reasons to choose deeper networks: (1) these networks can extract abstract representations of input at each layer, and (2) they have the ability to learn discriminative information that is significantly expressive and general. These models have already been used in previous studies like [197]. The description of selected IRNV-2 and DN-201 deep CNN models is given as follows.

InceptionResNetV2: In this case, the CNN model employs 164 layers deep which have been trained on a large number of images collected from ImageNet [249]. It is a hybrid model that combines the inception structure with residual association rules. This model utilizes images of dimensions 299 x 299 x 3, and it provides learned information and an estimated value for each class. A major benefit of IRNV-2 is the conversion of inception

modules to residual inception blocks as well as the addition of inception modules, and adding a new type of inception module (inception-A) after the stem module. A schematic view of InceptionResNetV2 architecture (compressed) is shown in Figure 5.3.

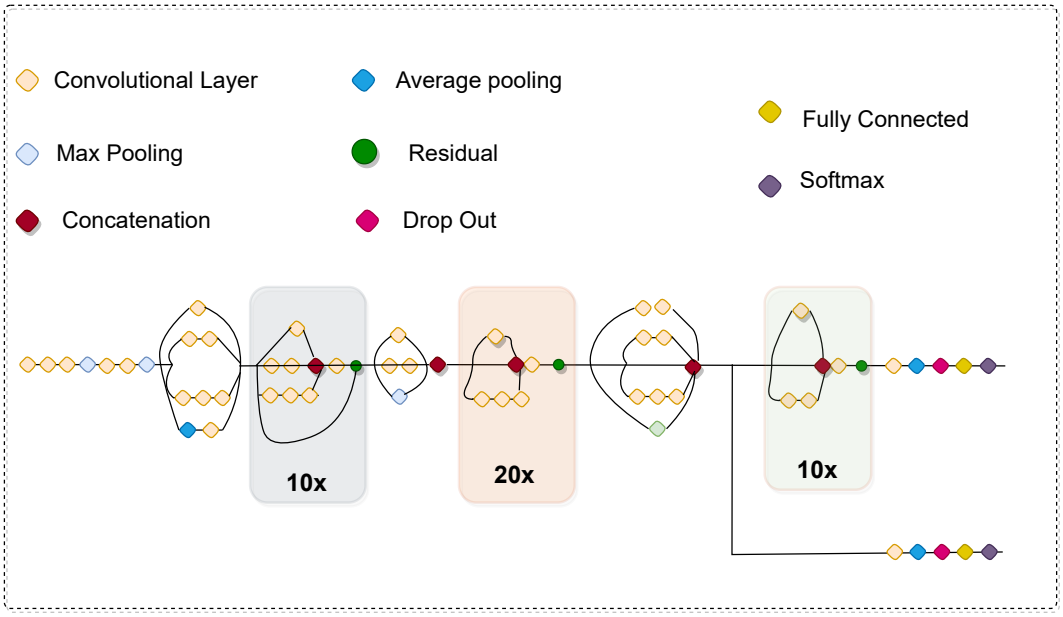


Figure 5.3: A block representation of InceptionRestNetV2, a pretrained model used for transfer learning. Figure reproduced from [197]

DenseNet201: It is a model with 201 deep layers, assumes an input size of 224x224x3, and has been evaluated on the following databases: SVHN, CIFAR-10, CIFAR-100, and ImageNet [250]. This network was designed to achieve deep and wide architecture based on CNN models that can be useful to enhance the performance of deep CNNs. Thus, DN is an advancement over ResNet (RN) that consists of dense connections between layers to transfer collective knowledge to allow features to be reused. In this way, the network layer obtains information from all preceding layers and sends it to all subsequent layers, which facilitates a maximum flow of information from one layer to the next, as well as enables feature reuse. In contrast to convolutional networks with l layers having l connections, DenseNet (DN) has $\frac{l(l+1)}{2}$ direct connections. Furthermore, DN can enhance performance by eliminating gradient problems, implicit deep supervision, model compactness, and reduction of parameter counts. Figure 5.4 illustrates the DenseNet201 architecture schematically.

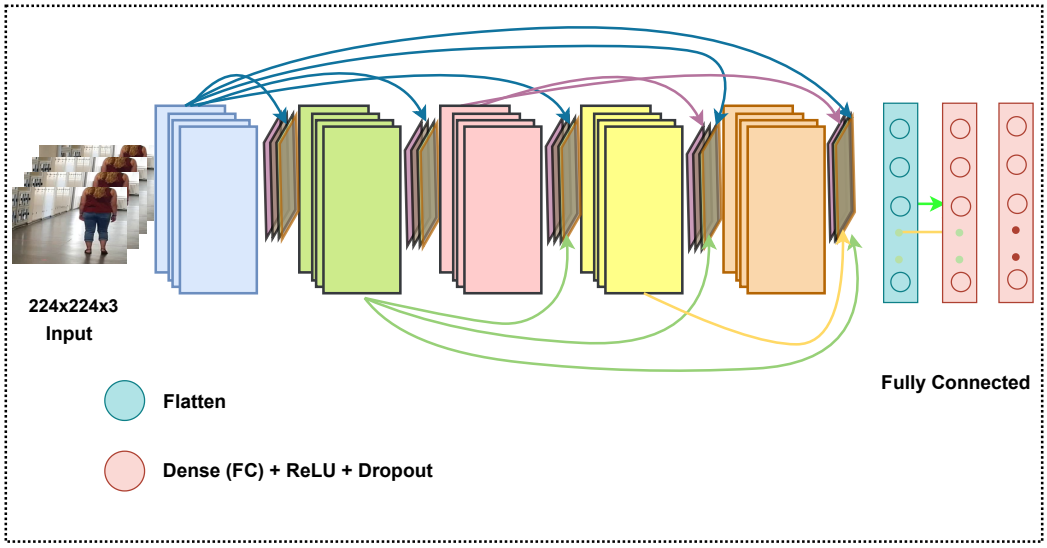


Figure 5.4: A block representation of Densenet201, a deep model with 201 deep layers trained on SVHN, CIFAR-10, CIFAR-100, and ImageNet databases [242].

5.3.3 Features Selection

The pre-trained deep NN extract features of high dimensions and ignore irrelevant information unrelated to robust modeling. Moreover, this irrelevant information not only decreases the classifier's overall performance but also increases computational cost. Therefore, one of the important factors in enhancing classification rates is the accessibility and utilization of distinct features from the extracted feature vector. Existing studies have applied various feature selection techniques for dimensionality reduction, including canonical correlation analysis (CCA), entropy, and principal component analysis (PCA) etc. As a result of these techniques, optimal features subsets (OFS) can also be selected from a large feature vector by discarding irrelevant information. Consequently, few techniques have been applied to compute the relationship between two or more representations of the same feature. In order to improve classifier performance without deteriorating computed features, it is a crucial step. This study chooses to use PCA and entropy-based features selection methods for the above mentioned reasons. Entropy and PCA are straightforward methods with the advantage that PCA minimizes reconstruction error without supposing the utilization of a diminished feature vector. As a result, we examine the contribution of each applied feature vector for the prediction of gait events using PCA and entropy-controlled feature vectors.

5.3.4 Training Strategies

The long short term memory network (LSTM) block was added to both models (IRNV-2, and DN-201) as described previously, to keep the temporal information which have shown good results in the previous study presented in chapter 4. Both models of the neural network were re-trained with the same training strategies to evaluate both networks' performance for predicting gait events. The same strategies were used to avoid any effect of the training method on the classifier's performance. Furthermore, we normalized the images over the complete dataset. For the training, we divided the total dataset as 60% for training, 20% for validation, and 20% for testing. We used both the datasets (gait dataset (including augmented data) and N2AI (*N2AI* dataset)) for the training of the NNs. The data augmentation methods were used on *N2AI* dataset to reduce the class imbalance in the dataset. The total number of images was 299527(144240+155287), out of which 197717 images were used for training, while validation and test images were 59905 each. The maximum number of training epochs we used was 100.

The training/validation accuracy and prediction performance of this approach will be discussed in the results section of this chapter.

5.3.5 Spatiotemporal Parameter Extraction

To analyze the gait, we need certain spatiotemporal parameters which are the basic building blocks of any gait analysis system. These parameters are already discussed in section 2.3, and here we will present the methodology to extract those parameters from the gait events predicted by the NNs.

As discussed earlier, the sequence of the input frames was considered while obtaining predictions. Therefore the predictions were also in sequence, i.e., the $Pred_1$ will be the label of F_1, $Pred_2$ will be the label of $F_3, \ldots\ldots, Pred_n$ will be the label of F_n in the image. Figure 5.5 shows this process schematically.

Once we have a series of labels in a sequence, here we are describing a new method of quantification of different parameters based this sequence. This method has not been reported in literature as per best of author's knowledge.

Gait Cycle: The first and important parameter is the gait cycle (GC) or stride interval (SI), the time between two consecutive same gait events of the identical foot. Here we are taking heel strike (HS) as the main event. We explain this with the help of the following example:

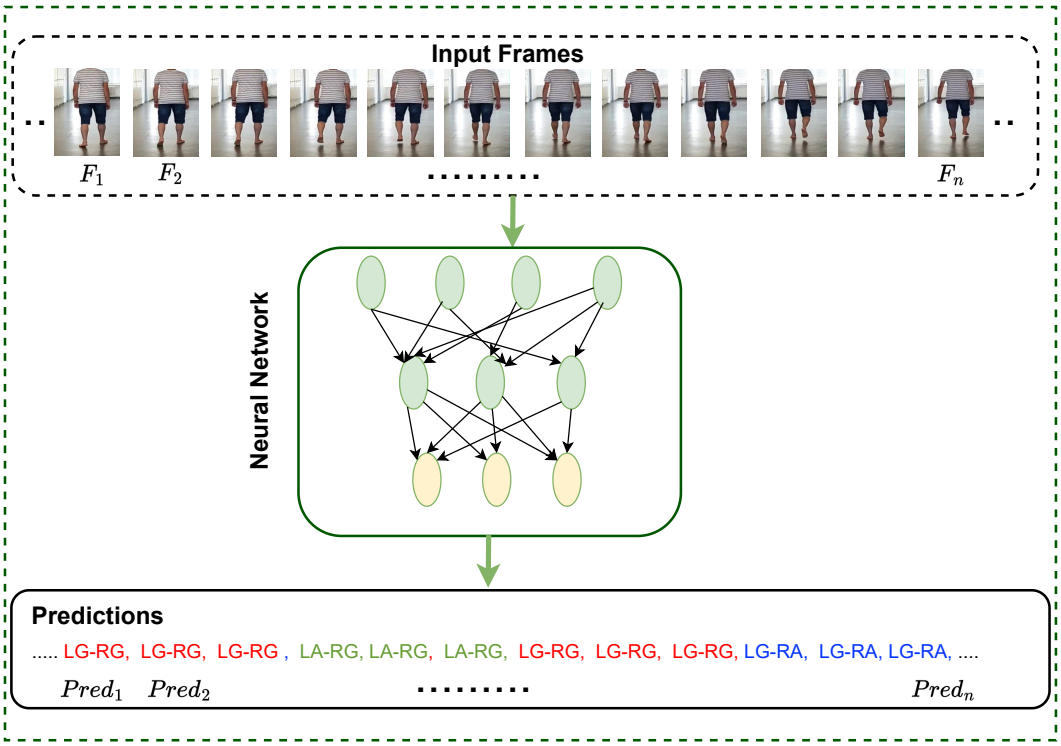


Figure 5.5: Process of getting the predictions for the input frames, for each frame we get one of the labels (LA-RG, LG-RA, or LG-RG)

The predicted label of frame F_t is LA-RG, and F_{t+1} is LG-RG; this implies the left foot, which was in the air previously, now touched the ground, so this is the heel strike (HS) of the left foot. Now both feet are on the ground, so it is (double support time). Next, the right foot will be lifted, and the labels of the frames will be LG-RA (swing time of right foot); once the right foot completes the step, it will touch the ground (right HS) and the labels will be LG-RG, please note during this time, the left foot is on the ground. Now left foot will leave the ground, changing the labels to LA-RG. When the left foot completes the step and touches the ground, changing the labels again to LG-RG, it will be the left foot's next HS. We assume this HS occurs at frame F_{t+15}, so the number of frames from F_{t+1} to F_{t+15} will be the frames in this gait cycle of the left foot (Figure 5.6).

Now we can calculate the left foot's GC duration/time by Equation (5.1).

$$GC_L = \frac{F_{No}(LHS_n) - F_{No}(LHS_{n-1})}{\text{fps}} \tag{5.1}$$

Step Time: The step time is the time between the same event of opposite feet. For example, the predicted label of frame F_t is LA-RG, and F_{t+1} is LG-RG; this implies the left foot, which was in the air previously, now touched the ground, so this is the HS of the left foot. Now both feet are on the ground. Next, the right foot will be lifted, and the labels of the frames

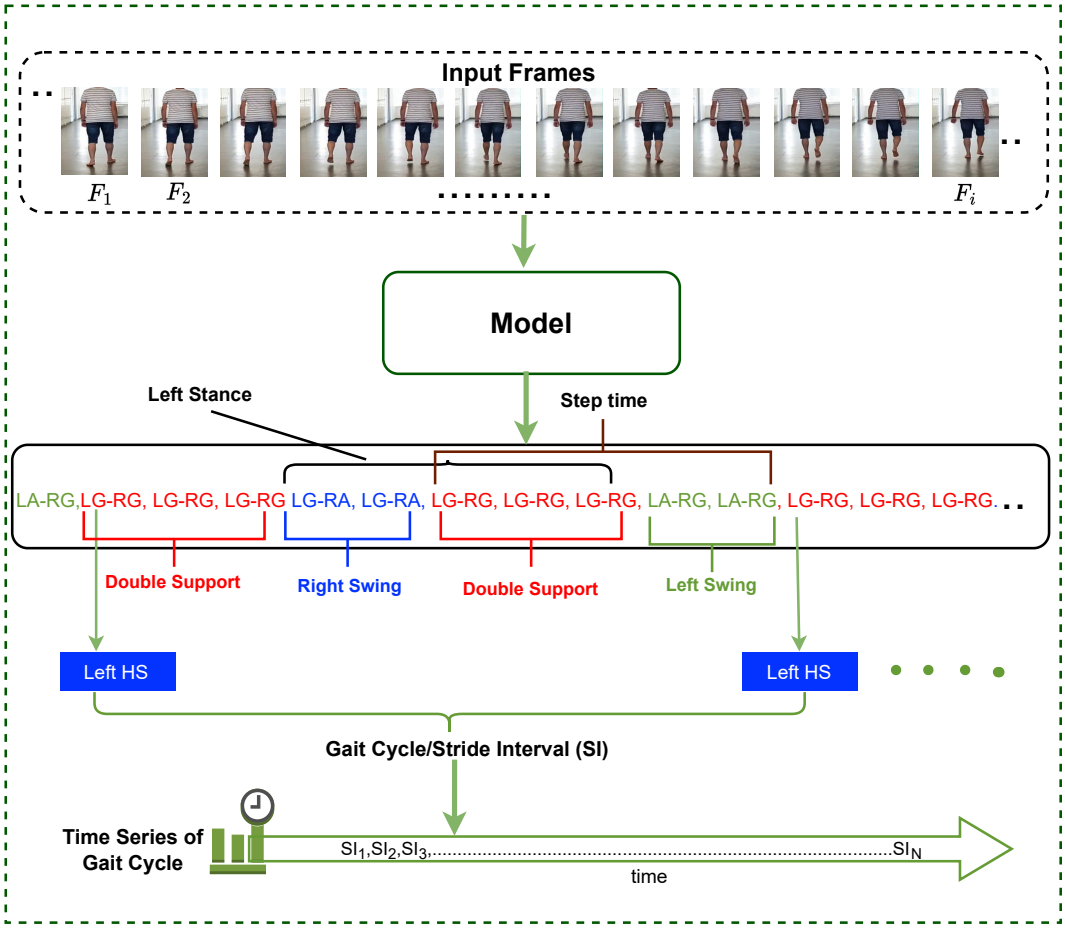


Figure 5.6: An illustration of proposed method to deduct different spatiotemporal parameters (GC,HS, swing time, stance time, double support) of the gait from the labels predicted by the NN; subsequent time series formation is also provided.

will be LG-RA; once the right foot completes the step, it will touch the ground (right HS and the labels will be LG-RG. The number of frames between these two events will be in the step-time of the right foot (Figure 5.6). And step time is calculated as (Equation (5.2)):

$$ST_R = \frac{F_{No}(RHS_n) - F_{No}(LHS_n)}{\text{fps}} \tag{5.2}$$

Swing (%): This is the period of time during which the foot is not in contact with the ground but in the air (Figure 5.6). It is calculated as (Equation (5.3)):

$$SwT_R(\%) = \frac{\sum_{F=1}^{N} Pred(LG-RA)}{\text{fps}} \times 100 \tag{5.3}$$

Stance (%): The phase of the cycle in which the foot is not in contact with the ground is called the stance phase of the cycle (Figure 5.6). It is calculated according to (Equation (5.4)):

$$StT_R(\%) = \frac{\sum_{F=1}^{N} Pred(LA-RG \vee LG-RG)}{\text{fps}} \times 100 \tag{5.4}$$

Double support (%): It is the time when both feet are placed on the ground during gait cycle(Figure 5.6). It is calculated according to (Equation (5.5)):

$$D.Support(\%) = \frac{\sum_{F=1}^{N} Pred(LG-RG)}{\text{fps}} \times 100 \tag{5.5}$$

Walking Time: It is the total time required to complete each walking task in minutes. It is calculated as:

$$W_{time}(min) = \frac{Count(TotalFrames)/\text{fps}}{60} \tag{5.6}$$

Cadence: It is the number of steps taken per minute is called cadence and is calculated as:

$$cadence = \frac{Count(HS)}{W_{time}} \tag{5.7}$$

Gait Velocity: It is defined as the distance traveled per second (m/s) and is given by:

$$vel = \frac{distance}{Count(TotalFrames)/\text{fps}} \tag{5.8}$$

Similarly other parameters will be calculated as follow:

$$Average\ Stride\ Length(m) = \frac{distance}{Count(GC)} \tag{5.9}$$

$$Average\ Step\ Length(m) = \frac{distance}{Count(HS)} \tag{5.10}$$

Gait Cycle Time Series: We can construct the time series by arranging them in a time-based manner after obtaining all the SIs from the predicted labels (Figure 5.6). This time series and other parameters can then be utilized for proper gait analysis. The time series-dependent gait analysis will be presented in the next part of this thesis.

Error Estimation:

The errors in classifying the labels of feet' positions will eventually affect the estimation of spatiotemporal parameters. The higher the variation in the prediction of frame labels, the higher the error will be in quantifying these parameters. As discussed earlier, the most relevant parameter in any gait analysis system is the gait cycle duration or stride interval. It

is very difficult to quantify the exact error in such a case, however, a systematic error or bias in the quantification of gait cycle duration could be calculated as follows:

$$e = \frac{v}{\text{fps}} \times \frac{t}{GCs} \tag{5.11}$$

where v is the variation in the frame's classification, t is the average stride interval duration and GCs are the total number of gait cycles. It could be observed that a higher frame rate will reduce systematic error.

5.4 Results

The result section is also discussed into two parts. In the first part, the performance of the models of transfer learning based gait event detection is presented, and in the second part, the results regarding extraction of spatiotemporal parameters from those gait events are discussed.

5.4.1 Performance of IRNV-2 and DN-201

Two experiments were conducted in which the proposed methods for the classification of feet's position were evaluated in terms of performance. In the first experiment, we used the dataset to retrain the IRNV-2 and calculated its performance on the dataset. In the second experiment, the DN-201 was trained with the same dataset and performance was evaluated. Both scenarios are discussed in the following sections.

The dataset (gait dataset +$N2AI$ dataset) we used, had a total of 299527(144240+155287) images. Out of this dataset, 197717 (60%) images were used for training, and 59905 images were used for each validation and testing. We performed re-training of IRNV-2 and DN-201classifiers.

Figure A.4 (a) shows the training and validation accuracy of the IRNV-2 model. The corresponding training and validation loss is shown in Figure A.4 (b). The training has risen up to 100 epochs with training accuracy (ACC) of 96.2%, validation accuracy of 92.9%, and validation loss of 0.11. In the case of the DN-201 model, the minimum validation loss was: 0.10, and maximum accuracy achieved was 97.1%, with validation accuracy of 93.8%. The results are shown graphically in Figure A.5(a,b).

The feature visualization process transforms the internal features of an image into visual patterns that can be perceived or recognized. Consequently, we can understand the learned features explicitly. In Figure 5.7, we visualize the feature maps extracted from the

convolutional layer of the IRNV-2 model to understand what input features are detected or preserved. It is generally expected that feature maps close to the input will detect small or fine-grained details, whereas feature maps near the output will capture more general features.

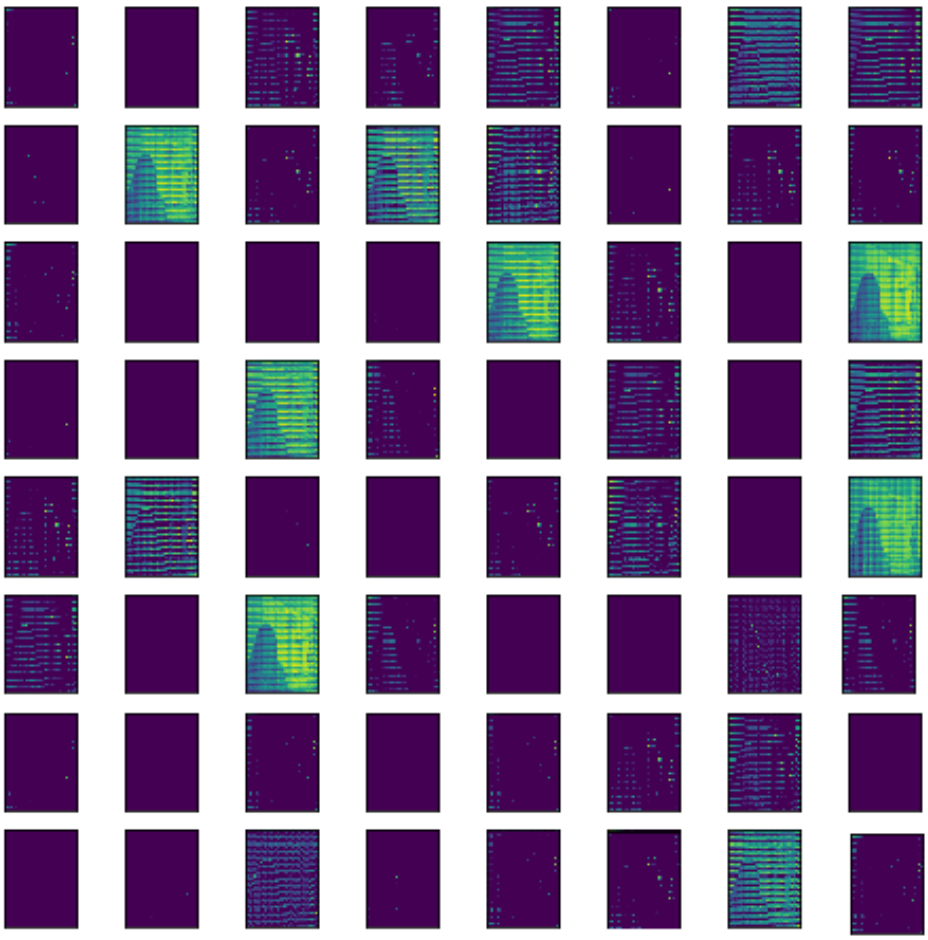


Figure 5.7: A visual representation of the feature maps extracted from the convolutional layer of the IRNV-2 model

Figure 5.8, visualizes the feature maps extracted from the convolutional layer of the DN-201 model to understand what input features are detected or preserved. It is observed that this model is detecting better and more relevant features as compared to the previous model, especially the features in lower filters. To measure and evaluate the performance of the discussed models on the test data, multiple metrics were used. We selected some measures of the overall metrics for multi-class datasets including confusion matrix, ACC, precision (PREC), sensitivity (SENS), F-1 score (F1-Score), specificity (SPEC), weighted average

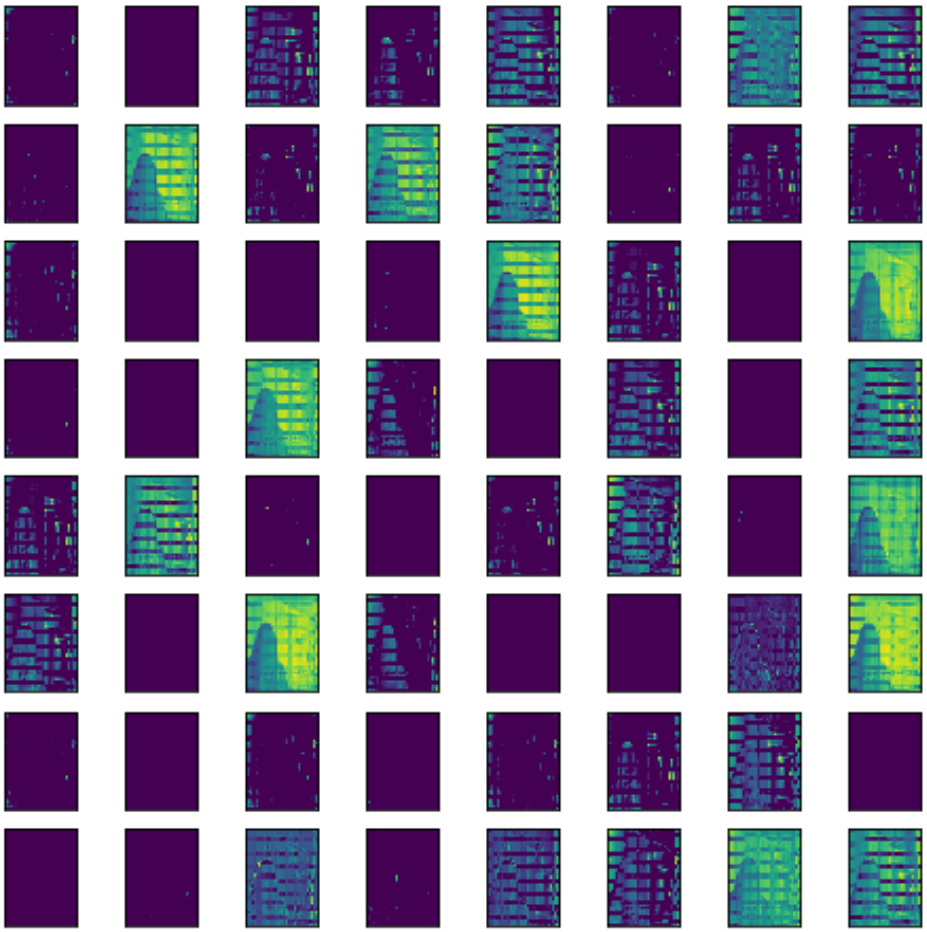


Figure 5.8: A visual representation of the feature maps extracted from the convolutional layer of the DN-201 model

precision ($PREC_{WAvg}$), weighted average recall (REC_{WAvg}), micro F-1 score ($F1_{Micro}$), and macro F-1 score ($F1_{Macro}$).

IRNV-2: The overall ACC of the IRNV-2 model was 92.5%. The $PREC_{WAvg}$ was 92.3% and REC_{WAvg} was 92.1%. The classifier predicted the classes with $F1_{Micro}$ of 92.2% and with $F1_{Macro}$ of 92.2%.

In the Table 5.2, the binary analysis of the classification is carried out; ACC PREC, SENS, SPEC, and F1-Score are shown. A comparison was made between the respective class and all others aggregated into one class. It means the table shows the classifier's ability to recognize the presence or absence of a specific gait event class among all classes of gait events.

The IRNV-2 model predicted the individual classes with a prediction ACC of 94.7%, 94.7%, and 92.8% for classes LA-RG, LG-RA, and LG-RG respectively. The PREC value

Table 5.2: Results of the classification performance of IRNV-2 and DN-201 on the test data from gait dataset with the measures of ACC, PREC, SENS, and F1-Score, rounded to 3-decimal places

	IRNV-2 Model					DN-201 Model				
	ACC	**PREC**	**SENS**	**SPEC**	**F1-Score**	**ACC**	**PREC**	**SENS**	**SPEC**	**F1-Score**
LA-RG	0.947	0.889	0.947	0.947	0.917	0.961	0.921	0.960	0.962	0.940
LG-RA	0.947	0.898	0.939	0.951	0.918	0.956	0.912	0.953	0.957	0.932
LG-RG	0.928	0.945	0.857	0.970	0.899	0.943	0.957	0.885	0.977	0.919

was 94.5% for LG-RG, almost 88.8% for classes LA-RG, and LG-RA. In contrast, the SENS value was smaller for class LG-RG than other two classes with a value of 94.7% and 93.9%. Regarding specificity, class LG-RG was predicted better than the other two classes with a value of 97%. The confusion matrix model showing the heatmap for IRNV-2 model is presented in Figure 5.9. The variations in prediction among classes were about 1-2 frames

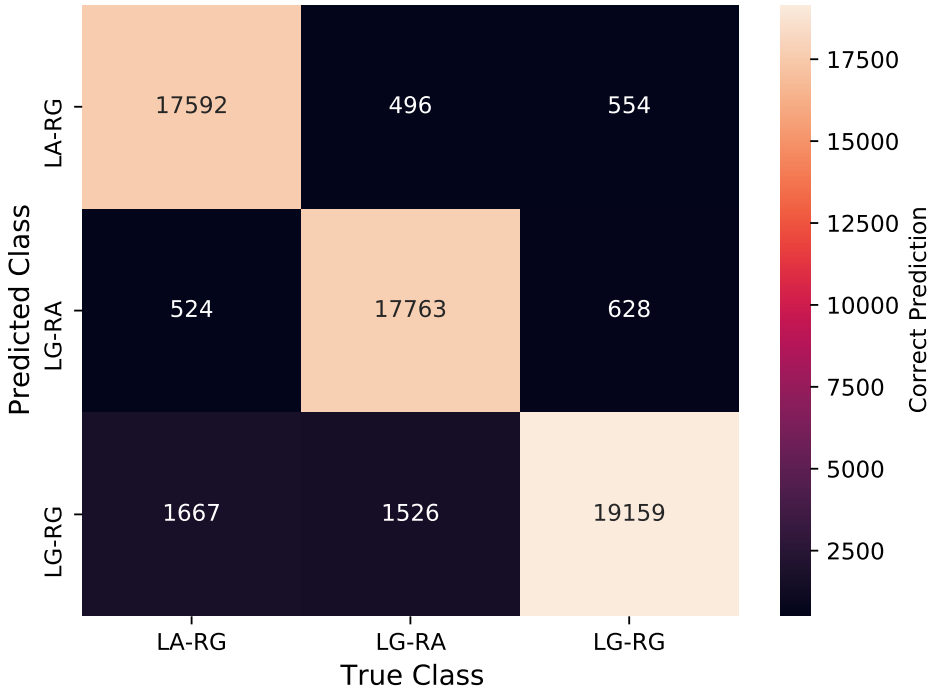


Figure 5.9: Confusion Matrix also showing heat map of IRNV-2 classifier on test data of gait dataset

and with almost equal in numbers. This variation was specifically observed while changing

the foot position from one phase to another. For example, if the previous position of both feet was on ground (LG-RG), then some of the initial frames of the next phase (either LA-RG, or LG-RA) are predicted as LG-RG. The same is valid while changing from the other two positions to the LG-RG position. These variations are shown in Figure 5.10.

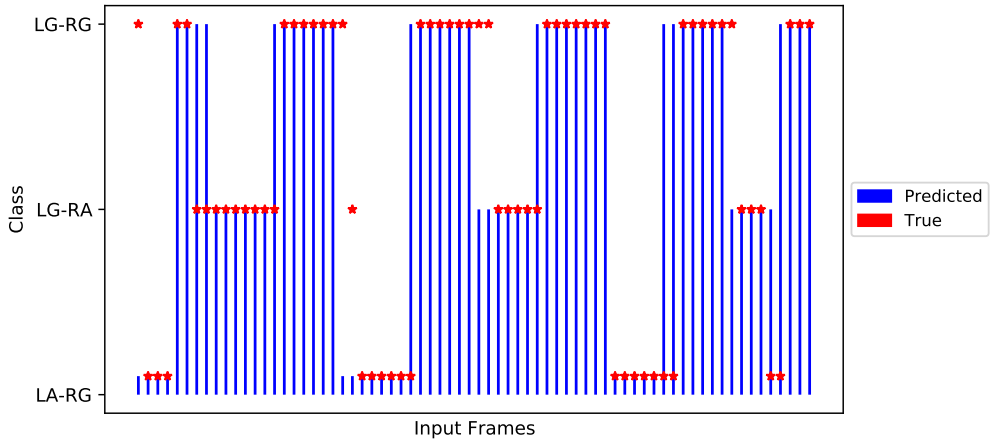


Figure 5.10: An illustration of the variations while predicting the labels for each frame of video using IRNV-2 model, there is a variation while changing the feet position

DN-201: The DN-201 model showed a little better performance in terms of accuracy and other measures compared to IRNV-2. This model predicted the true classes correctly with an overall ACC of 93.8% and an mean accuracy (AvACC) of 95.4%. The $PREC_{WAvg}$ was 93.2% ,REC_{WAvg} was 93.1%, $F1_{Micro}$ was 93.1% , and $F1_{Macro}$ was 93.5%.

Further, we also evaluated the binary classification for the CNN-LSTM model. The results are shown in Table 5.2, where binary analysis of the classification with measurements like ACC. PREC, SENS, SPEC, and F1-Score are shown. A comparison was performed between the respective class and all others aggregated into one class.

The accuracy in the case of individual classes was also improved compared to the other model. The highest accuracy was for class LA-RG, with a value of 96.15%, while other classes' accuracy was 95.60%, and 94.4%. In the case of precision, the class LG-RG with a value of 95.8% was significantly high compared to the other classes with values of 92.0% and 91.2%. A similar trend was observed for the specificity with values 97.69%, 96.17%, and 95.75% for classes LG-RG, LA-RG, and LG-RA respectively. However, the sensitivity value of class LG-Rg was lower compared to other classes, with a value of 88.47%. F1-Score for classes LA-RG, and LG-RA was 94.02% and 93.19% respectively. The confusion matrix for

DN-201 model showing the heatmap is presented in Figure 5.11. The variations in prediction

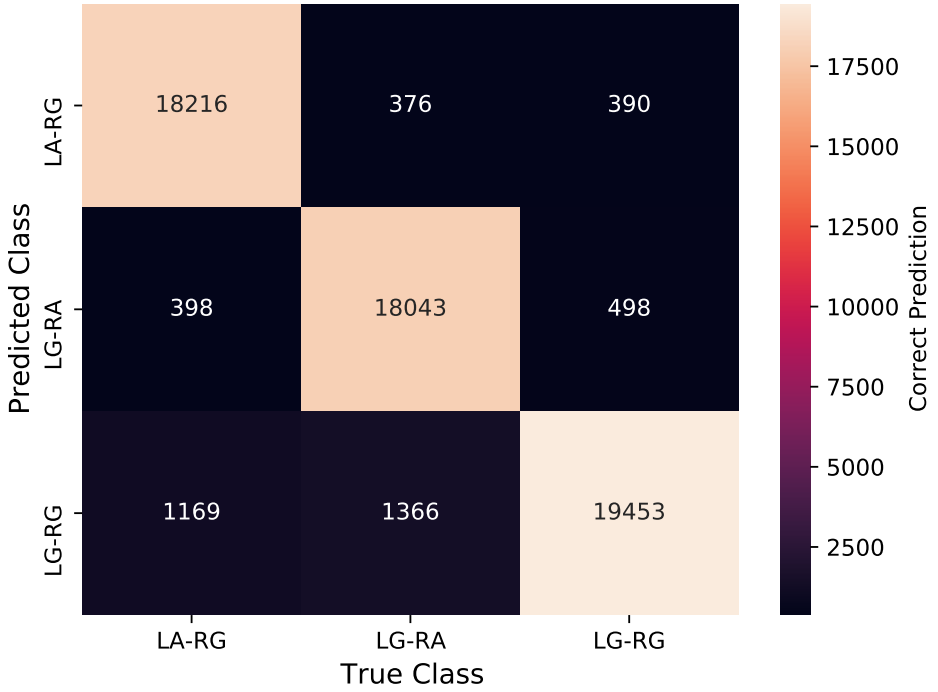


Figure 5.11: Confusion Matrix also showing heat map of DN-201 classifier on test data of gait dataset

among classes were also reduced as compared to the IRNV-2 classifier. At some points, there is a variation of about 2 frames but mostly within one frame variation. Like the previous model, it was also specifically seen while changing the foot position from one phase to another. For example, if the previous position of both feet was on the ground (LG-RG), then the initial 1 or 2 frames of the next phase (either LA-RG, or LG-RA) are predicted as LG-RG. The same is valid while changing from the other two positions to the LG-RG position. These variations are shown in Figure 5.12.

5.4.2 Validity of Gait Parameters Extraction

By incorporating the methodology presented in section 5.3, we calculated 12 gait parameters from the predicted labels of the NN. As suggested by [251], to compare the data of the same parameter obtained by the NN based method referred to as $predicted parameters(Pred\_Params)$ and the ground truth (observational) referred to as $true parameters(True\_Params)$, the abso-

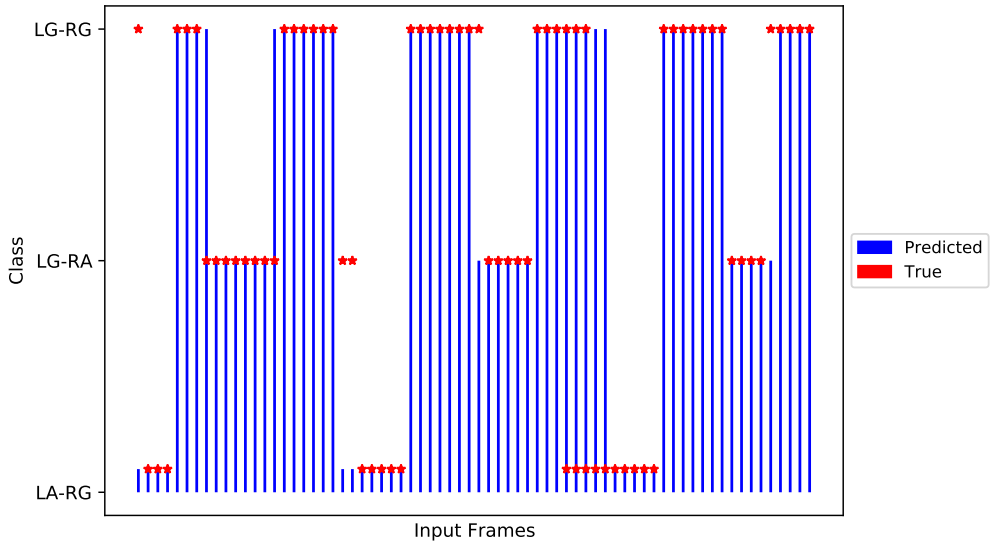


Figure 5.12: An illustration of the variation while predicting the labels for each frame of video using DN-201 model, there is a variation while changing the feet position

lute difference and Pearson's correlation coefficients (r) were computed. A Bland-Altman plot (BA plot) was used to visualize the amount and tendency of the deviations between the two systems. For the BA plot, we used the $Mean_{Diff} \pm (1.96 \times SD_{Diff})$ with $Mean_{Diff}$ being the mean difference between the $Pred\_Params$ and $True\_Params$; and SD_{Diff} being the standard deviation of the mean difference between the $Pred\_Params$ and $True\_Params$, for limits of agreement. In order to determine the degree to which a new test reproduces a gold-standard test, we used Lin's concordance correlation coefficient (LCC), which measures how well pairs of observations lie on a 45° line through the origin[252]. For all parameters, the strength of the agreement was determined based on the following criteria: Excellent (0.75–1.00), Good (0.60–0.74), Fair (0.40–0.59), and Poor (< 0.40) [253].

We selected 15 videos for predicting the labels and then quantifying different parameters from those labels. The statistics for the validity of all gait parameters calculated with the proposed system are summarized in Table 5.3. The scattered plots and Bland–Altman plots of all parameters are also shown in Figure 5.13 and Figure 5.14 respectively. NN method demonstrated excellent validity in 5 of 12 parameters, including different parameters (like number of steps, average stride length, walking time, number of the gait cycle, and average step length) (Table 5.3). A good validity was found for 5 spatiotemporal parameters (stance time, double support, cadence, step time, stride interval, and gait velocity). One parameter

Table 5.3: Various statistical correlation parameters of all gait spatiotemporal parameters quantified by NN based method and the ground truth.

Gait Parameter	LCC	r	95% limits of agreement		Significance
			Lower bound	Upper bound	
Number of Steps	0.818	0.845	0.587	0.947	<0.0001
Swing (percent)	0.574	0.678	0.254	0.883	0.0054
Stance (percent)	0.687	0.797	0.482	0.929	0.0003
Double Support (percent)	0.654	0.755	0.396	0.914	0.0011
Average Stride Length(m)	0.753	0.757	0.401	0.914	0.001
Walking Time (s)	0.79	0.936	0.816	0.979	<0.0001
Number of GCs	0.796	0.837	0.569	0.944	<0.0001
Average Step Length (m)	0.816	0.844	0.585	0.947	<0.0001
Cadence (steps/min)	0.734	0.817	0.524	0.937	0.0001
Step Time (s)	0.621	0.703	0.298	0.893	0.0034
Stride Interval(s)	0.621	0.703	0.298	0.893	0.0034
Gait Velocity (m/s)	0.735	0.771	0.428	0.92	0.0007

(swing time) was identified as fair with a little high significant value of 0.0054. All the parameters had high Pearson correlation coefficients ranging from moderately positive to very high positive correlations. (ranging from 0.67 to 0.93). The results presented here are for one walking condition, i.e., normal walk. The remaining results are included in Appendix B.

5.5 Discussion

In this work, we presented the use of transfer learning to improve the accuracy of feet' position classification from marker-less videos captured outside the strict laboratory environment. We also presented a novel method regarding the subsequent use of the detected events for the estimation of spatiotemporal parameters of the gait that are important for any gait analysis system.

As we already observed in the previous chapter, the accuracy and other performance metrics in the prediction of the labels of gait events were almost 91% on test data, which is a healthy improvement. With DN-201, we achieved up to 94% of overall accuracy and more

than 95% average accuracy. An important observation in the re-training of these pre-trained networks is the improvement in the validation loss and validation accuracy metrics. Although the use of transfer learning has not improved the prediction of gait events to a very high level as compared to the previous models, the crucial take-away was the reduction in the prediction variations. We can observe in Figure 5.12 that the variations are just one frame in most cases, and at some points, it is up to two variations. While in the previous models, it was mostly two-three frames on changing the foot position from one state to other. There was no significant difference between the performance of both pre-trained models used for the transfer of weight. As both are pre-trained with a large amount of data and have already learned the adequate weights for feature extractions, both gave almost identical results; however, DN-201 was slightly better. Regarding the variations in the prediction of a correct label of the foot position, we can argue that the toe is the last part of the foot that leaves the ground, and the tip of the toe is the last portion of the foot, and the farthest part also when recording from the back. Therefore, it is quite understandable that detecting the contact of the toe with the ground is very difficult at the last stages of the stance phase of the particular foot. The same is the case with the heel strike event when recording from the front. As the heel is the first part of the foot that touches the ground, and from the front, the heel is the last part of the foot that can be seen, therefore detecting a heel strike is also not easy, especially at the last stage of the swing phase of the foot.

One could argue that the placement of a camera on the side of the walking path may solve this problem, but in that case, we will have two feet in parallel while walking; one foot will hide the movement of the other foot. Two cameras will be required, one on each side of the walking area. Another disadvantage of the cameras on the sides is the field of view. The long field of view could be observed from the back and front, but from the side, it will be reduced, and multiple cameras on both sides will be required. This will increase the complexity of the system, and extra processing will be required for inter-camera calibration and the use of data further. Using a frontal or postern view to perform gait analysis reduces the physical space required for the test also. Although the marker-based and sensor-based systems provide gold standard accuracy in detecting the gait parameters [251, 254–256] compared to the accuracy of marker-less systems like in this study. However, the complexities in both the former systems, like strict laboratory environments, less portability, expensiveness, availability of trained staff, and others might make a trade-off with a simple marker-less system. Furthermore, the variation of one frame, and sometimes two frames, could also be ignored as the frame rate is very high, and the change in the position of the foot in the 30^{th} part of a second is too small to make some significant effect.

The study's second objective was to quantify various spatiotemporal parameters of the gait from the predicted labels through the proposed deep neural network based approach. We presented a comprehensive validation of the spatiotemporal gait parameters calculated from the gait dataset and demonstrated that they were reliably accurate.

In particular, the general and spatial gait parameters like 'the number of steps,' 'total gait cycles,' 'average step length,' 'average stride length,' and the temporal parameter 'walking time' showed excellent accuracy and high correlation with the ground truth. The other general parameters like 'cadence' and 'velocity' also showed a good degree of agreement. The 'stance time' and 'double support, and step time' temporal parameters also showed good accuracy. Only one parameter, i.e., 'swing time,' was in the fair range. There was a high positive correlation for almost all of the parameters.

The calculation of the gait parameters showed good or excellent validity except for swing time, which is again very important because of the variations in the predictions of the label. As we have already seen, there is a variation in predicting the correct label at the last stage of the stance phase and swing phase; therefore, that effect is also seen here. Therefore, the accuracy of most gait parameters relies totally on the accuracy of the prediction of gait events from the videos. Accordingly, the LCCs of stride interval and step time were found to be fair. Nevertheless, their correlation coefficients were high, suggesting this method's usability for clinical gait studies. Based on these results, it can be concluded that the calculation of the gait parameters by the proposed system is reasonable. Further, we can construct a time series containing values for each calculated temporal parameter, which can then be used to analyze the change in those parameters over time.

Considering the complexities of the existing gait analysis methods regarding their ease of use, availability, simplicity in applications, portability, and long-term data, the deep NN based method can provide simple and fairly good results for quantification of the parameters. It can be concluded that the proposed system is reasonably usable for clinical gait analysis. These results provide the scientific basis for conducting clinical gait analysis with the marker-free camera-based gait analysis method outside the laboratory. This could provide an interesting tool that allows gait analysis to take place outside of the laboratory in daily living conditions.

Although this study showed the great potential of the deep NN based method for clinical gait analysis to expand its application, the limits within which this method can provide accurate measurements need to be determined. We only counted the walking straight in two directions and did not count the turns. The turning also contains information; in some cases, it is important to have this information; therefore, it needs to be incorporated in future. Secondly, the determination of ground truth was observational, which may also affect the accuracy of good results. It could be expanded to use a sensor-based system for more

accurate ground truth determination. Although we tried to include gait data in different conditions, the data from real patients with gait disorders can further validate the results with a wide range of participants with different gait disorders. We also did not count any gender or age-based difference since sex difference affects gait characteristics . A large cohort study consisting of people with diverse backgrounds will clarify the reliability and usability of the system.

5.6 Conclusion

The present study showed that the transfer learning-based pre-trained NN improved the prediction of labels for different feet' positions and it especially reduced the variation in changing phase of the gait. The proposed method estimated spatiotemporal gait parameters with reasonable accuracy for gait data, providing a scientific basis for applying this system to clinical studies. Since this system is capable of being used both inside and outside the laboratory and has a great deal of versatility, it may open a new era of gait analysis research by being able to analyze a large number of subjects in various scenarios. To extend its application to patients bearing with gait disorders, further study with a focus on the determination of ground truth involving some sensor-based system is necessary to validate the range that this system can analyze accurately.

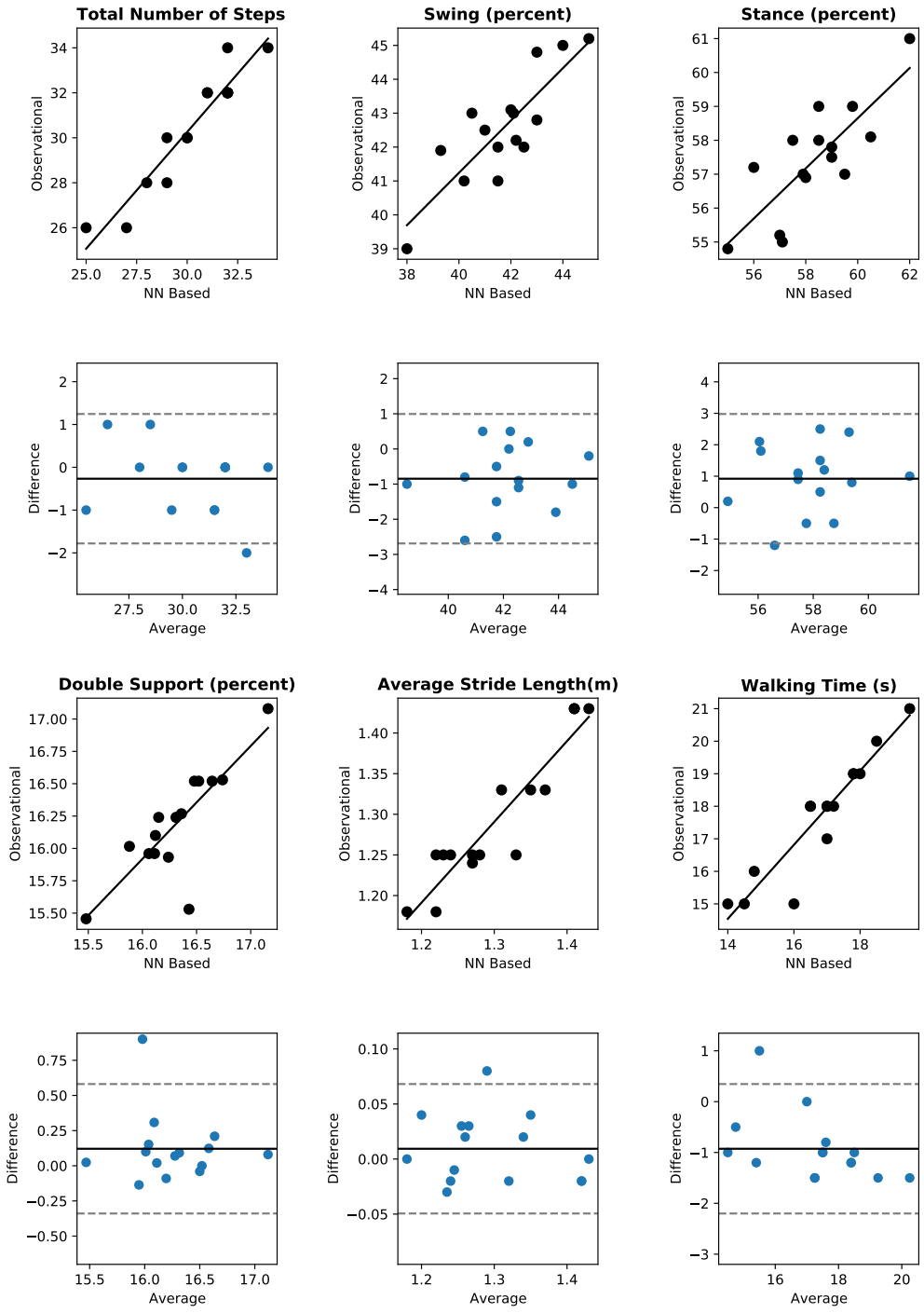


Figure 5.13: Scatter and Bland–Altman plots of the 6 different spatiotemporal parameters of the gait (number of steps, swing(%age), stance (%age), double support (%age), average stride interval (m), and walking time (s)) quantified by the NN based predicted labels and the ground truth. The upper plot is the scatter plot, and the lower plot is the Bland–Altman plot for each parameter. In the lower row, the Y-axis of the plot corresponds to the difference between the two measurements (Pred_Params and True_Params), whereas the X-axis is the mean of the two measurements. The solid lines show the average difference for the whole sample, and the dashed lines correspond to the 95% limits of agreement on both sides.

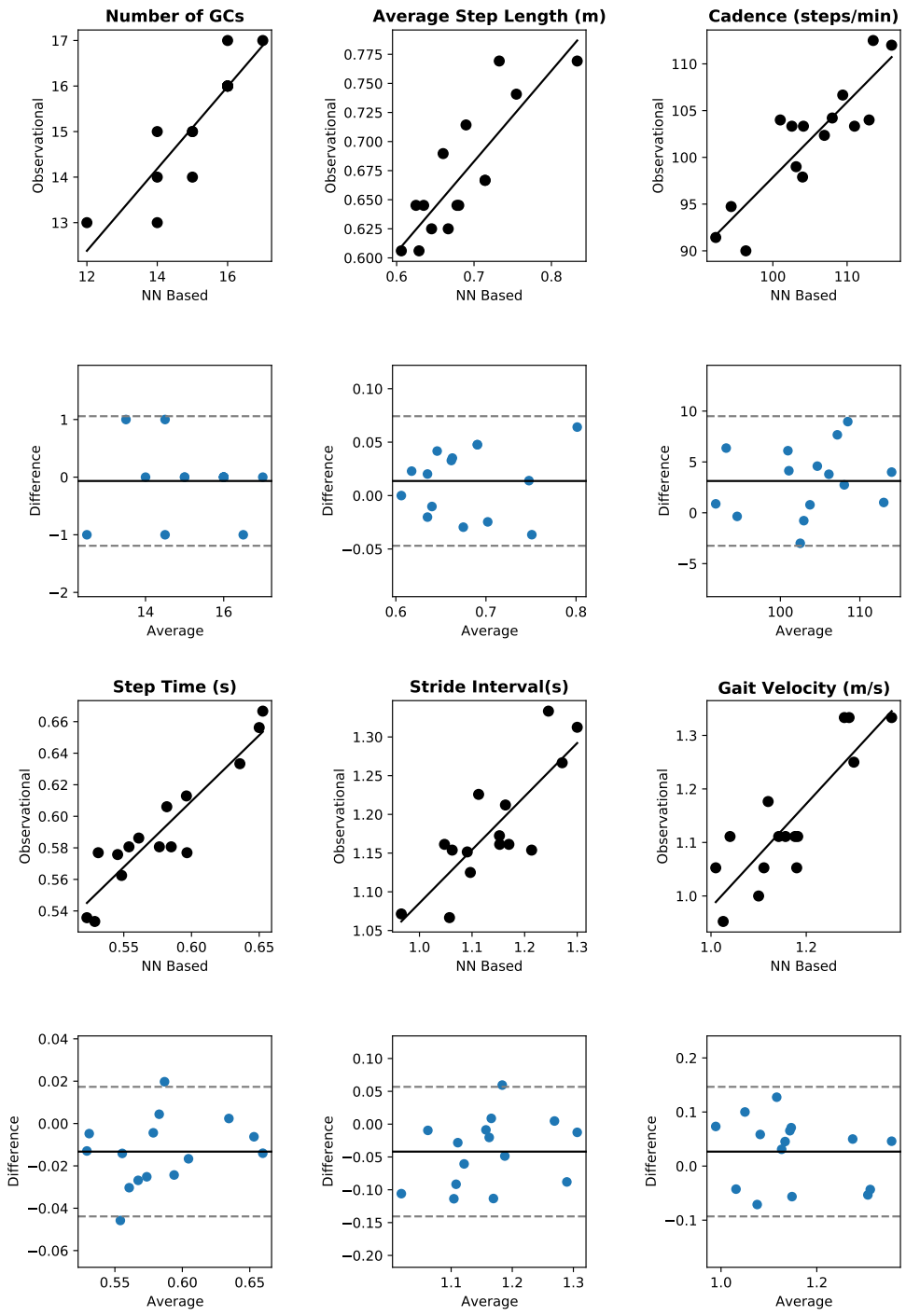


Figure 5.14: Scatter and Bland–Altman plots of the 6 different spatiotemporal parameters of the gait (number of GCs, average step length(m), cadence (steps/min), step time (s), stride interval (s), and gait velocity (ms)) quantified by the NN based predicted labels and the ground truth. The upper plot is the scatter plot, and the lower plot is the Bland–Altman plot for each parameter. In the lower row, the Y-axis of the plot corresponds to the difference between the two measurements (Pred_Params and True_Params), whereas the X-axis is the mean of the two measurements. The solid lines show the average difference for the whole sample, and the dashed lines correspond to the 95% limits of agreement on both sides.

PART III

TIME SERIES GAIT ANALYSIS

Chapter 6

Refinements to Entropy-Based Methods for the Analysis of Temporal Gait Dynamics

This chapter presents improvements to entropy-based methods that have been used for time series analysis of human gait dynamics, and have also proven to be highly effective for a wide range of biological time series analyses. It is discussed in the first section that a double-foot stride interval should be used rather than a single-foot stride interval. The second section focuses on the optimization of threshold values for symbolic entropy analysis methods for gait analysis. The last section of this chapter contains a discussion and conclusion.

6.1 Introduction & Background

It is widely acknowledged that human gait is a complex, nonlinear process [257–259] that incorporates sensory information from the cerebellum, motor cortex, and basal ganglia in addition to visual, vestibular, and proprioceptive feedback. Due to their limited ability to provide information on behavioral patterns, as well as irregular and complex behavioral fluctuations, traditional statistical analysis methods cannot be used to analyze this complex and dynamic system. Therefore, a nonlinear method is needed to analyze such complex patterns and derive valuable information from them.. In the last two decades, entropy analysis has become increasingly important in temporal-based gait research. The probability of the next state of a time series can be quantified based on knowledge of the current state of a time series. Entropy was first quantified in the 1800s in the field of thermodynamics and remains a probability-based concept today. Biological data can be quantified by approximate entropy

(ApEn), which was first proposed in 1991 [260]. sample entropy (SampEn) was introduced as the first modification to the method in order to address a bias in the calculation [261].

In 1998, entropy analysis was applied to human gait data for the first time [262]. An examination of entropy over multiple timescales was initiated in 2002 with the introduction of multiscale entropy (MsEn) [263]. Using ApEn, [264] investigated whether neurophysiological changes associated with aging affected the function of the nervous system that ensures human gait stability. Based on symbolic entropy (SyEn), symbol sequences extracted from the original stride interval time series were used to quantify the normalized corrected Shannon entropy (NCSE) of stride interval time series of normal and diseased subjects [265].

In [266], authors employed ApEn to analyze the variability of gait in elderly individuals as a means of assessing fall risk. This study examined the minimum foot clearances during treadmill walking of 14 healthy elderly subjects and ten elderly subjects with balance disorders and a history of falls (falls risk). This approach has also been applied to understanding developmental delays in motor-control function in infants [267]. Using ApEn, the authors examined the effect of running fitness on walking complexity [268].

During the walking of elderly subjects and those with knee osteoarthritis symptoms, [269] analyzed the variation in leg acceleration signals from cycle to cycle using SampEn. In their study, less variability was observed in the study group than in the control group. Using the three entropy algorithms ApEn, MsEn, and SampEn, in a study of healthy and active young adults, the authors investigated the differences between spontaneous walking and treadmill walking in terms of gait variability (stride interval time series)[270].

The SyEn method was applied in another study to quantify the complexity of stride interval time series data derived from subjects walking freely and under metronomically paced protocols [271]. For the classification of the same groups, the authors compared MsEn and SyEn and concluded that the later performed more effectively [272]. Following the development of these methods, many variations of MsEn-based methods have been proposed in recent years to quantify the complexity of multivariate time series data [273–275].

As discussed above, the gait analysis study has been conducted using several entropy-based measures of disorder and complexity. These methods normally take the time-series-based data, especially the time series of stride interval (SI) duration or gait cycle (GC) of one foot at a time, and analyze it in respect of complexity, variability, and other measures.

All entropy-based algorithms typically require parameter inputs that, if not selected properly, can lead to incorrect results being reported. For most entropy algorithms, three inputs are required. First, the length of the time series - usually abbreviated as N. Second, the relative distance between time series is used to determine whether patterns within them are similar - usually, abbreviated as r. Third, the data length is the number of measurements taken

over the entire time series to determine conditional probabilities. Usually, this parameter is abbreviated as m. There are other names for this parameter, including the pattern length, embedding dimension, segment length, vector length, word length, and pattern window.

There has been a mismatch among different studies regarding the use of different parameters. One example of how the selection of input parameters greatly affects the outcome is the difference between older and younger adults in terms of their gait patterns. In [276], authors reported for the first time that changing the tolerance for comparison led to different interpretations of young and older adults' gait. The authors reported that, compared with younger adults, older adults exhibit more random joint angle patterns during walking. However, accelerometry data demonstrate that age increases the likelihood of gait, indicating a more rigid walking pattern [269, 277]. According to [278], parameter input was also strongly associated with results between older and younger adults in walking. It is likely that some of these differences in findings between ages or pathological groups may be explained by the type of data used or the sampling rates during data collection [279, 280].

Entropy-based methods use either left or right foot gait data as input data when applying these methods. Data from both feet are not used simultaneously. It is possible that the classification between healthy and diseased gait could be improved by using both feet' data. The first refinement to entropy-based methods in this study is to determine how using both feet' gait cycles simultaneously in a single time series affects the outcome of the separation between controlled and diseased gait, using MsEn method. In this study, MsEn was used to examine gait time series in normal and neurodegenerative disease conditions.

A second refinement relates to the SyEn method, which is based on the quantification of threshold dependent normalized corrected Shannon entropy (TNCSE). Based on gait time series data in aging and diseased conditions, parameter selection (threshold value) optimization is presented and compared with existing threshold selection methods. We selected MsEn and SyEn because previous research has indicated that ApEn and SampEn are unable to capture the structural richness and component entanglement characterized by complex systems operating over a wide range of temporal scales. It is therefore recommended that multiscale entropy or a refined entropy method be used to quantify complexity [281].

6.2 Materials

When using entropy-based methods, data length must be considered. As described in previous chapters of this thesis, we collected only 20 m of data from volunteers for our pilot study. The number of stride intervals ranges between 15 and 33. The final application of

the proposed home-based gait analysis method gives long-term data; in that case, the time series analysis methods could be applied. However, in the studies presented in the previous chapters, the data needs to be larger to apply the time series analysis methods. Therefore, the data from the stride interval (SI)-time series database obtained from PhysioNet, an online database of physiological signals, was used. [282]. The database used in this study is the Gait Dynamics in neurodegenerative disease database [283]. This database collected data to gather SI time series from 16 healthy individuals and 48 patients with neurodegenerative diseases. This database was referred to as DB1 in this study and contained the SI time series of 16 healthy individuals. Twenty, fifteen, and thirteen patients with Huntington disease (HD), Parkinson disease (PD), and Amyotrophic Lateral Sclerosis (ALS), respectively, were in the pathologic group. During the experiment, healthy subjects and patients walked for an average of 200–270 steps with an interval of 1.0 seconds and 120–280 steps with an interval of 1.0–1.5 seconds, respectively. The demographic information related to the DB1 is presented in Table 6.1.

Table 6.1: Detailed demographic information about the subjects in database DB1 with Controlled, Parkinson, Huntington and Amyotrophic lateral sclerosis diseased subjects.

Parameter	Controlled	Parkinson's disease	Huntington disease	Amyotrophic Lateral Sclerosis
Age (years)	38.69 $\pm$18.73	67.20 $\pm$10.69	47.37 $\pm$12.51	55.62 $\pm$12.83
Height (m)	1.833 $\pm$0.087	1.87 $\pm$0.152	1.8437 $\pm$0.089	1.7446 $\pm$0.950
Weight (kg)	66.81 $\pm$11.08	75.07 $\pm$16.90	73.47 $\pm$16.24	77.12 $\pm$21.15
Speed (m/s)	1.354 $\pm$0.160	0.999 $\pm$0.202	1.15 $\pm$0.349	1.054 $\pm$0.218

More detailed information about the database can be found in [282, 283]. Figure 6.1 shows the stride interval time series of healthy, HD, and PD subjects. The data is selected from DB1 and the first 200 points are used from healthy and HD subjects because in PD time series, the total data points are 200.

6.3 Methods

In the following sections, two widely used entropy-based methods, MsEn and SyEn, for gait analysis are discussed. These methods, along with the proposed refinements, are presented below.

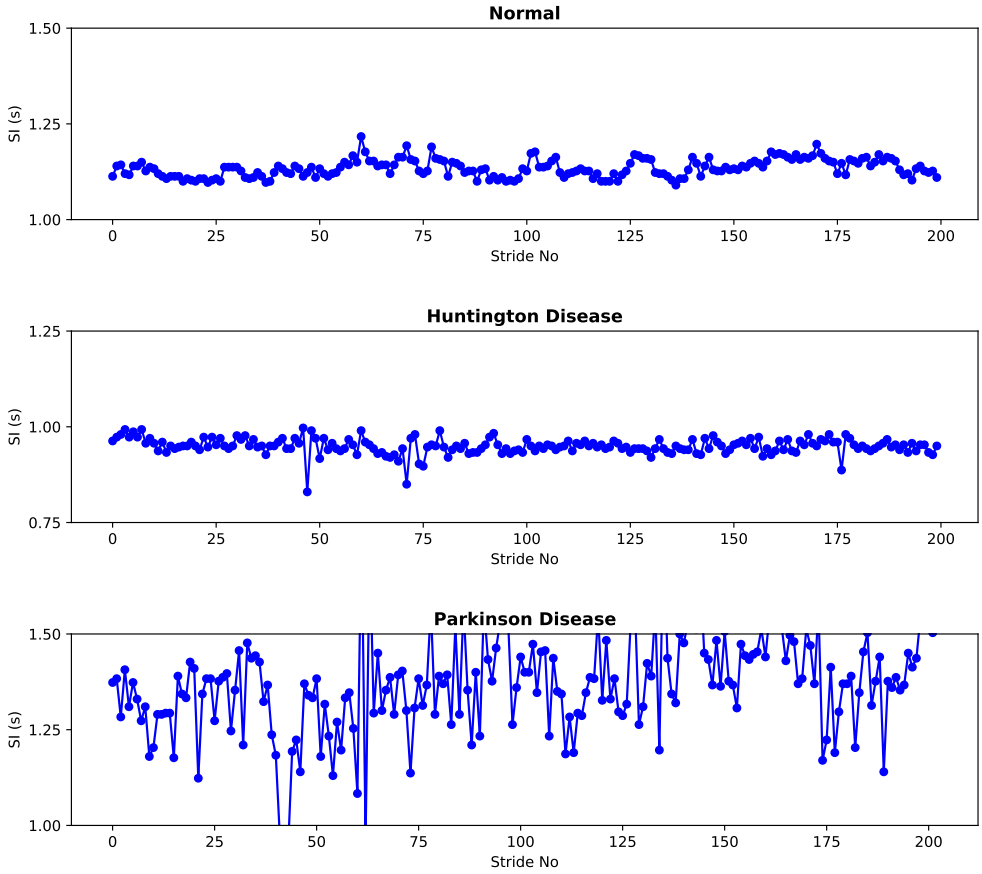


Figure 6.1: A representation of stride interval time series (200 points) of three subjects from DB1 [282, 283], the upper plot is SIs of Normal walking, middle plot is SIs of a patient with Huntington Disease, and the lower plot is SIs of a patient with Parkinson Disease

6.3.1 Multiscale Entropy for Gait Time Series Analysis

To analyze a stride interval time series $SI = \{SI_1, SI_2, \ldots\ldots, SI_N\}$ with a total length N, the first step is to successively increase the number of data points and then average them in non-overlapping windows to construct consecutive coarse-grained time series. This average is done based on a scale factor, denoted by ω.

Using equation Equation (6.1), each element of a coarse-grained time series C_j^{ω} is calculated:

$$C_j^{\omega} = \frac{1}{\omega} \sum_{i=(j-1)\omega+1}^{j\omega} SI_i \tag{6.1}$$

where ω denotes the scale factor, j represents the index of coarse-grained time series and the value of i ranges from 1 to N/s. The minimum value of ω is 2 because, at $\omega = 1$, the resultant time series will be just the original time series.

The resultant coarse-grained stride interval times series(CG series) for different scale factor values are presented in Figure 6.2

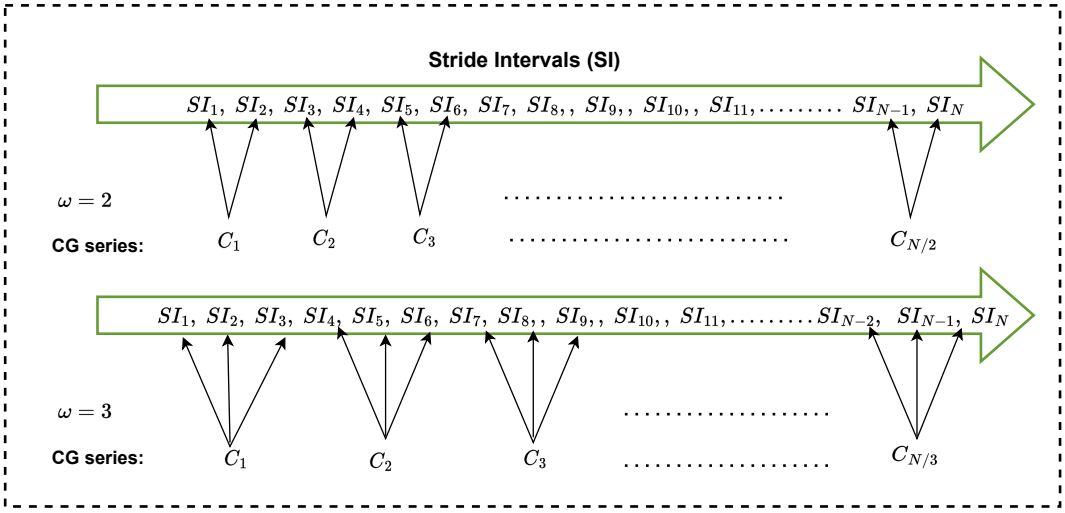


Figure 6.2: Process of converting original stride interval time series into coarse-grained stride interval time series at scale 2 and 3

In the second step, SampEn is calculated and plotted as a function of the scale factor for each coarse-grained time series based on different scales.

The quantification of SampEn using the coarse-grained series C_i^{ω} is as follows [284].

For simplicity, we assume $C_j = \{c_1, c_2, c_3, \ldots\ldots, c_T\}$, where T is the total length of this series. Two vectors $x_m(a)$ and $x_m(b)$ are constructed from C_j, both with a pattern length m. In the sequence $x_m(a)$, the vector starts at element c_a of the series, and in the sequence $x_m(b)$, the vector starts at element c_b. Then the distance between the two vectors is calculated which is the maximum difference between their corresponding elements. This distance is $d[x_m(a), x_m(b)]$. After this, the value of d is compared with a parameter r, which is called similarity criteria. The two vectors will be considered as similar if the distance is less than the similarity criteria i.e. $(d < r)$.

Let X_m be the set of vectors $\{x_m(1), x_m(2), \ldots\ldots x_m(T-m+1)\}$, with pattern length m, and maximum number of patterns $T-m+1$. For each vector $x_m(a)$, count the number of other vectors $x_m(b)$ with b is between 1 to $T-m$, so that $d[x_m(a), x_m(b)] < r$. Let V_i be the

total number of vectors. The function $V_m(a)$ is defined as:

$$V_m(a) = \frac{V_a}{T-m} \tag{6.2}$$

The probability that two vectors of pattern length m are similar will be:

$$V_m = \frac{1}{T-m} \sum_{a=1}^{T-m} V_m(a) \tag{6.3}$$

Then the pattern length is increased to $m+1$ and same procedure is applied to obtain $V_{m+1}(a)$ and V_{m+1}.

$$V_{m+1}(a) = \frac{V_a}{T-m+1} \tag{6.4}$$

$$V_{m+1} = \frac{1}{T-m+1} \sum_{a=1}^{T-m+1} V_{m+1}(a) \tag{6.5}$$

Then the the sample entropy will be the ratio of the above quantities calculated as:

$$SampEn(r,m) = -ln\left[\frac{V_{m+1}}{V_m}\right] \tag{6.6}$$

It is essential to select m and r correctly in order to estimate SampEn. By default, r is set at 10 to 25% of the standard deviation of the time series, and $m = 2$. Based on different scale factors (ω), the value of SampEn is calculated, known as MsEn.

6.3.2 Refinement to MsEn method for Gait Analysis

The entropy-based methods for analyzing stride interval time series consider each foot's stride interval individually, as discussed earlier. It is more appropriate to use the data from both feet simultaneously, since both feet are used at the same time when walking. Figure 6.3 illustrates this method in which the corresponding elements from both time series are placed one after another, thus including the data from both feet as well as increasing the total amount of data. This is followed by estimating the multiscale entropy using the quantification of SampEn method stated above over multiple scales.

6.3.3 Symbolic Entropy for Gait Time Series Analysis

The process of quantifying symbolic time series analysis of stride intervals starts with transforming the original time series into a series of discretized codes (symbols). The symbolization of data can be done in various ways. Nevertheless, depending on the data, we must set the type and number of symbols. A higher number of symbols can be used for long

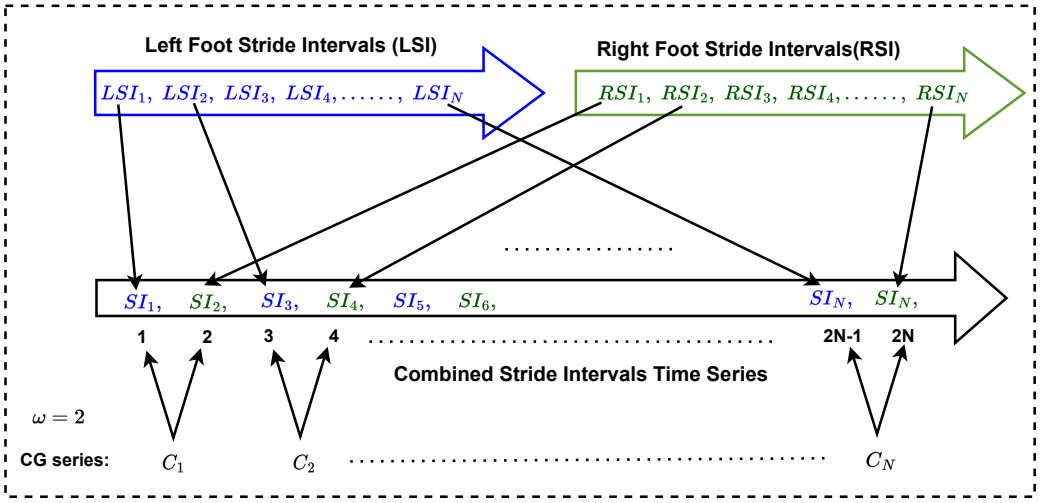


Figure 6.3: Process of combining both feet's stride interval into one series and then applying multiscale entropy method

time series, whereas fewer symbols are desired for shorter time series. The most widely used method is to transform the original time series into a sequence of '1' and '0' (symbol series) depending on a certain threshold. In the next step, TNCSE of the symbol series is calculated using the steps suggested in [265].

Data Symbolization: Symbolization of the unprocessed (original) stride interval time series requires a partition of time series, and the selection of partition (threshold) is the preliminary point. Symbolization depends upon the 'τ' (threshold value). A threshold value can be set by using various approaches. The most common method is to split the original time series into two bins and allot the value '0' or '1' to every bin with respect to its happening [285]. The number of symbols is called the quantization level 'Q'. If we have a stride interval time series $SI = \{SI_1, SI_2, \ldots\ldots, SI_N\}$ with a total length N, we can translate these SI values into symbols (Sy) by using the following criterion in Equation (6.7).

$$sy_i = \begin{cases} 1 & \text{if } SI_i \geq \tau \\ 0 & \text{otherwise} \end{cases} \tag{6.7}$$

The following method can also be used to convert time series: the corresponding symbol is '1' if the difference between consecutive SIs is greater than a threshold, or '0'; if the difference between consecutive SIs is less than a threshold.

$$sy_i = \begin{cases} 1 & \text{if } SI_i - SI_{i+1} \geq \tau \\ 0 & \text{if } SI_i - SI_{i+1} > \tau \end{cases} \tag{6.8}$$

There is a transformation to this symbolization provided in [285] in which instead of two symbols, the original time series is converted into 4 symbols based on an average of time

series $\overline{SI}$ and a special parameter 'a', called weight parameter using following criterion.

$$sy_i = \begin{cases} 0 & \text{if } SI_i > (1+a)\overline{SI} \\ 1 & \text{if } SI_i > \overline{SI} \text{ and } SI_i < (1+a)\overline{SI} \\ 2 & \text{if } SI_i > (1-a)\overline{SI} \text{ and } SI_i > \overline{SI} \\ 3 & \text{if } SI_i \leq (1-a)\overline{SI} \end{cases} \tag{6.9}$$

However, the most used method of symbolization is threshold dependent.

Symbol Sequence Formation: The next phase of the symbolic entropy calculation is the formation of symbol sequences (words) using the symbol series. This is done by gathering sets of symbols together in temporal order using a fixed-length (word length L) template, which slides stepwise along the symbol series. The symbol sequences are built using any finite number of consecutive symbol values in the series which is called word series. Then each possible sequence is converted to the corresponding decimal value. This series will be called code series. In the next step, the entropy is calculated by plotting the histogram of the probability of occurrence of different codes in the code series. The process of symbolization, symbol sequence formation and code series development is presented in Figure 6.4.

Quantification of TNCSE: By using the probability of code series, TNCSE is calculated using following equations: [265].

$$\text{Shannon entropy (SEn)} = -\sum_{i=0}^{Q^L-1} p_i \, log_2 pi \tag{6.10}$$

where Q is the quantization level, L is the word length, and p is the histogram of the symbol-sequence frequencies. The estimation of SEn are affected by random error in numbers and also by systematic or bias [286]. Therefore a correction in SEn was given by [287], as following:

$$\text{CSEn} = \text{SEn} + \frac{C_r - 1}{2Mln2} \tag{6.11}$$

M is the total size of the word series (number of words), and C_r is the number of occurring words in all possible words. CSEn will be maximum for some word-length L and quantization-level 'Q' when all words occur with identical distribution in the series. It is given below:

$$\text{CSEn}^{max} = -log_2\left(\frac{1}{M}\right) + \left(\frac{M-1}{2Mln2}\right) \tag{6.12}$$

Equation (6.12) illustrates that the maximum CSEn value for two different word lengths will not be identical. Similarly, when L increases, M also increases, and therefore the maximum CSEn value will also increase. Therefore, the values of CSEn for two different word lengths cannot be compared at the same threshold value and quantization level. To

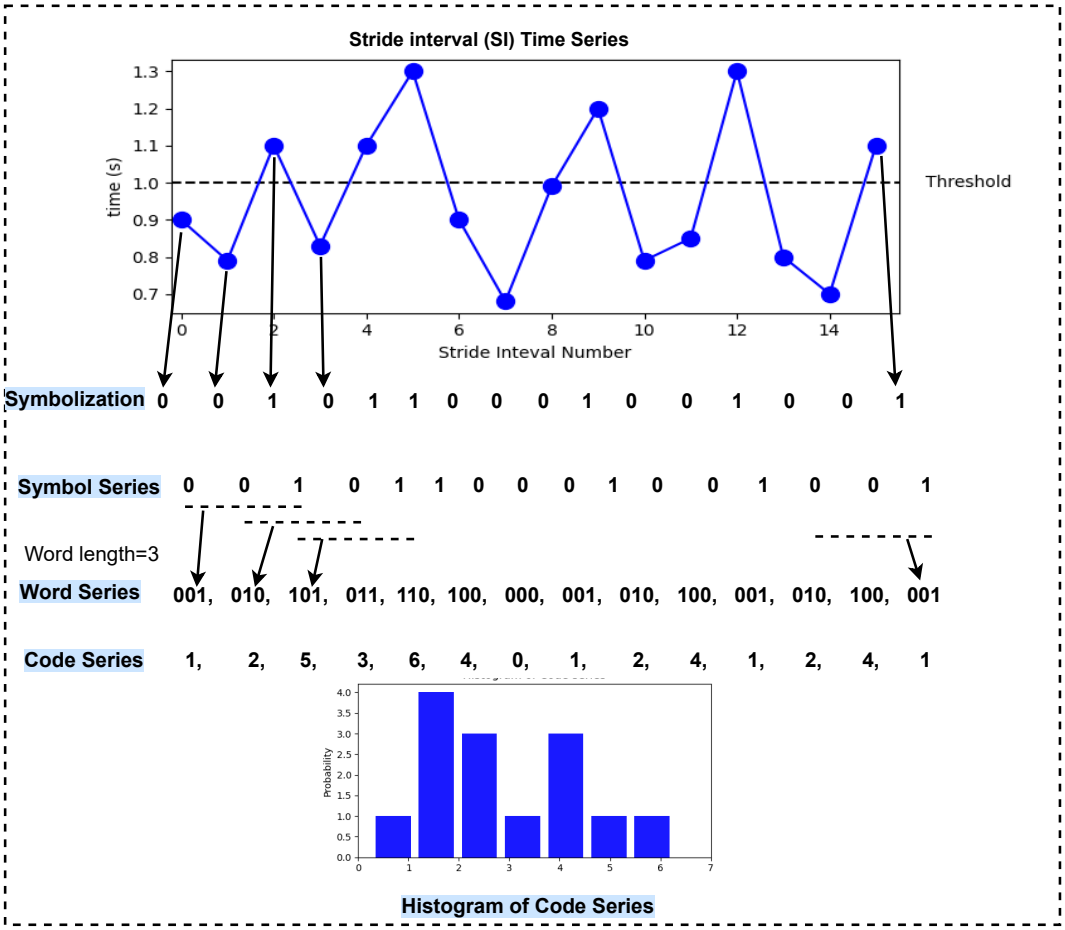


Figure 6.4: A schematic representation of converting the original stride interval time series into symbol series and subsequent formation of word series and code series

solve this problem, TNCSE was proposed by [265]. The normalizing factor in TNCSE will be the maximum CSEn value at some word length L and quantization level Q. TNCSE is given below:

$$\text{TNCSE(L,Q)} = \frac{CSE}{CSE^{max}} \tag{6.13}$$

Any word length or quantization level will result in a TNCSE value between 0 and 1.

6.3.4 Refinement to SyEn Method

As it has already been discussed, the quantification of SyEn depends on certain threshold values (τ) for symbolization. The threshold value selection is an important phase of symbolization, as the data points in the time series are either below or above the threshold. If the quantization level is set to two (0 or 1), then the small value of the threshold will result

in more ones than the number of zeros, while a high threshold value will result in a fewer number of ones. If the symbol series contains more zeros (or more ones), the resultant entropy value will be small, and it will not be a true representation.

In different studies like [265, 271, 272], the threshold is chosen as an arbitrary value, and SyEn is calculated using those arbitrary values. Therefore, a refinement to this threshold value selection is proposed, by using other time series statistical values. These values include average stride interval (ASI), variation in stride interval (VSI), median of SI time series (SI_{med}), and mode of SI time series (SI_{mod}). ASI is the averaged value of the SI time series, VSI, is the variance of the SI time-series. Gait is not a simple, linear system; variations in stride intervals and arbitrary irregularities should also be considered when analyzing gait. A normal gait pattern shows variations that are not purely chaotic. Instead, each stride interval is dependent on previous stride intervals; therefore, it is pertinent to use VSI as a threshold measure. The other two values are SI_{med}, the median of the time series, and SI_{mod}, the mode of the SI time series. These are calculated using the following equations:

$$\text{ASI} = \frac{1}{N}\sum_{i=1}^{N} SI_i \tag{6.14}$$

$$\text{VSI} = \frac{\sum_{i=1}^{N}(SI_i - \mu)^2}{N} \tag{6.15}$$

$$SI_{med} = \begin{cases} SI_{\frac{N+1}{2}} & \text{if } N \text{ is odd} \\ \frac{1}{2}\left(SI_{\frac{N}{2}} + SI_{\frac{N}{2}+1}\right) & \text{if } N \text{ is even} \end{cases} \tag{6.16}$$

$$SI_{mod} = \max\left(\mathrm{f}_{SI}\right) \tag{6.17}$$

where N is the length of the time series, μ is the mean, and f is the frequency of data points in the time series. Hence the symbolization method discussed above will be modified, and we can rewrite the Equation (6.7) as follow

$$sy_i = \begin{cases} 1 & \text{if } SI_i \geq T \\ 0 & \text{otherwise} \end{cases} \tag{6.18}$$

where T is one of the values from ASI, VSI, SI_{med}, and SI_{mod}.

The SyEn values using arbitrary threshold values, and the methods proposed here are compared and presented in the results section.

6.4 Results

Both methods were used to analyze SI time series in DB1. Then both methods were applied to the proposed refinements, and a comparison was made. In the following sections, the outcomes of both methods are discussed.

6.4.1 Multiscale Entropy Analysis of Stride Interval Time Series

Generally, entropy values are expressed in terms of complexity, with normal time series being considered more complex than diseased or aged time series. First, we compared the complexity of normal and neurodegenerative-diseased data sets using the MsEn method. For coarse-graining the original SI time series, we used scales of 1 to 15 and determined the corresponding entropy value on each scale. Using the coarse-grained time series, MsEn is plotted against the scale factor. As recommended by many studies, the parameters m and r were set to 2 and 0.15, respectively. Each element of the coarse-grained time series is obtained by averaging the number of data points depending on the scale factor.

A comparison of the MsEn of the controlled time series and neurodegenerative diseased stride interval time series extracted from the left foot is shown in Figure 6.5. Rather than

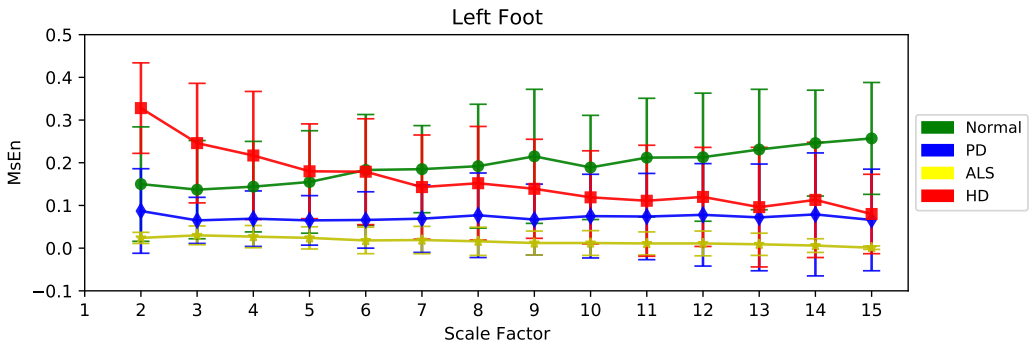


Figure 6.5: An analysis of the MsEn of controlled walking time series derived from the left foot of healthy subjects and those with neurodegenerative diseased conditions. These curves represent the lines connecting the mean $\pm$ standard deviation values of multiscale entropy.

representing the MsEn curve for one particular time series, the curves shown represent lines connecting the mean values of MsEn for all diseased and physiological time series. The

physiologic time series are assigned the lowest entropy value for scale 1, which corresponds to traditional (single scale) SampEn. We performed statistical tests to determine the significant difference between the subjects. The degree of separation between groups at all scales was quantified using p-values (95% confidence interval). The corresponding MsEn values for all of the 4 types of stride interval time series and the p-values are shown in Table 6.2. The

Table 6.2: Mean ± Standard Deviation values of Multiscale entropy for controlled, and neurodegenerative diseased conditions and corresponding p-values, using stride interval time series of **left foot** at scale 1 to 15

Scale	Multi-Scale Entropy (Mean ± Standard Deviation)				p-values		
	Normal	ALS	HD	PD	Nor Vs ALS	Nor Vs HD	Nor Vs PD
1	0.15±0.134	0.024±0.013	0.328±0.106	0.087±0.099	0.0005	0.0061	0.0018
2	0.137±0.115	0.03±0.022	0.246±0.140	0.065±0.054	0.0072	0.0053	0.0019
3	0.144±0.106	0.027±0.026	0.217±0.150	0.069±0.065	0.0065	0.0042	0.0040
4	0.155±0.120	0.024±0.026	0.18±0.111	0.065±0.058	0.0079	0.0085	0.0035
5	0.183±0.130	0.018±0.031	0.179±0.124	0.066±0.066	0.0074	0.0057	0.0036
6	0.185±0.102	0.019±0.032	0.143±0.122	0.069±0.079	0.0056	0.0054	0.0044
7	0.192±0.145	0.016±0.033	0.152±0.133	0.077±0.099	0.0011	0.0019	0.0059
8	0.215±0.157	0.012±0.028	0.139±0.116	0.067±0.083	0.0006	0.0064	0.0064
9	0.189±0.122	0.012±0.029	0.119±0.109	0.075±0.098	0.0005	0.0614	0.0016
10	0.212±0.139	0.011±0.027	0.111±0.109	0.074±0.101	0.0005	0.0067	0.0070
11	0.213±0.150	0.011±0.029	0.12±0.110	0.078±0.12	0.0053	0.0056	0.0048
12	0.231±0.141	0.009±0.026	0.096±0.116	0.072±0.125	0.0005	0.0098	0.0054
13	0.246±0.124	0.006±0.016	0.113±0.135	0.079±0.144	0.0005	0.0031	0.0045
14	0.257±0.131	0.001±0.004	0.08±0.093	0.066±0.119	0.0002	0.0055	0.0040
15	0.253±0.140	0.001±0.002	0.087±0.110	0.069±0.163	0.0006	0.0065	0.0031

Figure 6.5 shows that HD has the highest entropy value at scale 1, but also a high standard deviation. At scales below 5, the normal stride interval series has lower entropy values than HD. However, at higher scales, the MsEn value for normal walking is higher than for all diseased subjects. ALS-diseased subjects have the lowest MsEn values at all scales, and the fluctuation at different scales is also very minimal. Similarly, there is no big difference in PD entropy values at a range of scales. Overall, we can determine that the healthy stride interval has higher entropy values at a higher scale factor. When comparing normal and ALS-diseased subjects, normal walking is not statistically significantly different from ALS at lower scales. The classification between the groups is improved at scales greater than 6. The maximum degree separation between normal and ALS time series was obtained at

scale = 14. The p-values for Normal compared to PD subjects were significant at almost all scales, which shows a good degree of separation. At most of the scales, the Normal and HD subject's entropy values did not show a statistically significant difference; however, for some scales, it is less than 5%, which is a statistically significant result. Consequently, we can say that while using the stride interval time series of the left foot only, the MsEn did not show a clear degree of separation between groups on all scales; however, it showed good results at some scales.

Next, we applied the MsEn method to analyze the stride interval time series derived from the right foot of the subjects in DB1 in normal and neurodegenerative diseased conditions. We kept all the parameters the same, i.e., scales=1-15, $m = 2$, and $r = 0.15$. The results showing mean±standard deviation curves are plotted in Figure 6.6. The corresponding

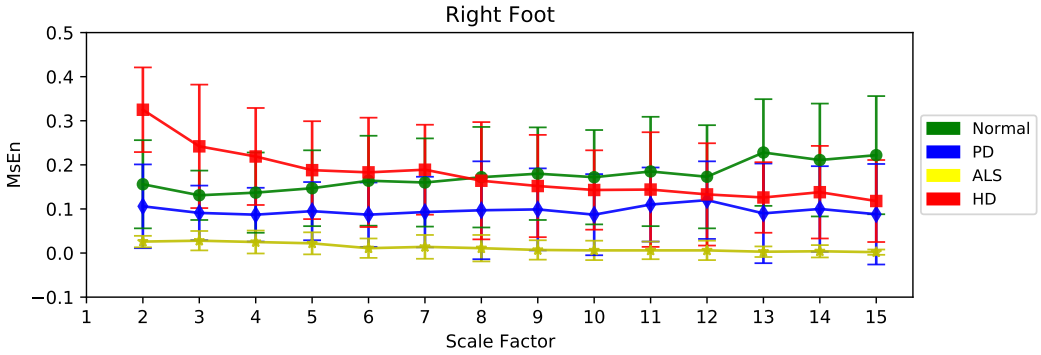


Figure 6.6: An analysis of the MsEn of controlled walking time series derived from the right foot of healthy subjects and those with neurodegenerative diseased conditions. These curves represent the lines connecting the mean ± standard deviation values of multiscale entropy.

MsEn values for normal, ALS,HD, and PD stride interval time series derived from the right foot, along with the p-values showing the degree of separation among normal and other diseased groups are shown in Table 6.3. Figure 6.6 shows that the stride interval of the right foot leads to similar results as the entropy values of the left foot. Normal walking has higher entropy values at higher scales above 8. However, HD also has higher entropy values at lower scales than all other conditions. HD also has a higher standard deviation, which shows that its values are spread to two extents. At scales 12-15, the entropy values of HD and PD are similar. The dispersion in entropy values of PD was low. ALS has the lower entropy values assigned on all scales, in the case of the stride interval of the right foot. When comparing normal and ALS-diseased subjects, normal walking is statistically significantly different from ALS at almost all lower scales, but it was not significant at scales higher than

Table 6.3: Mean ± Standard Deviation values of Multiscale entropy for controlled, and neurodegenerative diseased conditions and corresponding p-values, using stride interval time series of **right foot** at scale 1 to 15

Scale	Multi-Scale Entropy (Mean ± Standard Deviation)				p-values		
	Normal	ALS	HD	PD	Nor Vs ALS	Nor Vs HD	Nor Vs PD
1	0.156±0.100	0.026±0.013	0.325±0.096	0.0106±0.095	0.0051	0.0011	0.0068
2	0.131±0.056	0.028±0.022	0.242±0.140	0.091±0.062	0.0012	0.0011	0.0104
3	0.137±0.091	0.025±0.026	0.219±0.110	0.087±0.061	0.0045	0.0056	0.0102
4	0.147±0.086	0.022±0.025	0.188±0.111	0.095±0.066	0.0048	0.0063	0.0069
5	0.164±0.102	0.011±0.022	0.183±0.124	0.087±0.073	0.005	0.0061	0.0023
6	0.16±0.100	0.014±0.027	0.189±0.102	0.093±0.08	0.0054	0.0061	0.0031
7	0.172±0.114	0.011±0.03	0.164±0.133	0.077±0.111	0.0044	0.0074	0.0044
8	0.18±0.105	0.007±0.022	0.152±0.116	0.099±0.093	0.0041	0.0075	0.003
9	0.172±0.107	0.006±0.022	0.143±0.090	0.097±0.092	0.0041	0.0062	0.0062
10	0.185±0.124	0.006±0.02	0.144±0.130	0.110±0.084	0.0045	0.0061	0.0064
11	0.173±0.117	0.006±0.022	0.133±0.116	0.120±0.088	0.0055	0.0064	0.0117
12	0.228±0.121	0.003±0.012	0.126±0.080	0.090±0.113	0.0056	0.0059	0.0017
13	0.211±0.128	0.004±0.014	0.138±0.105	0.010±0.097	0.0056	0.0024	0.0014
14	0.222±0.134	0.002±0.006	0.118±0.093	0.088±0.114	0.0056	0.0049	0.0017
15	0.205±0.127	0±0	0.126±0.100	0.075±0.103	0.0106	0.0727	0.0032

10. The classification between normal and HD diseased subjects is statistically significant either at lower or higher values, but not at the middle values of the scale factor. The degree of separation between normal and PD subjects is high at higher scales i.e. greater than 10. In general, there is mixed behavior in the separation between normal and diseased conditions. The results of both feet show that we cannot draw conclusions about the classification of normal and diseased subjects when using MsEn. For some scales, there is a clear separation among the groups but for others, there is no clear degree of separation.

Therefore, we next applied the MsEn method by combining both feet data into one-time series. It was combined as one stride interval of the left foot, then the next stride interval of the right foot, and so on. The size of the time series was doubled after merging the stride interval data of both feet. All other parameters were kept to avoid any effect of the parameter change. The pattern length (m) was 2, the similarity criteria (r) was 0.15, and the scale factor was 1-15. It is worth noting that while using one-foot data, we got some values of MsEn as '0' in some individual time series at some scales. This was due to the length of the time series because once the scale increased, the size of the coarse-grained time series decreased. For example, if a time series consists of 300 points, the coarse-grained series has a size of 20 at scale=15 - decreasing the size to very low-level results in the '0' value for MsEn. Therefore

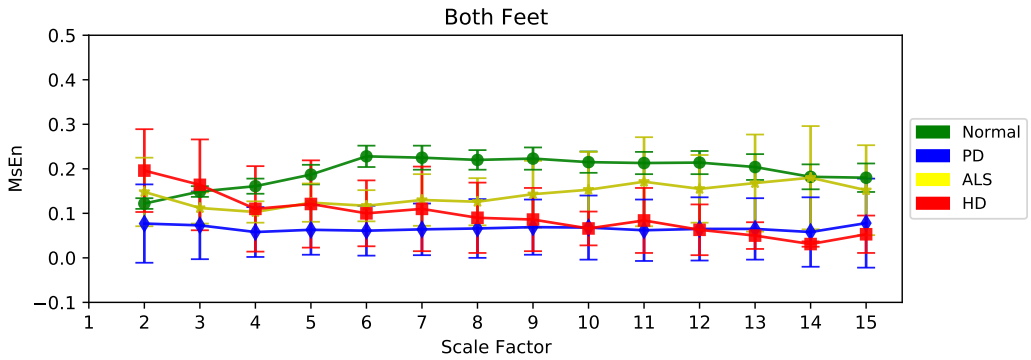


Figure 6.7: An analysis of the MsEn of controlled walking time series derived from the left and right feet (combined) of healthy subjects and those with neurodegenerative diseased conditions. These curves represent the lines connecting the mean $\pm$ standard deviation values of multiscale entropy.

merging the left and right foot data has also overcome this problem. The mean+standard deviation values of the MsEn at different scales in normal and neurodegenerative diseased conditions are plotted in Figure 6.7. We performed statistical tests to determine the significant difference between the subjects using both feet data. The degree of separation between groups at all scales was quantified using p-values (95% confidence interval). The corresponding MsEn values for all of the 4 types of stride interval time series and the p-values using both feet stride interval are shown in Table 6.4.

The figure shows that normal walking was assigned higher entropy values at almost all scales after scale=2. It is also in line with the general considerations of complexity that the normal physiologic time series has higher complexity than the pathological time series. Although on scales 1, and 2, it was less than the entropy value of HD, the difference was not that high. Another important finding is that normal walking also has low variations. In contrast to the left and right feet individually, the ALS entropy values are higher than HD and PD entropy values at scales above 4. Even at scale 15, the entropy value of ALS is higher than normal. Entropy values of HD decreased almost linearly with an increase in the scale factor. However, the entropy values of PD remained almost the same at all scales, with a little variation in upper and lower limits.

The p-values also showed statistically significant results while comparing normal walking with neurodegenerative diseased conditions walking, using stride intervals of both feet at the same time. The Table 6.4 shows that the entropy values of normal walking are significantly different from the entropy values of ALS except at very low or very high scales. At scales 1,3,14, and 15, it was not less than 5%, and the maximum degree of separation was obtained at scale -10 with a value of 0.0003. When comparing normal with HD, there were almost the

Table 6.4: Mean ± Standard Deviation values of Multiscale entropy for controlled, and neurodegenerative diseased conditions and corresponding p-values, using stride interval time series of **Both feet** at scale 1 to 15

Scale	Multi-Scale Entropy (Mean ± Standard Deviation)				p-values		
	Normal	ALS	HD	PD	Nor Vs ALS	Nor Vs HD	Nor Vs PD
1	0.122±0.012	0.148±0.077	0.196±0.093	0.077±0.088	0.0057	0.0053	0.0012
2	0.149±0.012	0.112±0.035	0.164±0.102	0.073±0.076	0.0053	0.0052	0.0024
3	0.161±0.017	0.103±0.024	0.11±0.096	0.058±0.056	0.0009	0.0001	0.0041
4	0.187±0.022	0.124±0.043	0.121±0.098	0.063±0.056	0.0006	0.0003	0.0019
5	0.228±0.024	0.117±0.035	0.1±0.074	0.061±0.056	0.0010	0.0003	0.0076
6	0.225±0.027	0.13±0.058	0.11±0.095	0.064±0.058	0.0013	0.0001	0.0026
7	0.22±0.022	0.126±0.053	0.09±0.079	0.066±0.066	0.0034	0.0001	0.0049
8	0.223±0.025	0.143±0.075	0.086±0.071	0.069±0.062	0.0009	0.0005	0.0021
9	0.215±0.024	0.153±0.085	0.066±0.038	0.068±0.072	0.0004	0.0001	0.0041
10	0.213±0.025	0.171±0.1	0.084±0.073	0.062±0.069	0.0003	0.0025	0.0046
11	0.214±0.026	0.155±0.076	0.063±0.057	0.065±0.071	0.0008	0.0001	0.0038
12	0.204±0.029	0.168±0.109	0.05±0.03	0.065±0.069	0.0004	0.0002	0.0045
13	0.182±0.028	0.18±0.116	0.031±0.006	0.058±0.078	0.0005	0.0043	0.0016
14	0.178±0.032	0.175±0.101	0.053±0.042	0.078±0.1	0.0055	0.0044	0.0051
15	0.164±0.019	0.18±0.099	0.027±0.016	0.061±0.081	0.0052	0.0051	0.0041

same results with a maximum degree of separation with a value of 0.0001 at scale 3,6,7,9, and 11. PD's entropy values were also statistically significantly different than the entropy values of normal walking conditions using both feet data simultaneously at all scales. Hence, combining both feet data in a single time series performed well in the classification of normal subjects with neurodegenerative diseased subjects, compared to single foot data.

6.4.2 Symbolic Entropy Analysis of Stride Interval Time Series

As in the [265, 271, 272] studies, the symbolic entropy method was used to quantify TNCSE based on stride interval time series derived from sixteen healthy subjects walking freely at normal speed, twenty HD subjects, fifteen PD subjects, and thirteen ALS subjects, considering various arbitrary threshold values. These variables' values were kept constant to avoid being affected by a change in the arbitrary values. We set the quantization level (Q) to two ($0 and 1$) and the word length (L) to three. The TNCSE was calculated based on the mean values of all subjects. A comparison was made between the entropy values of normal

walking time series and those of diseased walking time series at different threshold values (τ). In Figure 6.8, walking in the free state is compared with walking in patients suffering from ALS, HD, and PD. As compared to neurodegenerative diseased walking, normal walking

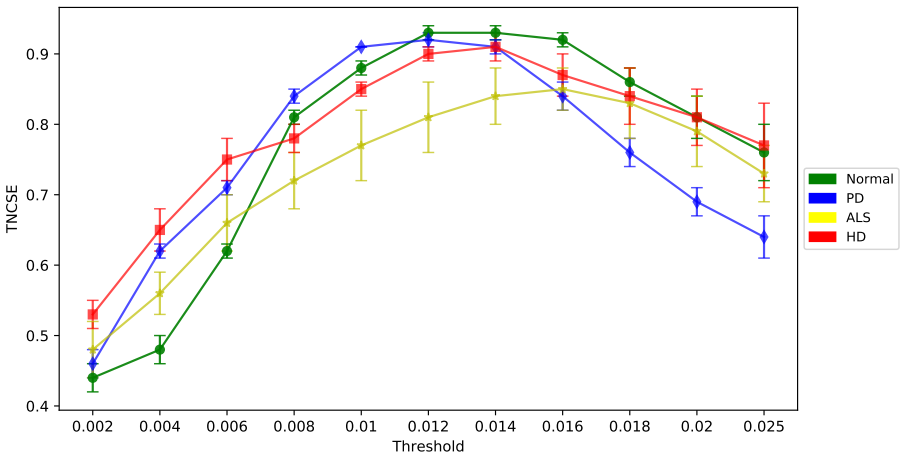


Figure 6.8: Comparison of the SyEn values in term of TNCSE of controlled walking time series derived from the left and right feet (combined) of healthy subjects and those with neurodegenerative diseased conditions. These curves represent the lines connecting the mean $\pm$ standard deviation values of symbolic entropy.

has lower TNCSE values at smaller threshold values. On the other hand, normal walking has higher TNCSE values at larger threshold values. Unlike the other walking conditions, normal walking also has a larger variation in entropy values. The highest value of TNCSE was 0.93 for normal walking at thresholds of 0.012 and 0.014. However, an increase in threshold beyond 0.016 lowered the TNCSE values for normal walking. In Figure 6.9, a box plot depicts how the results of setting arbitrary threshold values result in the distribution of TNCSE values over different walking scenarios.

The boxplot shows that there is no clear degree of separation among different groups, so threshold values were selected in a second step based on the refinement proposed in this study. The SyEn method was used to calculate the TNCSE of normal and neurodegenerative diseased walking conditions. These threshold values include ASI, VSI, SI_{med}, and SI_{mod}. After applying these thresholds, we got a mean value of TNCSE across individual subjects. The values of TNCSE using the refined method for selection of threshold is presented in Figure 6.10.

Figure 6.10 shows that entropy values for normal walking are higher when using ASI or VSI as thresholds, however, SI_{med} and SI_{mod} did not show the good value of TNCSE

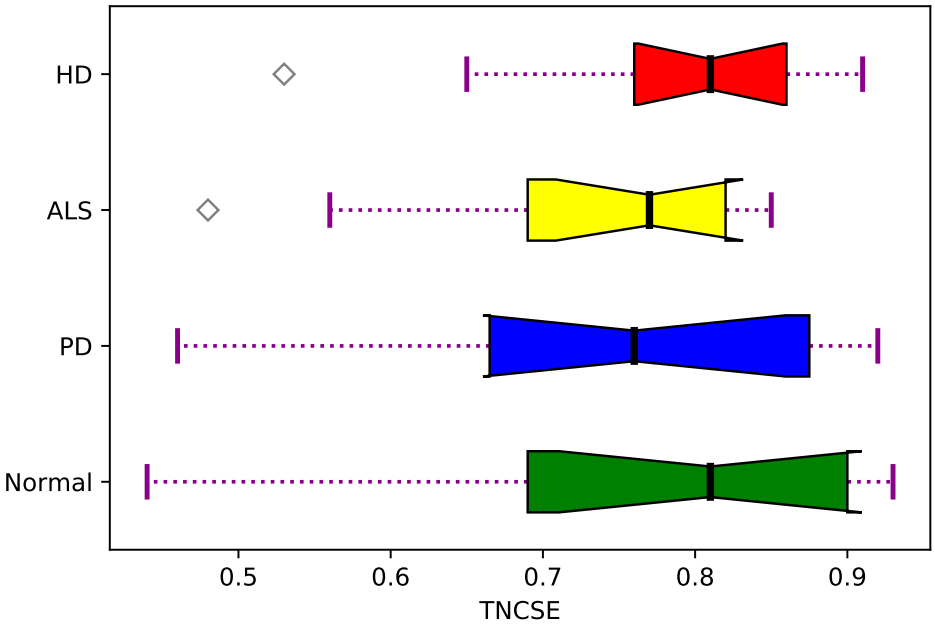


Figure 6.9: Boxplot showing the distribution of TNCSE values while using arbitrary values as the threshold (τ), when applied to stride interval time series of controlled walking and neurodegenerative diseased walking conditions.

for walking in healthy conditions. Also, the separation among the different groups was clearer in case of ASI and VSI. In Table 6.5, the mean$\pm$standard deviation values for TNCSE are presented together with p-values that compare normal walking with neurodegenerative diseased walking on basis of arbitrary threshold selection, and also threshold selection using the proposed method.

For arbitrary thresholds, the maximum values of TNCSE are at a threshold between 0.012 and 0.016 (0.93,0.92) for the stride interval representing normal gait under free conditions. However, at the same time, it is maximum for walking in HD and PD (0.92,0.91). The p-values indicate that the degree of separation between normal and ALS walking is obtained at thresholds o.010 to 0.018 (p-values less than 0.005). We got statistically significant p-values (0.05 or less) separating healthy from diseased conditions at thresholds o.010 to 0.018, but not at all scales. Similarly, at threshold value=0.004, we got a p-value less than 0.005 comparing healthy with diseased conditions. On the other hand, ASI and VSI showed higher TNCSE values for walking under free conditions (0.95 and 0.97) and lower TNCSE values under diseased conditions. The p-values (0.001, 0.002, and 0.003) using these thresholds showed a clear degree of separation. The difference between the two groups (normal and diseased) is quite evident in this case.

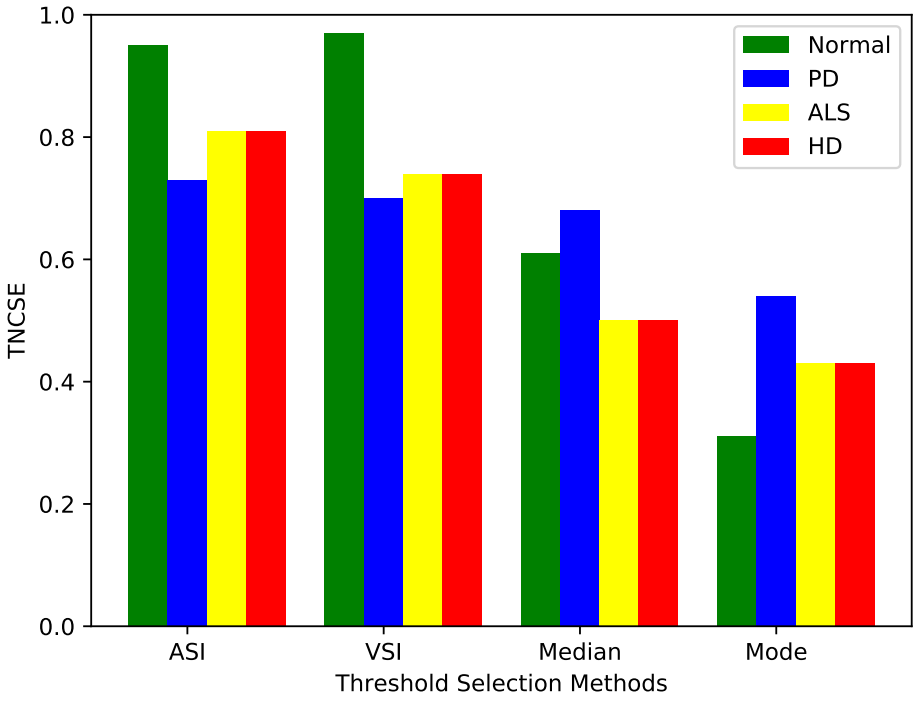


Figure 6.10: Mean TNCSE values using different statistical threshold values for walking under normal and neurodegenerative diseased conditions.

6.5 Discussion

Gait analysis using mathematical algorithms is a fascinating field of research. Many linear and non-linear methods have been proposed for the analysis of human locomotion. These methods are usually applied to the time series derived from the stride intervals during walking. Due to the complex and non-linear nature of the stride interval time series, gait analysis has been conducted using several entropy-based measures of disorder and complexity. In this study, we presented two popular methods used for gait analysis, including multiscale entropy and symbolic entropy. However, there are some limitations in the application of these methods, which may lead to inaccurate classification of gait. Therefore, we have also presented the refinements to these methods and compared both with and without refinements. Validation of these methods requires gait data, which is not always easy to obtain in clinical conditions. However, PhysioNet [282] provides researchers with a useful and easily accessible database to facilitate the development and evaluation of innovative analytical methods. We have also used the data from the PhysioNet in normal and neurodegenerative diseased conditions.

Table 6.5: Mean ± Standard Deviation values of Symbolic entropy (TNCSE) for controlled, and neurodegenerative diseased conditions and corresponding p-values, at different thresholds

Threshold	Symbolic Entropy (Mean ± Standard Deviation)				p-values		
	Normal	ALS	HD	PD	Nor Vs ALS	Nor Vs HD	Nor Vs PD
0.002	0.44±0.02	0.48±0.04	0.53±0.02	0.46±0.02	0.001	0.002	0.006
0.004	0.48±0.02	0.56±0.03	0.65±0.03	0.62±0.01	0.003	0.003	0.003
0.006	0.62±0.01	0.66±0.04	0.75±0.03	0.71±0.01	0.007	0.004	0.003
0.008	0.81±0.01	0.72±0.04	0.78±0.02	0.84±0.01	0.006	0.006	0.004
0.010	0.88±0.01	0.77±0.05	0.85±0.01	0.91±0	0.002	0.005	0.003
0.012	0.93±0.01	0.81±0.05	0.9±0.01	0.92±0.01	0.003	0.004	0.006
0.014	0.93±0.01	0.84±0.04	0.91±0.02	0.91±0.01	0.001	0.001	0.004
0.016	0.92±0.01	0.85±0.03	0.87±0.03	0.84±0.02	0.002	0.001	0.002
0.018	0.86±0.02	0.83±0.05	0.84±0.04	0.76±0.02	0.001	0.004	0.001
0.020	0.81±0.03	0.79±0.05	0.81±0.04	0.69±0.02	0.006	0.006	0.001
0.025	0.76±0.04	0.73±0.04	0.77±0.06	0.64±0.03	0.005	0.006	0.001
ASI	0.95±0.01	0.73±0.01	0.81±0.02	0.83±0.01	0.001	0.003	0.003
VSI	0.97±0.02	0.7±0.03	0.74±0.02	0.72±0.01	0.001	0.002	0.001
Median	0.61±0.07	0.68±0.04	0.5±0.06	0.55±0.05	0.007	0.007	0.007
Mode	0.31±0.07	0.54±0.05	0.43±0.06	0.25±0.09	0.006	0.006	0.006

Those suffering from neurodegenerative diseases and the elderly commonly experience gait dysfunction. Therefore, the database with healthy and neurodegenerative diseased subjects is a useful dataset for validating the mathematical methods of gait analysis.

The first refinement presented here concerns the multiscale entropy method for time series analysis. Initially, this method was applied to heart rate variability analysis using RR (RR) interval time series [263, 288]. In recent years, many studies have also utilized it for stride interval time series analyses [272, 289, 290]. Compared to the RR interval time series, gait involves the repetitive movement of both feet simultaneously. When walking, both feet move simultaneously, resulting in two sets of time series, one for the left foot and one for the right foot. However, in the existing studies, the entropy-based methods utilize either left or right foot gait data as input data during the application of these methods. Data from both feet are not used at the same time.

Our findings also indicated that using the MsEn method to single-foot stride interval time series did not significantly separate both groups at various scales. When using both feet data simultaneously, the entropy values for the healthy controls were high across a wide range of scales. This supports the general phenomenon of complexity analysis that

normal physiologic time series represent more complex behavior than diseased time series. There are two possible reasons for improving the classification. One, of course, is more information is added to the time series when both feet are used. Sometimes, a patient may have a specific problem that affects one leg, and using only that leg data might result in improper results. The second reason is that the amount of data increases when using both feet. During the coarse-graining process, the length of the series is decreased depending on the scale factor. For a time series, a high scale factor results in a very short coarse-grained time series. Therefore, with double-foot data, more data is added to the time series, which overcomes the problem of short-time series.

The second refinement in this study is proposed in the application of the symbolic entropy method while selecting the threshold values. The SyEn is a useful method for complexity analysis in different fields, especially biomedical signals. Applying a threshold-dependent symbolic entropy method to calculate the TNCSE indicates that the time series share common random underlying dynamics at different threshold values. A significant difference between healthy controls and neurodegenerative disease conditions is shown by this method. When TNCSE is high, it suggests that a walking time series is more normal and natural. When TNCSE is low, it indicates that the walking time series is abnormal or diseased. While using arbitrary thresholds, time series representing normal walking showed higher values of TNCSE at t=0.012 to 0.014. However, at these threshold values, the TNCSE values for PD and HD were also elevated. In this case, no variation in TNCSE values was observed.

Using an averaged stride interval as a threshold value showed higher TNCSE values during normal walking, and a clear difference can be observed while comparing it with different diseased subjects. However, this method also showed high variability in TNCSE in almost all cases, which is a serious concern when the length of the time series is short. Variation of stride interval as a threshold value gave significant results, and a clear degree of separation between the two groups (healthy and diseased) was seen. TNCSE values were lower in the case of the latter group, and the variation was large. The other two measures (median and mode) could neither provide significant results in the individual walking condition nor when comparing the two groups.

It was found that ASI or VSI threshold values lead to lower p-values compared to arbitrary threshold values when describing the significance of results. By using the former method, maximum separation is achieved between the healthy and diseased groups.

The "loss of complexity" theory was proposed by [291, 292], which assumed that aging, degeneration, or disease is associated with a reduction in the complexity of signals generated by different physiological systems, including heart rate variability and blood pressure fluctuations. This phenomenon is observed in many diseases. Our findings complemented

this phenomenon when healthy individuals were compared with individuals suffering from neurodegenerative disease, using TNCSE with a threshold value as VSI.

From the results, it is evident that some information may be lost when the mean value is subtracted (as employed in studies [265, 271, 272] from each data point in a time series when correlation among individual variables values is significant. In comparison to other arbitrary threshold selection methods, the proposed method of symbolization using VSI as a threshold value outperformed other methods of symbolization while applying symbolic entropy.

However, there are still some limitations. First, due to the small sample size, heterogeneous clinical manifestation, and staging, it is difficult to obtain a meaningful clinical correlation with the analyzed data. Analysis using these methods shows no correlation between disease severity and analysis results. Second, the age and gender distributions of the current dataset do not match well among the different gait groups. It is unclear how these two factors influence gait complexity.

6.6 Conclusion

The present study aimed to propose two refinements to existing entropy-based gait time series analysis methods. We assessed altered gait variability in healthy and neurodegenerative disease patients using the existing methods and the proposed refinements. To calculate MsEn, the data of both feet before the coarse grain were used in the multiscale entropy analysis method. The results indicated that the MsEn values showed a clearer separation between healthy and diseased subjects compared to a single foot. We also presented a method for selecting threshold values for data symbolization in the symbolic entropy analysis method. The results indicated that the proposed method using VSI as a threshold value improved the classification between healthy and neurodegenerative diseased subjects. The proposed methods could be applied in a variety of biological time series analysis fields.

PART IV

FINAL REMARKS

Chapter 7

Conclusion

The recent developments in computer vision based on deep learning methods have proved to be an exciting field of research in many biomedical problems. New methods of deep learning have focused on designing, implementing, and deploying complex and diverse applications, which can now be seen in various fields including biomedical engineering. Successful case studies result from a prudent, carefully devised fundamental concept that has transformed the way real-world problems are sensed, formalized, and unravelled at an increased level that is machine-centred and automatic. With the technological advancement and availability of high computational power and modern graphic processing units, deep learning has been shown to provide excellent results in the analysis and recognition of medical images and events in recent years.

Gait analysis requires data acquisition and extraction tools of the gait features. For gait analysis and feature extraction, various wearable and non-wearable solutions are proposed in the literature. This dissertation aimed to propose a new method of human gait events detection using marker-less videos captured outside the strict laboratory environments, and thereby contribute novel approaches to the state-of-the-art in this field. The thesis also sought to refine existing mathematical methods for analyzing stride interval time series.

The gait parameters are usually recorded in a gait laboratory under controlled conditions, using either wearable systems or non-wearable systems using floor sensors or multiple cameras. On the other hand, it is known that patients move very consciously and, therefore, unnaturally under a strict laboratory environment. Therefore, there is a risk that the data collected in this way are subject to bias. Furthermore, there is also a need for flexible gait analysis methods that can be used, for example, in day-to-day hospital care. The overall goal of this study was to provide a basis for the development of an inexpensive and easy-to-implement method of obtaining spatiotemporal gait parameters that will then be used to distinguish between normal and abnormal gait patterns.

For this purpose, in the first phase, deep learning models based on convolutional neural network (CNN) were developed, which can extract useful features from the images to classify feet' positions. This CNN was supported by long short term memory network (LSTM) to include the temporal information and keep the walking sequence. Data were collected from volunteers in different walking conditions in the first study. To overcome the problem of the small dataset and class imbalance problem, the data augmentation method was used to enlarge the dataset and maintain the class balance among different classes representing the feet' positions. Furthermore, we utilized an online dataset available for research purposes.

The accuracy of CNN using AlexNet [103] architecture was not adequate enough to be used for extraction of the gait parameters. The variations in predicting the class labels were also high, especially at the last part of the stance and the swing phase of the gait cycle. Adding the LSTM layer to the same CNN-based model improved the accuracy of predicting the feet' positions. We observed that deep learning methods could be used to estimate the gait events from marker-less videos captured under home conditions. Further, gait events can be detected from this classification of feet' positions and be used for a gait analysis system. We also conclude that, in the problems like ours where the sequence and timing of images are important, only CNN may not be an efficient solution all the time. LSTM networks are well suited for processing time-series data where the order is particularly critical, such as sequences of gait data. In essence, LSTM networks benefit from recurrence by utilizing information extracted from a previous cell and forwarding it to the next cell in the network. Thus, merging two or more models is an efficient alternative for problems like this. Hence, hybrid models are useful because they integrate different architectures to improve efficiency by considering the advantages of each architecture.

In the second phase, the important spatiotemporal parameters of the gait were extracted from the gait events predicted by the neural network (NN) using our proposed method. These parameters are useful to analyze the gait dynamics in various filed, including health care and recognition. Before extracting these parameters, we used transfer learning to improve the prediction of gait events by employing deep models because the accuracy and other performance metrics in the prediction of classes in the final model on test data were almost 91% in the previous phase. We also collected new data on gait videos in walking with normal and other simulated pathological conditions. For transfer learning, we used two famous pretrained NNs, inceptionresnetv2 (IRNV-2) and densenet201 (DN-201) [249, 250]. We used these networks' weights for retraining the NN on new data. Transfer learning-based pre-trained NN improved the prediction of labels for different feet' positions. It especially reduced the variations in the predictions in the last stage of the gait swing and stance phases. We were able to achieve an accuracy of 94% in predicting the class labels.

Consequently, the improved model was used for the quantification of spatiotemporal parameters of the gait. We extracted 12 different parameters, including basic parameters; like speed, the number of steps, the number of gait cycles (GCs), and cadence; spatial parameters like step length and stride length; and temporal parameters like step time, walking time, stride time, swing phase, stance phase, and double support phase. Our proposed method estimated spatiotemporal gait parameters for gait data. In particular, the general and spatial gait parameters like 'the total number of steps,' 'total gait cycles,' 'average step length,' 'average stride length,' and the temporal parameter 'walking time' showed excellent accuracy and high correlation with the ground truth. The other general parameters like 'cadence' and 'velocity' also showed a good degree of agreement. The 'stance time' and 'double support, and step time' temporal parameters also showed good accuracy. Only one parameter, i.e., 'swing time,' was in the fair range. There was a high positive correlation in almost all of the parameters. These results provided a scientific basis for applying the proposed methodology and system to clinical studies.

Gait cannot be monitored from one-time events; it requires the parameter values over time. Especially if long-time data is available, the gait analysis will be more accurate. Therefore, the gait parameters result in a time series having values at a certain period. Then the resultant time series could be used for subjective gait analysis. As the third contribution in this dissertation, our objective was to improve the existing mathematical methods of time series analysis of temporal gait data. For this purpose, we proposed two refinements to existing entropy-based methods for stride interval time series analysis. We assessed altered gait variability in healthy and neurodegenerative disease patients to validate the proposed methods. We utilized both feet' data before coarse-graining in the multiscale entropy analysis method to calculate multiscale entropy (MsEn). The results indicated that the MsEn values showed a clearer separation between healthy and diseased subjects compared to a single foot. We also presented a method for selecting threshold values for data symbolization in the symbolic entropy analysis method. The results indicated that the proposed method of variation in stride interval (VSI) as a threshold value improved the classification between healthy and neurodegenerative diseased subjects.

In conclusion, this thesis has developed methods that can partially overcome the existing difficulties in gait analysis regarding the simplicity of the data collection approach, long-term data availability, comfortability for subjects in terms of non-wearable systems, and portability. These can also be used to avoid strict laboratory environments, which create extra pressure on the subjects, and they might not be able to provide their actual gait conditions because of dual tasks (data recording and thinking about the procedure). Using a frontal view to perform gait analysis reduces the physical space required for the test. By using these

methods, better artificial intelligence-based medical support systems can be designed, which will assist physicians with clinical routines, such as diagnosis, therapy, and monitoring, thereby reducing clinical workload and improving patient safety.

Having demonstrated these performances of the proposed models on gait videos, it is apparent that the temporal information encoded between the gait events may not be accurately detected by only CNN based models; therefore, combining the CNN based models with LSTM based models could potentially increase the classification of feet' position. Hybrid models are characterized by the integration of different architectures to improve efficiency by considering the advantages of each architecture. It is pertinent to note that when comparing the accuracy improvements obtained with the processing time of the LSTM and CNN-LSTM models, the time difference is not that significant, thus making the suggested architecture suitable for the detection of feet' positions in gait.

The inbuilt capability of pre-trained CNNs is also helpful in extracting the relevant features from the images of the gait. It is generally expected that feature maps close to the input will detect small or fine-grained details, whereas feature maps near the output will capture more general features. It is observed that the proposed models detected better and more relevant features, especially in the lower filters in starting layers.

While comparing with some prior studies of gait analysis using camera-based technologies where the authors used OpenPose to investigate features of walking or other human movement patterns [163, 206–211], we anticipate that the methodology used in these studies may be capable of producing more accurate results, and the accuracy of the final model is a little low. Nevertheless, it is noteworthy that marker-less video-based gait event detection (without pose estimation) is best suited to the clinic and home-based gait analysis, and requires no other equipment than a smartphone camera for data collection. In contrast, many other methods require expensive, inaccessible, and less portable equipment.

However, this thesis also has certain limitations, like the availability of real patient data, which may be a better representation and could be used for inter-group classification. This study involved healthy volunteers who walked according to a pattern designed to mimic pathologic gaits. It is unclear whether this computer vision-based system is effective at classifying patients' gaits in a clinical setting. We did not include the turns while walking, which have great potential to represent certain gait parameters because difficulty in turning is also a symptom of some gait-related problems. We only considered the controlled and diseased groups while using time series analysis and did not make any inter-group analysis in the case of diseased subjects.

Chapter 8

Outlook

Based on the proposed study, we have outlined some findings and set guidelines for future work. Although the computed results are promising, there is a room to enhance the performance of the models in detecting gait events. In this regard, fine-tuned CNN-LSTM models for extracting deep features from different convolutional neural networks (CNNs) can improve classification accuracy and event recognition rates in the future.

However, it is also imperative that enough original data be available to test and develop new methods in the future. This work can also be extended by recording videos in real-time home scenarios where multiple persons are walking in a room or corridor. Other potential future work directions include collecting real patient data from different hospitals and carrying an analysis in inter-diseased conditions. This data will help to distinguish different gait related diseases. This work may be supported using floor sensors or foot switches for ground truth determination while recording the gait videos.

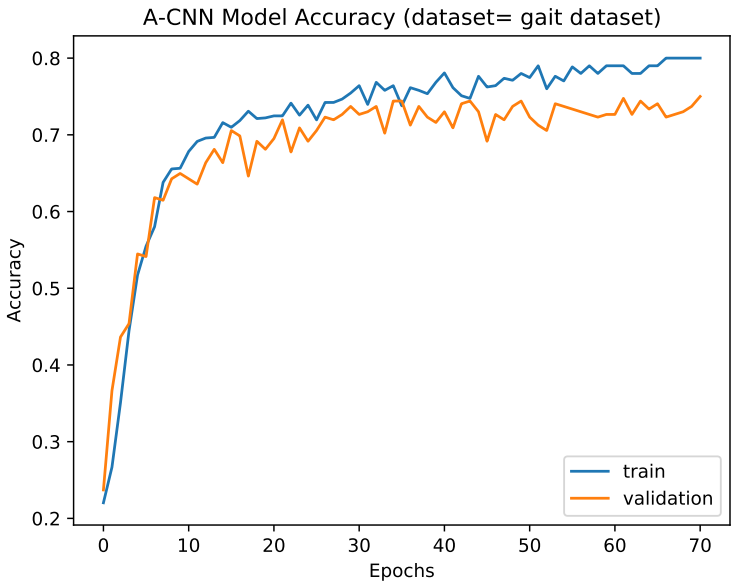
Another future work could be the use of time taken to make turns while walking because in certain diseases turning contains useful information. Despite the fact that we included gait freezing in terms of the total time taken, this could also be further investigated in terms of after how many steps or after how much time the patient's gait freezes.

Appendix A

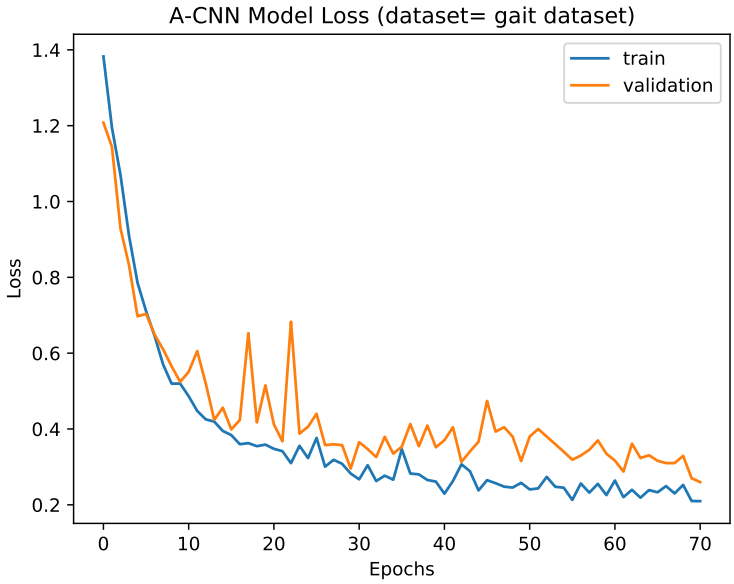
Results of Feet' Positions Detection: Training and Validation

Training and Validation Accuracy and Loss of neural networks (NNs)

Here we are presenting the training and validation accuracy and loss curves of different deep NNs, we trained with gait data.

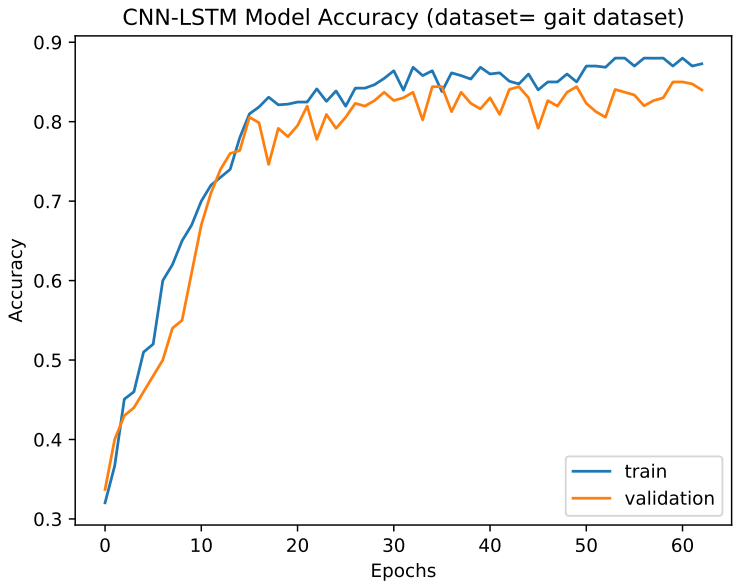


(a) A-CNN training and validation accuracy

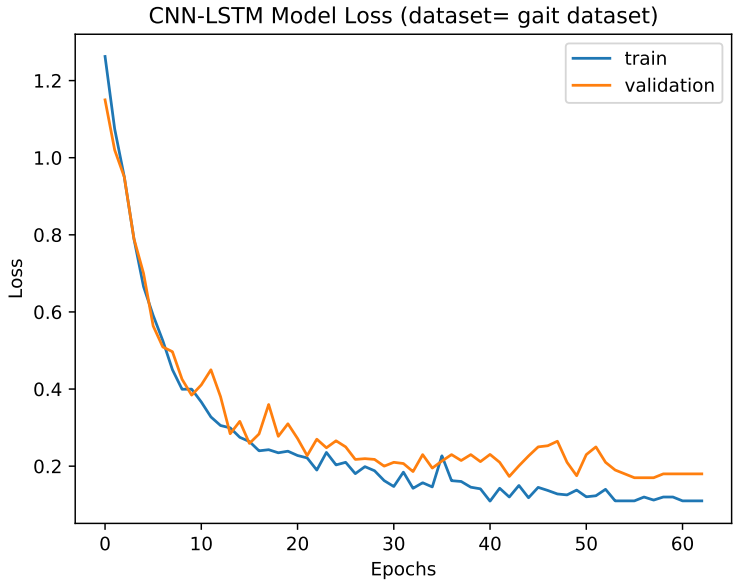


(b) A-CNN training and validation loss

Figure A.1: A-CNN model training and validation performance when using only gait data collected from videos. training and validation accuracy (a), training and validation loss (b)

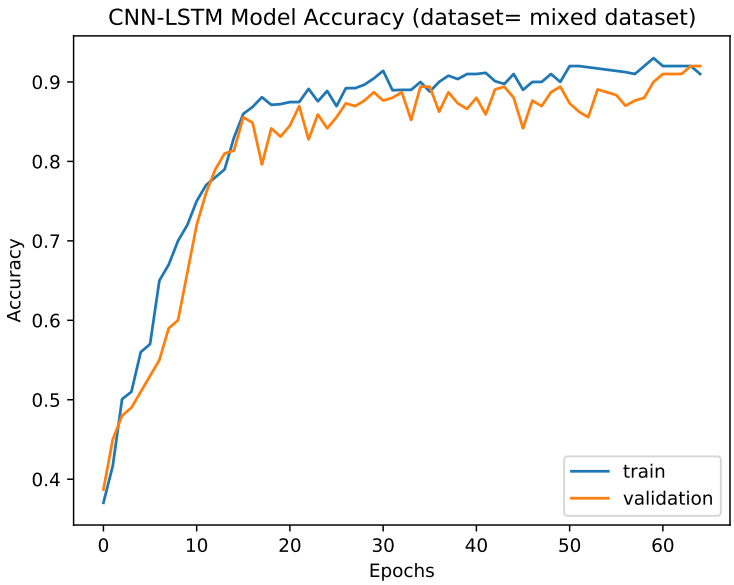


(a) CNN-LSTM training and validation accuracy

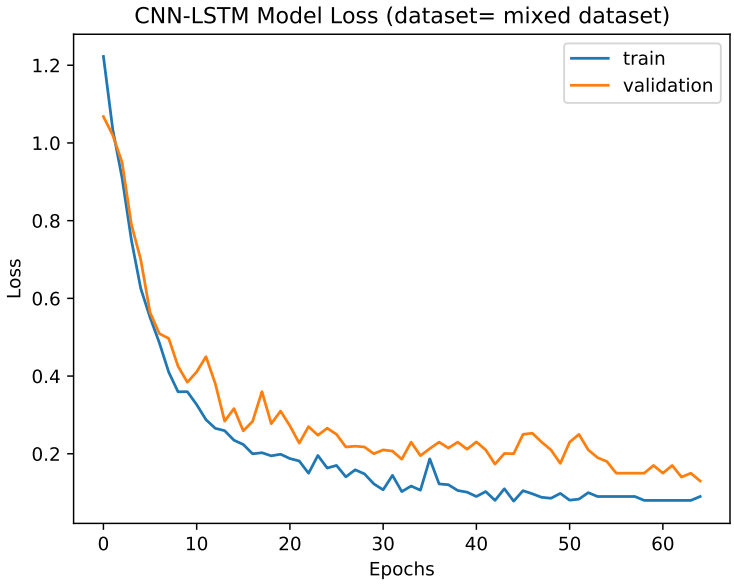


(b) CNN-LSTM training and validation loss

Figure A.2: CNN-LSTM model training and validation performance when using only gait data collected from videos. training and validation accuracy (a), training and validation loss (b)

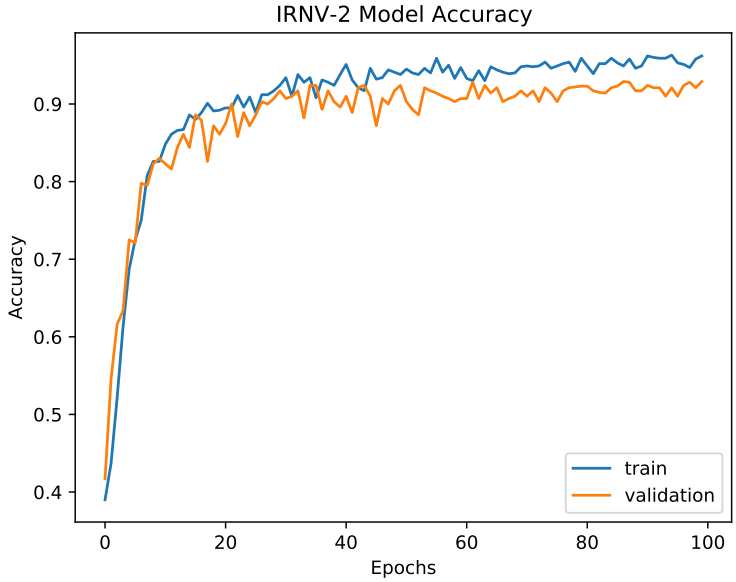


(a) CNN-LSTM training and validation accuracy

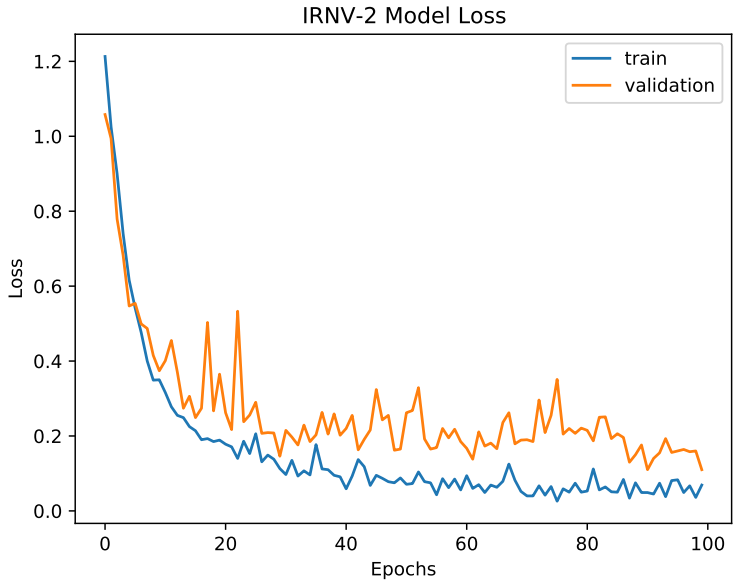


(b) CNN-LSTM training and validation loss

Figure A.3: CNN-LSTM model training and validation performance when using mixed dataset. training and validation accuracy (a), training and validation loss (b)

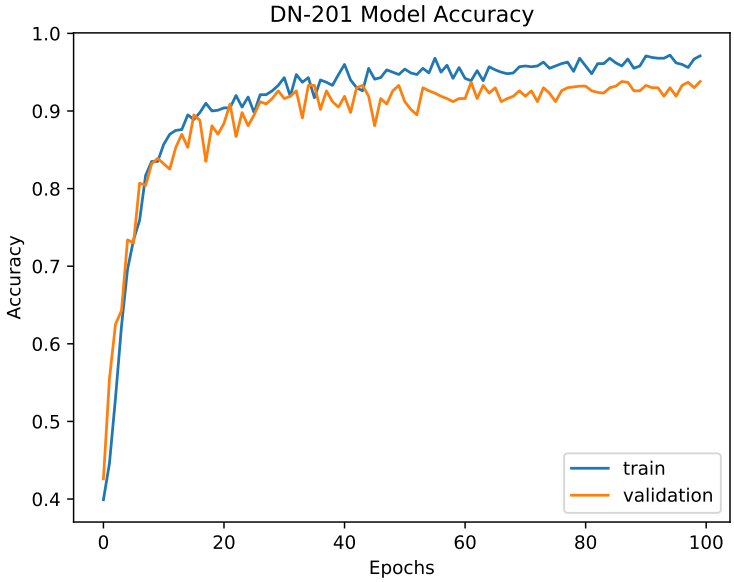


(a) inceptionresnetv2 (IRNV-2) training and validation accuracy

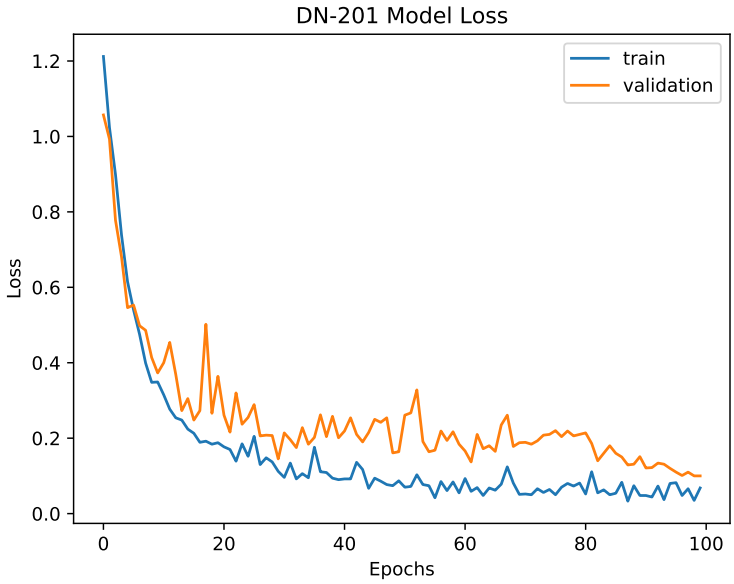


(b) inceptionresnetv2 (IRNV-2) training and validation loss

Figure A.4: inceptionresnetv2 (IRNV-2) model training and validation performance . training and validation accuracy (a), training and validation loss (b)



(a) densenet201 (DN-201) training and validation accuracy



(b) densenet201 (DN-201) training and validation loss

Figure A.5: densenet201 (DN-201) model training and validation performance . training and validation accuracy (a), training and validation loss (b)

Appendix **B**

Quantification of Spatiotemporal Parameters

Spatiotemporal parameters estimation:

Here we are presenting the correlations between the spatiotemporal parameters obtained from neural networks (NNs) based gait events and ground truth (through observation), in diseased conditions.

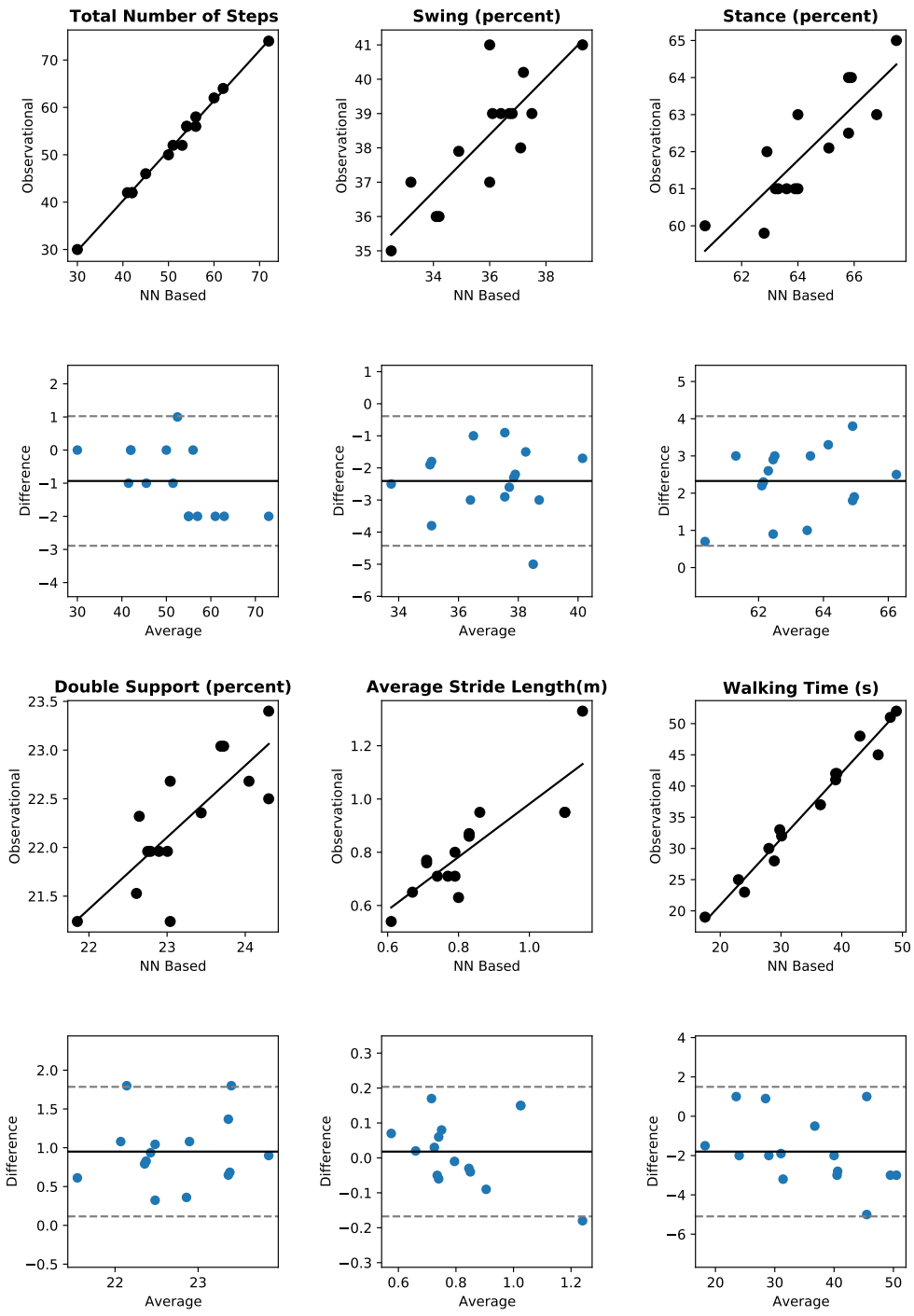


Figure B.1: Scatter and Bland–Altman plots of the 6 different spatiotemporal parameters of the gait (number of steps, swing(%age), stance (%age), double support (%age), average stride interval (m), and walking time (s)) quantified by the NN based predicted labels and the ground truth in **alzheimer's disease walking (**AD_Walk**)** conditions. The upper plot is the scatter plot, and the lower plot is the Bland–Altman plot for each parameter. In the lower row, the Y-axis of the plot corresponds to the difference between the two measurements (predicted parameters (Pred_Params) and true parameters (True_Params)), whereas the X-axis is the mean of the two measurements. The solid lines show the average difference for the whole sample, and the dashed lines correspond to the 95% limits of agreement on both sides.

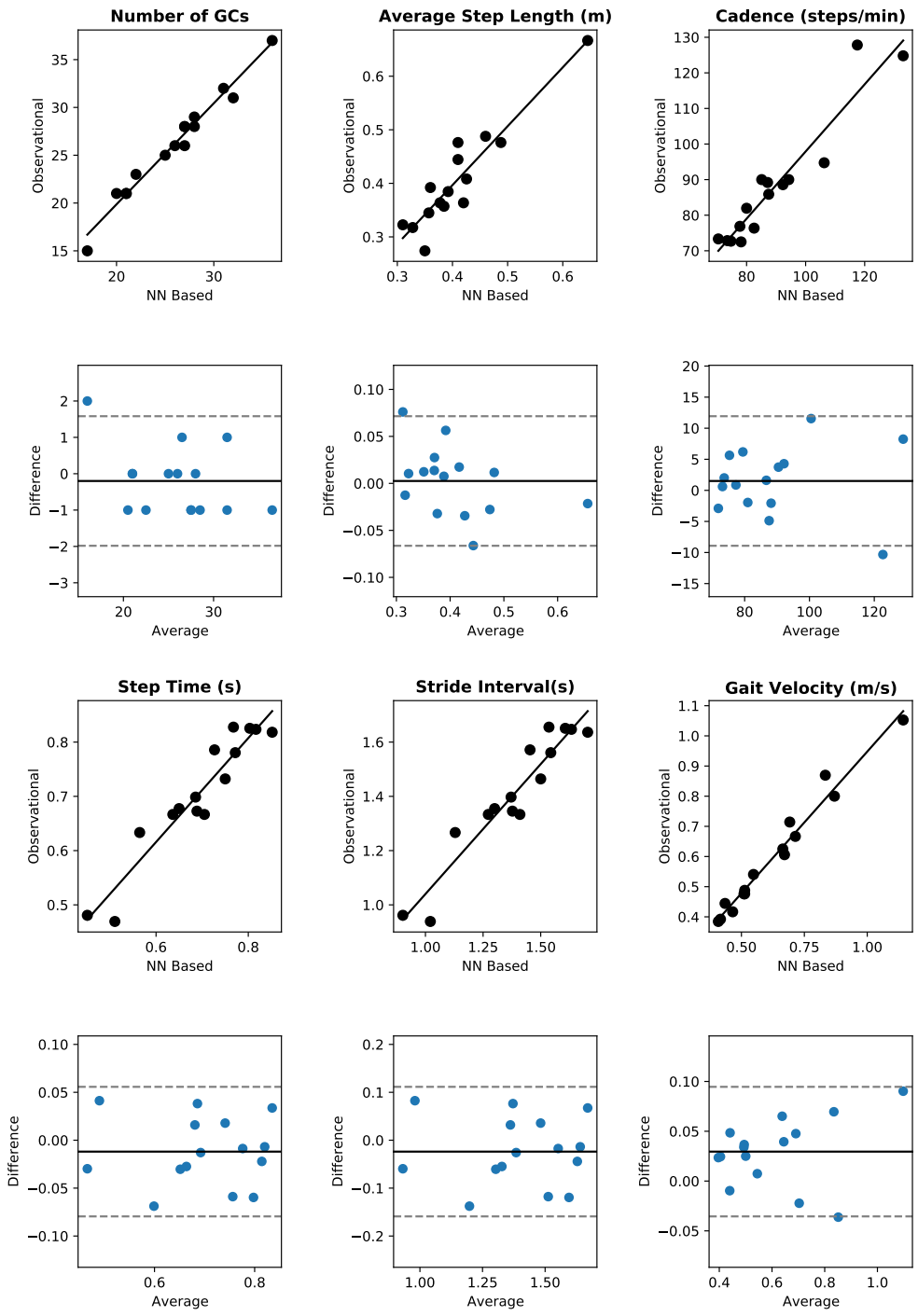


Figure B.2: Scatter and Bland–Altman plots of the 6 different spatiotemporal parameters of the gait (number of gait cycles (GCs), average step length(m), cadence (steps/min), step time (s), stride interval (s), and gait velocity (ms)) quantified by the NN based predicted labels and the ground truth in **alzheimer's disease walking (*AD_Walk*)** conditions. The upper plot is the scatter plot, and the lower plot is the Bland–Altman plot for each parameter. In the lower row, the Y-axis of the plot corresponds to the difference between the two measurements (predicted parameters (Pred_Params) and true parameters (True_Params)), whereas the X-axis is the mean of the two measurements. The solid lines show the average difference for the whole sample, and the dashed lines correspond to the 95% limits of agreement on both sides.

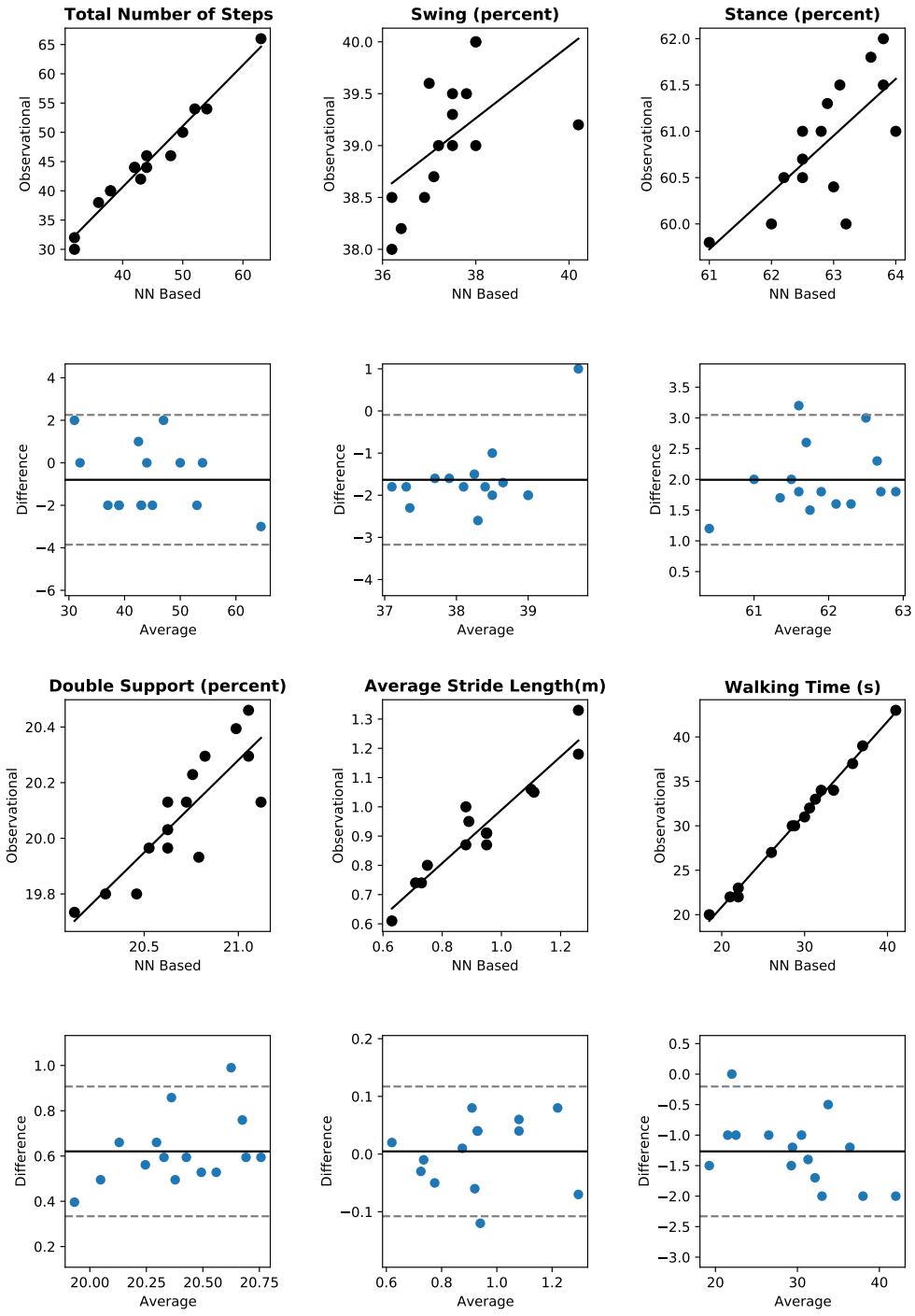


Figure B.3: Scatter and Bland–Altman plots of the 6 different spatiotemporal parameters of the gait (number of steps, swing(%age), stance (%age), double support (%age), average stride interval (m), and walking time (s)) quantified by the NN based predicted labels and the ground truth in **NPH walking (**NPH_Walk**)** conditions. The upper plot is the scatter plot, and the lower plot is the Bland–Altman plot for each parameter. In the lower row, the Y-axis of the plot corresponds to the difference between the two measurements (predicted parameters (Pred_Params) and true parameters (True_Params)), whereas the X-axis is the mean of the two measurements. The solid lines show the average difference for the whole sample, and the dashed lines correspond to the 95% limits of agreement on both sides.

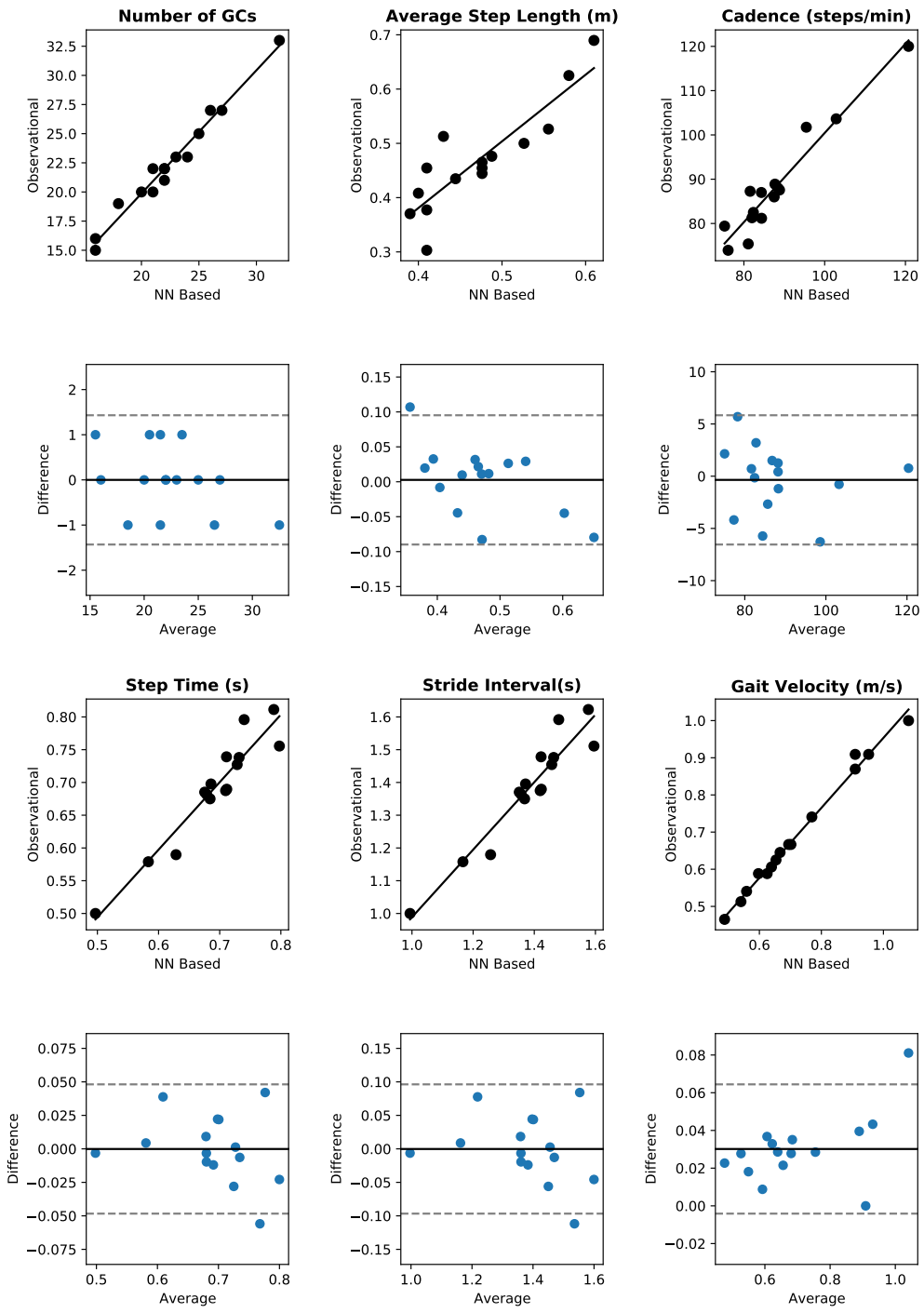


Figure B.4: Scatter and Bland–Altman plots of the 6 different spatiotemporal parameters of the gait (number of gait cycles (GCs), average step length(m), cadence (steps/min), step time (s), stride interval (s), and gait velocity (ms)) quantified by the NN based predicted labels and the ground truth in **NPH walking (*NPH_Walk*)** conditions. The upper plot is the scatter plot, and the lower plot is the Bland–Altman plot for each parameter. In the lower row, the Y-axis of the plot corresponds to the difference between the two measurements (predicted parameters (Pred_Params) and true parameters (True_Params)), whereas the X-axis is the mean of the two measurements. The solid lines show the average difference for the whole sample, and the dashed lines correspond to the 95% limits of agreement on both sides.

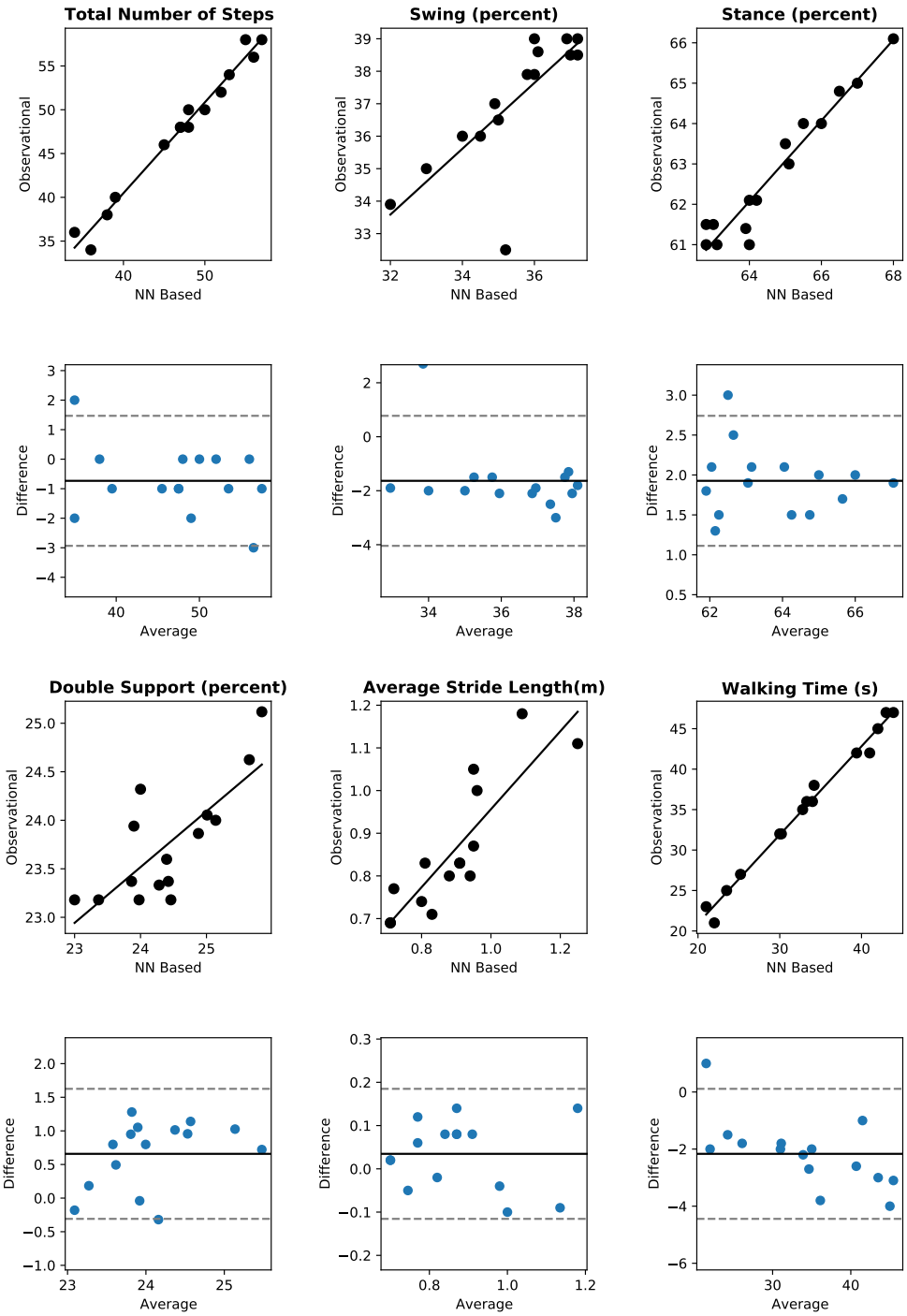


Figure B.5: Scatter and Bland–Altman plots of the 6 different spatiotemporal parameters of the gait (number of steps, swing(%age), stance (%age), double support (%age), average stride interval (m), and walking time (s)) quantified by the NN based predicted labels and the ground truth in **injury walking (*Injury_Walk*)** conditions. The upper plot is the scatter plot, and the lower plot is the Bland–Altman plot for each parameter. In the lower row, the Y-axis of the plot corresponds to the difference between the two measurements (predicted parameters (Pred_Params) and true parameters (True_Params)), whereas the X-axis is the mean of the two measurements. The solid lines show the average difference for the whole sample, and the dashed lines correspond to the 95% limits of agreement on both sides.

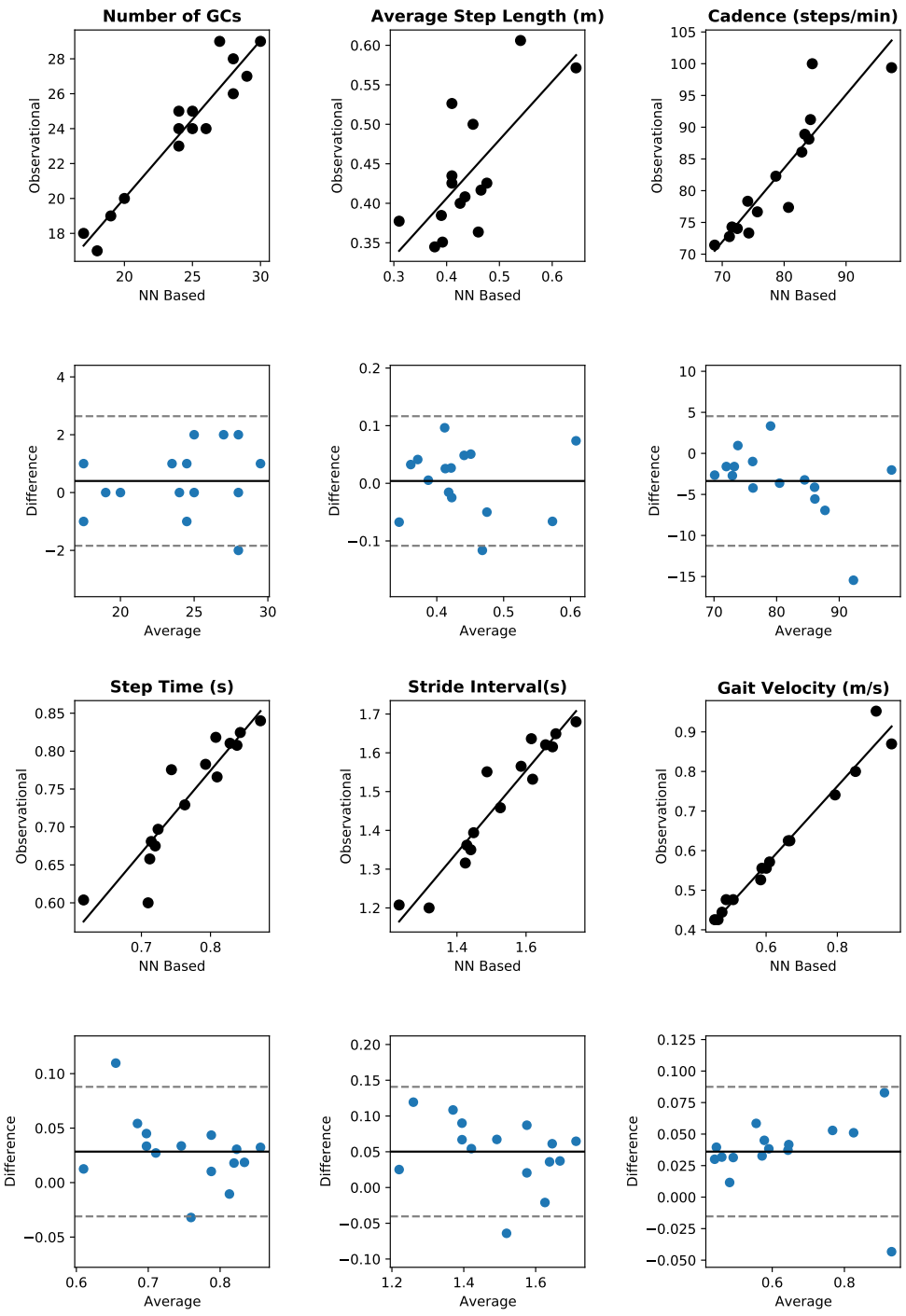


Figure B.6: Scatter and Bland–Altman plots of the 6 different spatiotemporal parameters of the gait (number of gait cycles (GCs), average step length(m), cadence (steps/min), step time (s), stride interval (s), and gait velocity (ms)) quantified by the NN based predicted labels and the ground truth in **injury walking (*Injury_Walk*)** conditions. The upper plot is the scatter plot, and the lower plot is the Bland–Altman plot for each parameter. In the lower row, the Y-axis of the plot corresponds to the difference between the two measurements (predicted parameters (Pred_Params) and true parameters (True_Params)), whereas the X-axis is the mean of the two measurements. The solid lines show the average difference for the whole sample, and the dashed lines correspond to the 95% limits of agreement on both sides.

List of Figures

List of Tables

References

[1] S. uk Ko, G. J. Jerome, E. M. Simonsick, et al., "Differential gait patterns by history of falls and knee pain status in healthy older adults: Results from the baltimore longitudinal study of aging," *Journal of Aging and Physical Activity*, vol. 26, pp. 577–582, 10 2018.

[2] A. Ardalan, N. Yamane, A. K. Rao, et al., "Analysis of gait synchrony and balance in neurodevelopmental disorders using computer vision techniques." *Health informatics journal*, vol. 27, p. 14604582211055650.

[3] S. Rupprechter, G. Morinan, Y. Peng, et al., "A clinically interpretable computer-vision based method for quantifying gait in parkinson's disease," *Sensors*, vol. 21, p. 5437, 8 2021.

[4] B. Moreland, R. Kakara, and A. Henry, "Trends in nonfatal falls and fall-related injuries among adults aged 65 years - united states, 2012-2018." *MMWR. Morbidity and mortality weekly report*, vol. 69, pp. 875–881, 7 2020.

[5] C. S. Florence, G. Bergen, A. Atherly, et al., "Medical costs of fatal and nonfatal falls in older adults." *Journal of the American Geriatrics Society*, vol. 66, pp. 693–698, 2018.

[6] Y. K. Haddad, G. Bergen, and C. S. Florence, "Estimating the economic burden related to older adult falls by state," *Journal of Public Health Management and Practice*, vol. 25, pp. E17–E24, 3 2019.

[7] M. Yamada and N. Ichihashi, "Predicting the probability of falls in community-dwelling elderly individuals using the trail-walking test." *Environmental health and preventive medicine*, vol. 15, pp. 386–91, 11 2010.

[8] E. Schönfelder, A. Osmanovic, L. H. Müschen, et al., "Costs of illness in amyotrophic lateral sclerosis (als): a cross-sectional survey in germany," *Orphanet Journal of Rare Diseases*, vol. 15, p. 149, 12 2020.

[9] G. Palestra, M. Rebiai, E. Courtial, et al., "Evaluation of a rehabilitation system for the elderly in a day care center," *Information*, vol. 10, p. 3, 12 2018.

[10] L. Lonini, Y. Moon, K. Embry, et al., "Video-based pose estimation for gait analysis in stroke survivors during clinical assessments: A proof-of-concept study," *Digital Biomarkers*, vol. 6, pp. 9–18, 1 2022.

[11] F. I. MAHONEY and D. W. BARTHEL, "Functional evaluation: The barthel index." *Maryland state medical journal*, vol. 14, pp. 61–5, 2 1965.

[12] P. W. Tipton, "Dissecting parkinsonism: cognitive and gait disturbances," *Neurologia i Neurochirurgia Polska*, vol. 55, pp. 513–524, 12 2021.

[13] K.-D. Ng, S. Mehdizadeh, A. Iaboni, et al., "Measuring gait variables using computer vision to assess mobility and fall risk in older adults with dementia." *IEEE journal of translational engineering in health and medicine*, vol. 8, p. 2100609, 2020.

[14] C. Clinic. Gait Disorders-online, Accessed: 5 oct 2022. https://my.clevelandclinic.org/health/symptoms/21092-gait-disorders.

[15] R. A. Mann and J. Hagy, "Biomechanics of walking, running, and sprinting," *The American Journal of Sports Medicine*, vol. 8, pp. 345–350, 9 1980.

[16] J. Perry and J. M. Burnfield, *Gait analysis: Normal and pathological function*, 2nd ed., vol. 9. Journal of Sports Science Medicine, 1992.

[17] C. Tunca, N. Pehlivan, N. Ak, et al., "Inertial sensor-based robust gait analysis in non-hospital settings for neurological disorders," *Sensors (Switzerland)*, vol. 17, 4 2017.

[18] I. Blazquez and Hernandez. (2002) Foot Pronation and Supination-online, Accessed: 5 oct 2022. https://www.angelfire.com/la/Ivan/gait.html.

[19] Physiopedia. Gait-online. https://www.physio-pedia.com/Gait.

[20] A. Maksimovic, R. Hanewinckel, V. J. Verlinden, et al., "Gait characteristics in older adults with diabetes and impaired fasting glucose: The rotterdam study," *Journal of Diabetes and its Complications*, vol. 30, pp. 61–66, 1 2016.

[21] M. W. Whittle, *Normal gait. gait analysis: An introduction. edited by whittle mw*. Oxford, Butterworth-Heinemann, 1991.

[22] K. Base. Spatiotemporal Gait Parameters-online, Accessed: 5 oct 2022. https://help.plantiga.com/spatiotemporal-gait-parameters.

[23] M. Bertoli, A. Cereatti, D. Trojaniello, et al., "Estimation of spatio-temporal parameters of gait from magneto-inertial measurement units: multicenter validation among parkinson, mildly cognitively impaired and healthy older adults," *BioMedical Engineering OnLine*, vol. 17, p. 58, 12 2018.

[24] S. S. Kuys, N. M. Peel, K. Klein, et al., "Gait speed in ambulant older people in long term care: A systematic review and meta-analysis," *Journal of the American Medical Directors Association*, vol. 15, pp. 194–200, 3 2014.

[25] N. M. Peel, S. S. Kuys, and K. Klein, "Gait speed as a measure in geriatric assessment in clinical settings: A systematic review," *The Journals of Gerontology: Series A*, vol. 68, pp. 39–46, 1 2013.

[26] J. Looper and L. S. Chandler, "How do toddlers increase their gait velocity?" *Gait Posture*, vol. 37, pp. 631–633, 4 2013.

[27] R. Cham, S. A. Studenski, S. Perera, et al., "Striatal dopaminergic denervation and gait in healthy adults," *Experimental Brain Research*, vol. 185, pp. 391–398, 3 2008.

[28] T. V. BARREIRA, P. T. KATZMARZYK, W. D. JOHNSON, et al., "Cadence patterns and peak cadence in us children and adolescents," *Medicine Science in Sports Exercise*, vol. 44, pp. 1721–1727, 9 2012.

[29] G. Durward, C. N. Pugh, L. Ogunremi, et al., "Detection of risk of falling and hip fracture in women referred for bone densitometry," *The Lancet*, vol. 354, pp. 220–221, 7 1999.

[30] R. L. Cromwell, P. M. Meyers, P. E. Meyers, et al., "Tae kwon do: An effective exercise for improving balance and walking ability in older adults," *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 62, pp. 641–646, 6 2007.

[31] J. Xu, T. Bao, U. H. Lee, et al., "Configurable, wearable sensing and vibrotactile feedback system for real-time postural balance and gait training: proof-of-concept," *Journal of Neuro-Engineering and Rehabilitation*, vol. 14, p. 102, 12 2017.

[32] A. M. de-la Herran, B. Garcia-Zapirain, and A. Mendez-Zorrilla, "Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications," *Sensors*, vol. 14, pp. 3362–3394, 2 2014.

[33] J. A. Simancek, "Assessment," pp. 12–25, 2013.

[34] PMRknowledge. Biomechanic of Gait and Treatment of Abnormal Gait Patterns-online, Accessed on 05 oct 2022. https://now.aapmr.org/biomechanic-of-gait-and-treatment-of-abnormal-gait-patterns.

[35] A. Esquenazi and M. Talaty, "Gait analysis: Technology and clinical applications," *Physical medicine and rehabilitation*, pp. 93–108, 2000.

[36] Y. P. Ivanenko, N. Dominici, and F. Lacquaniti, "Development of independent walking in toddlers." *Exercise and sport sciences reviews*, vol. 35, pp. 67–73, 4 2007.

[37] D. H. Sutherland, R. Olshen, L. Cooper, et al., "The development of mature gait." *The Journal of bone and joint surgery. American volume*, vol. 62, pp. 336–53, 4 1980.

[38] W. Pirker and R. Katzenschlager, "Gait disorders in adults and the elderly : A clinical guide." *Wiener klinische Wochenschrift*, vol. 129, pp. 81–95, 2 2017.

[39] A. D. Gitler, P. Dhillon, and J. Shorter, "Neurodegenerative disease: models, mechanisms, and a new hope." *Disease models mechanisms*, vol. 10, pp. 499–502, 2017.

[40] M. Bahlo, R. Bowden, P. Czabota, et al. Neurodegenerative disorders-online,Accessed on 05 oct 2022. https://www.wehi.edu.au/research-diseases/development-and-ageing/neurodegenerative-disorders.

[41] Parkinson Disease-online, Accessed on 05 oct 2022. https://www.coursehero.com/file/26691760/DPdocx/?cv=1.

[42] Parkinson's Disease-online, Accessed on 05 oct 2022. https://lsom.uthscsa.edu/neurosurgery/clinical-practice/parkinsons-disease/.

[43] R. Selzler, J. R. Green, and R. Goubran, "Neurodegenerative disease prediction based on gait analysis signals acquired with force-sensitive resistors." IEEE, 10 2018, pp. 122–125.

[44] R. Baker, A. Esquenazi, M. G. Benedetti, et al., "Gait analysis: Clinical facts," *European Journal of Physical and Rehabilitation Medicine*, vol. 52, pp. 560–574, 8 2016.

[45] L. di Biase, A. D. Santo, M. L. Caminiti, et al., "Gait analysis in parkinson's disease: An overview of the most accurate markers for diagnosis and symptoms monitoring," *Sensors*, vol. 20, p. 3529, 6 2020.

[46] A. Mirelman, P. Bonato, R. Camicioli, et al., "Gait impairments in parkinson's disease," *The Lancet Neurology*, vol. 18, pp. 697–708, 7 2019.

[47] A. Suppa, A. Kita, G. Leodori, et al., "l-dopa and freezing of gait in parkinson's disease: Objective assessment through a wearable wireless system," *Frontiers in Neurology*, vol. 8, 8 2017.

[48] J. M. Hausdorff, "Gait dynamics in parkinson's disease: Common and distinct behavior among stride length, gait variability, and fractal-like scaling," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 19, p. 026113, 6 2009.

[49] M. Plotnik, N. Giladi, Y. Balash, et al., "Is freezing of gait in parkinson's disease related to asymmetric motor function?" *Annals of Neurology*, vol. 57, pp. 656–663, 5 2005.

[50] M. Plotnik, N. Giladi, and J. M. Hausdorff, "Bilateral coordination of walking and freezing of gait in parkinson's disease." *The European journal of neuroscience*, vol. 27, pp. 1999–2006, 4 2008.

[51] R. Bartsch, M. Plotnik, J. W. Kantelhardt, et al., "Fluctuation and synchronization of gait intervals and gait force profiles distinguish stages of parkinson's disease," *Physica A: Statistical Mechanics and its Applications*, vol. 383, pp. 455–465, 9 2007.

[52] G. Yogev, M. Plotnik, C. Peretz, et al., "Gait asymmetry in patients with parkinson's disease and elderly fallers: when does the bilateral coordination of gait require attention?" *Experimental Brain Research*, vol. 177, pp. 336–346, 2 2007.

[53] O. Blin, A. Ferrandez, and G. Serratrice, "Quantitative analysis of gait in parkinson patients: increased variability of stride length," *Journal of the Neurological Sciences*, vol. 98, pp. 91–97, 8 1990.

[54] S. Frenkel-Toledo, N. Giladi, C. Peretz, et al., "Treadmill walking as an external pacemaker to improve gait rhythm and stability in parkinson's disease," *Movement Disorders*, vol. 20, pp. 1109–1114, 9 2005.

[55] J. M. Hausdorff, M. E. Cudkowicz, R. Firtion, et al., "Gait variability and basal ganglia disorders: Stride-to-stride variations of gait cycle timing in parkinson's disease and huntington's disease," *Movement Disorders*, vol. 13, pp. 428–437, 5 1998.

[56] P. Foundaiton. Statistics: Who has Parkinson-online, Accessed on 05 oct 2022. https://www.parkinson.org/understanding-parkinsons/statistics.

[57] R. Anand, K. D. Gill, and A. A. Mahdi, "Therapeutics of alzheimer's disease: Past, present and future." *Neuropharmacology*, vol. 76 Pt A, pp. 27–50, 1 2014.

[58] A. Association. Alzheimer's and Dementia-online, Accessed on 05 oct 2022. https://www.alz.org/alzheimers-dementia/what-is-alzheimers.

[59] C. Goodman and K. Fuller, *Pathology implications for the physical therapist*, 5th ed. Elsvier, 2020.

[60] A. H. Ropper, M. A. Samuels, J. P. Klein, et al., *Principles of neurology*, 11th ed. McGraw Hill, 2019.

[61] T. A. Ala and W. H. Frey, "Validation of the nincds-adrda criteria regarding gait in the clinical diagnosis of alzheimer disease. a clinicopathologic study." *Alzheimer disease and associated disorders*, vol. 9, pp. 152–9, 1995.

[62] S. Della and S. D. Sala, "Walking difficulties in patients with alzheimer's disease might originate from gait apraxia," pp. 196–201, 2004.

[63] J.-L. Fuh, K.-N. Lin, S.-J. Wang, et al., "Neurologic diseases presenting with gait impairment in the elderly," *Journal of Geriatric Psychiatry and Neurology*, vol. 7, pp. 89–92, 4 1994.

[64] N. B. Alexander, J. M. Mollo, B. Giordani, et al., "Maintenance of balance, gait patterns, and obstacle clearance in alzheimer's disease," *Neurology*, vol. 45, pp. 908–914, 5 1995.

[65] J. Verghese, R. B. Lipton, C. B. Hall, et al., "Abnormality of gait as a predictor of non-alzheimer's dementia," *New England Journal of Medicine*, vol. 347, pp. 1761–1768, 11 2002.

[66] J. C. Morris, E. H. Rubin, E. J. Morris, et al., "Senile dementia of the alzheimer's type: An important risk factor for serious falls," *Journal of Gerontology*, vol. 42, pp. 412–417, 7 1987.

[67] "2021 alzheimer's disease facts and figures," *Alzheimer's Dementia*, vol. 17, pp. 327–406, 3 2021.

[68] D. Watt, "Huntington's disease: A disorder of families. by s. e. folstein. (pp. 251; £25.00.) johns hopkins university press: Baltimore. 1989." *Psychological Medicine*, vol. 20, pp. 728–731, 8 1990.

[69] About Huntington's Disease and Related Disorders-online. https://www.hopkinsmedicine.org/psychiatry/specialty_areas/huntingtons_disease/patient_family_resources/education_whatis.html.

[70] W. C. Koller and J. Trimble, "The gait abnormality of huntington's disease." *Neurology*, vol. 35, pp. 1450–4, 10 1985.

[71] J. M. Hausdorff, S. L. Mitchell, R. Firtion, et al., "Altered fractal dynamics of gait: reduced stride-interval correlations with aging and huntington's disease," *Journal of Applied Physiology*, vol. 82, pp. 262–269, 1 1997.

[72] S. C. Warby, A. Montpetit, A. R. Hayden, et al., "Cag expansion in the huntington disease gene is associated with a specific and targetable predisposing haplogroup." *American journal of human genetics*, vol. 84, pp. 351–66, 3 2009.

[73] C. AJ, "Huntington's chorea; a clinical problem in east anglia."

[74] O. W. Quarrell, A. Tyler, M. P. Jones, et al., "Population studies of huntington's disease in wales." *Clinical genetics*, vol. 33, pp. 189–95, 3 1988.

[75] S. A. Simpson and A. W. Johnston, "The prevalence and patterns of care of huntington's chorea in grampian." *The British journal of psychiatry : the journal of mental science*, vol. 155, pp. 799–804, 12 1989.

[76] P. S. Harper, "The epidemiology of huntington's disease." *Human genetics*, vol. 89, pp. 365–76, 6 1992.

[77] NIH. Huntington's Disease-online,Accessed on. 05 oct 2022. https://www.ninds.nih.gov/health-information/disorders/huntingtons-disease.

[78] R. H. Brown, "Amyotrophic lateral sclerosis," *Archives of Neurology*, vol. 54, p. 1246, 10 1997.

[79] J. M. Hausdorff, A. Lertratanakul, M. E. Cudkowicz, et al., "Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis," *Journal of Applied Physiology*, vol. 88, pp. 2045–2053, 6 2000.

[80] A. Hirano, "Neuropathology of als: an overview." *Neurology*, vol. 47, pp. S63–6, 10 1996.

[81] A. O. E. R. I. (OERI). Impact of Disease States and Neural Damage on Motor Control-online, Accessed on. 05 oct 2022. https://socialsci.libretexts.org/Bookshelves/Psychology/Biopsychology_(OERI)_-_DRAFT_for_Review/093A_Movement/9.053A_Impact_of_Disease_States_and_Neural_Damage_on_Motor_Control.

[82] B. J. Goldfarb and S. R. Simon, "Gait patterns in patients with amyotrophic lateral sclerosis." *Archives of physical medicine and rehabilitation*, vol. 65, pp. 61–5, 2 1984.

[83] A. Assoiciation. Wo Gets ALS?-online,Accessed on. 05 oct 2022. http://webco.alsa.org/site/PageServer/?pagename=CO_1_WhoGets.html.

[84] A. Association. Types of Dementia-online, Accessed on. 05 oct 2022. https://www.alz.org/alzheimers-dementia/what-is-dementia/types-of-dementia/normal-pressure-hydrocephalus.

[85] H. Stolze, J. Kuhtz-Buschbeck, H. Drücke, et al., "Gait analysis in idiopathic normal pressure hydrocephalus – which parameters respond to the csf tap test?" *Clinical Neurophysiology*, vol. 111, pp. 1678–1686, 9 2000.

[86] H. Stolze, "Comparative analysis of the gait disorder of normal pressure hydrocephalus and parkinson's disease," *Journal of Neurology, Neurosurgery Psychiatry*, vol. 70, pp. 289–297, 3 2001.

[87] H. Association. Hydrocephalus is relentless-online, Accessed on 05 oct 2022. https://www.hydroassoc.org/.

[88] M. T. KHAN, "The scope of deep learning based architecture for employee productivity in smart offices," PhD thesis, University of Peshawar.

[89] A. Sano, S. Taylor, A. W. McHill, et al., "Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: Observational study." *Journal of medical Internet research*, vol. 20, p. e210, 2018.

[90] S. Thejaswini, K. M. R. Kumar, S. Rupali, et al., *Eeg based emotion recognition using wavelets and neural networks classifier*, 2018.

[91] Z. Yin, M. Zhao, Y. Wang, et al., "Recognition of emotions using multimodal physiological signals and an ensemble deep learning model." *Computer methods and programs in biomedicine*, vol. 140, pp. 93–110, 3 2017.

[92] P. Schmidt, A. Reiss, R. Duerichen, et al., "Wearable affect and stress recognition: A review," 11 2018.

[93] C. Maaoui and A. Pruski, "Emotion recognition through physiological signals for human-machine communication," 9 2010.

[94] J. Kim, "Bimodal emotion recognition using speech and physiological changes," 6 2007.

[95] D. Kulic and E. A. Croft, "Affective state estimation for human–robot interaction," *IEEE Transactions on Robotics*, vol. 23, pp. 991–1000, 10 2007.

[96] W. Yang, M. Rifqi, C. Marsala, et al., "Physiological-based emotion detection and recognition in a video game context," pp. 194–201, 2018.

[97] R. Mahajan, "Emotion recognition via eeg using neural network classifier," pp. 429–438, 2018.

[98] R. Jenke and A. Peer, "A cognitive architecture for modeling emotion dynamics: Intensity estimation from physiological signals," *Cognitive Systems Research*, vol. 49, pp. 128–141, 6 2018.

[99] R. Bhuyar and S. Ansari, "Design and implementation of smart office automation system," *International Journal of Computer Applications*, vol. 151, pp. 37–42, 10 2016.

[100] J. R. Williamson, T. F. Quatieri, C. J. Smalt, et al., "Using eeg to discriminate cognitive workload and performance based on neural activation and connectivity*."

[101] T. Ojha, S. Misra, and N. S. Raghuwanshi, "Wireless sensor networks for agriculture: The state-of-the-art in practice and future challenges," *Computers and Electronics in Agriculture*, vol. 118, pp. 66–84, 10 2015.

[102] C. Röcker, "Services and applications for smart office environments - a survey of state-of-the-art usage scenarios," *International journal of humanities and social sciences*, vol. 4, pp. 51–67, 2010.

[103] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," F. Pereira, C. J. Burges, L. Bottou, et al., Eds., vol. 25. Curran Associates, Inc., 2012.

[104] M. Lescroart, P. Agrawal, and J. Gallant, "Both convolutional neural networks and voxel-wise encoding models of brain activity derived from convnets represent boundary-and surface-related features," *Journal of Vision*, vol. 16, p. 756, 9 2016.

[105] L. Deng, J. Li, J.-T. Huang, et al., "Recent advances in deep learning for speech research at microsoft." IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 5 2013.

[106] F. Clement, C. Couprie, L. Najman, et al., "Learning hierarchical features for scene labeling."

[107] A. Ullah, "A knowledge extraction framework for smart health using deep learning," PhD thesis, University of Engineering and Technology Taxila, Pakistan, 2020.

[108] J. Kim and E. André, "Emotion recognition based on physiological changes in music listening." *IEEE transactions on pattern analysis and machine intelligence*, vol. 30, pp. 2067–83, 12 2008.

[109] S. Alhagry, A. Aly, and R. A., "Emotion recognition based on eeg using lstm recurrent neural network," *International Journal of Advanced Computer Science and Applications*, vol. 8, 2017.

[110] C. L. Lisetti and F. Nasoz, "Using noninvasive wearable computers to recognize human emotions from physiological signals," *EURASIP Journal on Advances in Signal Processing*, vol. 2004, p. 929414, 12 2004.

[111] J. Sorinas, M. D. Grima, J. M. Ferrandez, et al., "Identifying suitable brain regions and trial size segmentation for positive/negative emotion recognition," *International Journal of Neural Systems*, vol. 29, p. 1850044, 3 2019.

[112] H. Huang, Q. Xie, J. Pan, et al., "An eeg-based brain computer interface for emotion recognition and its application in patients with disorder of consciousness," *IEEE Transactions on Affective Computing*, vol. 12, pp. 832–842, 2021.

[113] X. Li, R. La, Y. Wang, et al., "Eeg-based mild depression recognition using convolutional neural network." *Medical biological engineering computing*, vol. 57, pp. 1341–1352, 6 2019.

[114] A. Wachter, "Generation of artificial image and video data for medical deep learning applications," PhD thesis, Karlsruher Institut für Technologie (KIT), 2022.

[115] J. Cai, W. Chen, and Z. Yin, "Multiple transferable recursive feature elimination technique for emotion recognition based on eeg signals," *Symmetry*, vol. 11, p. 683, 5 2019.

[116] A. de Santos Sierra, C. S. Ávila, J. G. Casanova, et al., "A stress-detection system based on physiological signals and fuzzy logic," *IEEE Transactions on Industrial Electronics*, vol. 58, pp. 4857–4865, 2011.

[117] P. Li, H. Liu, Y. Si, et al., "Eeg based emotion recognition by combining functional connectivity network and local activations," *IEEE Transactions on Biomedical Engineering*, vol. 66, pp. 2869–2881, 10 2019.

[118] N. K. Al-Qazzaz, M. K. Sabir, and K. Grammer, "Correlation indices of electroencephalogram-based relative powers during human emotion processing." ACM Press, 2019, pp. 64–70.

[119] N. Maisonneuve, M. L. L. Stevens, M. Niessen, et al., "Citizen noise pollution monitoring atgentive-attentive agents for collaborative learners view project agt international; lionhead view project citizen noise pollution monitoring," 2009.

[120] S. G. Santur and Y. Santur, "Knowledge mining approach for healthy monitoring from pregnancy data with big volumes," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 4, pp. 141–145, 12 2016.

[121] Y. Santur, M. Karaköse, and E. Akin, "Random forest based diagnosis approach for rail fault inspection in railways," 12 2016, pp. 745–750.

[122] Y. Santur, M. Karakose, I. Aydin, et al., "Imu based adaptive blur removal approach using image processing for railway inspection." IEEE, 5 2016, pp. 1–4.

[123] R. Chalasani and J. C. Principe, "Deep predictive coding networks," 1 2013.

[124] D. C. Cireşan, U. Meier, J. Masci, et al., "Flexible, high performance convolutional neural networks for image classification." AAAI Press, 2011, pp. 1237–1242.

[125] X. Hao and G. Zhang, "Deep learning," *Encyclopedia with Semantic Computing and Robotic Intelligence*, vol. 01, p. 1630018, 3 2017.

[126] D. S. W. Ting, L. R. Pasquale, L. Peng, et al., "Artificial intelligence and deep learning in ophthalmology." *The British journal of ophthalmology*, vol. 103, pp. 167–175, 2019.

[127] L. Deng and D. Yu, "Deep convex net: a scalable architecture for speech pattern classification." ISCA, 8 2011, pp. 2285–2288.

[128] L. Deng, X. He, and J. Gao, "Deep stacking networks for information retrieval." IEEE, 5 2013, pp. 3153–3157.

[129] J. Li, H. Chang, and J. Yang, "Sparse deep stacking network for image classification," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 29, 3 2015.

[130] W. Liu, Z. Wang, X. Liu, et al., "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11–26, 4 2017.

[131] A. Sohail, "Breast cancer analysis using histopathological images and deep convolutional neural networks," PhD thesis, Pakistan Institute of Engineering and Applied Sciences Nilore, Islamabad, Pakistan, 2021.

[132] J. Günther, P. M. Pilarski, G. Helfrich, et al., "First steps towards an intelligent laser welding architecture using deep neural networks and reinforcement learning," *Procedia Technology*, vol. 15, pp. 474–483, 2014.

[133] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 5 2015.

[134] N. Akhtar and U. Ragavendran, "Interpretation of intelligence in cnn-pooling processes: a methodological survey," *Neural Computing and Applications*, vol. 32, pp. 879–898, 2 2020.

[135] C.-Y. Lee, P. W. Gallagher, and Z. Tu, "Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree," A. Gretton and C. C. Robert, Eds., vol. 51. PMLR, 10 2016, pp. 464–472.

[136] T. Yoshioka, N. Ito, M. Delcroix, et al., "The ntt chime-3 system: Advances in speech enhancement and recognition for mobile multi-microphone devices." IEEE, 12 2015, pp. 436–443.

[137] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review." *Neural computation*, vol. 29, pp. 2352–2449, 2017.

[138] J. Gu, Z. Wang, J. Kuen, et al., "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, 5 2018.

[139] J.-L. Zhou, H.-W. Sun, D.-H. Yin, et al., "Deprotonation or protonation: The coordination properties, crystal structures and spectra of cobalt (ii) complex with 1-(2-pyridylazo)-2-acenaphthequinol ligand," *Journal of Molecular Structure*, vol. 1134, pp. 63–66, 4 2017.

[140] G. E. Hinton, N. Srivastava, A. Krizhevsky, et al., "Improving neural networks by preventing co-adaptation of feature detectors," 7 2012.

[141] A. Khan, A. Sohail, U. Zahoora, et al., "A survey of the recent architectures of deep convolutional neural networks," *Artificial Intelligence Review*, vol. 53, pp. 5455–5516, 12 2020.

[142] C. Szegedy, W. Liu, Y. Jia, et al., "Going deeper with convolutions," 9 2014.

[143] C. Szegedy, V. Vanhoucke, S. Ioffe, et al., "Rethinking the inception architecture for computer vision." IEEE, 6 2016, pp. 2818–2826.

[144] K. He, X. Zhang, S. Ren, et al., "Deep residual learning for image recognition," 12 2015.

[145] S. Zagoruyko and N. Komodakis, "Wide residual networks." British Machine Vision Association, 2016, pp. 87.1–87.12.

[146] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Training very deep networks," C. Cortes, N. Lawrence, D. Lee, et al., Eds., vol. 28. Curran Associates, Inc., 2015.

[147] S. Xie, R. Girshick, P. Dollar, et al., "Aggregated residual transformations for deep neural networks." IEEE, 7 2017, pp. 5987–5995.

[148] G. Huang, Z. Liu, L. van der Maaten, et al., "Densely connected convolutional networks," 8 2016.

[149] S. Hochreiter, "Untersuchungen zu dynamischen neuronalen netzen," PhD thesis, Technische Universität Munich, 1991.

[150] LSTM_Cell.svg-online Accessed: 5 oct 2022, licensed under the creative commons attribution-share alike 4.0 international license. http://en.wikipedia.org/wiki/Long_short_term_memory.

[151] L. Shao, F. Zhu, and X. Li, "Transfer learning for visual categorization: A survey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, pp. 1019–1034, 5 2015.

[152] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, p. 9, 12 2016.

[153] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, pp. 1345–1359, 10 2010.

[154] J. Lu, V. Behbood, P. Hao, et al., "Transfer learning using computational intelligence: A survey," *Knowledge-Based Systems*, vol. 80, pp. 14–23, 5 2015.

[155] H.-C. Shin, H. R. Roth, M. Gao, et al., "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning." *IEEE transactions on medical imaging*, vol. 35, pp. 1285–98, 2016.

[156] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 623–656, 10 1948.

[157] T. D. Schneider and T. D. Schneider, "Information theory primer with an appendix on logarithms pdf version information theory based fur binding sites modeling view project molecular information theory view project information theory primer with an appendix on logarithms," 2013.

[158] Z. Ali and S. Bhaskar, "Basic statistical tools in research and data analysis," *Indian Journal of Anaesthesia*, vol. 60, p. 662, 2016.

[159] A. Luque, A. Carrasco, A. Martín, et al., "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognition*, vol. 91, pp. 216–231, 7 2019.

[160] A. P. King and R. J. Eckersley, "Inferential statistics ii: Parametric hypothesis testing," pp. 91–117, 2019.

[161] R. Baker, "The history of gait analysis before the advent of modern computers," *Gait and Posture*, vol. 26, pp. 331–342, 9 2007.

[162] G. A. Borelli, *On the movement of animals*. Springer Berlin Heidelberg, 1989.

[163] J. Stenum, C. Rossi, and R. T. Roemmich, "Two-dimensional video-based analysis of human gait using pose estimation," *PLoS Computational Biology*, vol. 17, 4 2021.

[164] A. Esteva, B. Kuprel, R. A. Novoa, et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–118, 2 2017.

[165] G. Haskins, U. Kruger, and P. Yan, "Deep learning in medical image registration: a survey," *Machine Vision and Applications*, vol. 31, p. 8, 2 2020.

[166] F. Alam and S. U. Rahman, "Challenges and solutions in multimodal medical image subregion detection and registration," *Journal of Medical Imaging and Radiation Sciences*, vol. 50, pp. 24–30, 3 2019.

[167] M. H. Hesamian, W. Jia, X. He, et al., "Deep learning techniques for medical image segmentation: Achievements and challenges," *Journal of Digital Imaging*, vol. 32, pp. 582–596, 8 2019.

[168] H. Shin, "Deep convolutional neural network-based hemiplegic gait detection using an inertial sensor located freely in a pocket," *Sensors (Basel, Switzerland)*, vol. 22, 3 2022.

[169] G. Litjens, T. Kooi, B. E. Bejnordi, et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 12 2017.

[170] A. Saboor, T. Kask, A. Kuusik, et al., "Latest research trends in gait analysis using wearable sensors and machine learning: A systematic review," *IEEE Access*, vol. 8, pp. 167 830–167 864, 2020.

[171] A. M. de-la Herran, B. Garcia-Zapirain, and A. Mendez-Zorrilla, "Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications," *Sensors*, vol. 14, pp. 3362–3394, 2 2014.

[172] M. O. Derawi, P. Bours, and K. Holien, "Improved cycle detection for accelerometer based gait authentication," 2010.

[173] M. Yang, H. Zheng, H. Wang, et al., "igait: an interactive accelerometer based gait analysis system." *Computer methods and programs in biomedicine*, vol. 108, pp. 715–23, 11 2012.

[174] A. Bourke and G. Lyons, "A threshold-based fall-detection algorithm using a bi-axial gyroscope sensor," *Medical Engineering Physics*, vol. 30, pp. 84–90, 1 2008.

[175] S. R. Hundza, W. R. Hook, C. R. Harris, et al., "Accurate and reliable gait cycle detection in parkinson's disease." *IEEE transactions on neural systems and rehabilitation engineering : a*

publication of the IEEE Engineering in Medicine and Biology Society, vol. 22, pp. 127–37, 1 2014.

[176] Y. Wahab, A. Zayegh, R. K. Begg, et al., "Design of mems biomedical pressure sensor for gait analysis." IEEE, 11 2008, pp. 166–169.

[177] M. Engin, A. Demirel, E. Z. Engin, et al., "Recent developments and trends in biomedical sensors," *Measurement*, vol. 37, pp. 173–188, 3 2005.

[178] J. Wu and B. Wu, "The novel quantitative technique for assessment of gait symmetry using advanced statistical learning algorithm," *BioMed Research International*, vol. 2015, pp. 1–7, 2015.

[179] Y. Chen and Y. Xue, "A deep learning approach to human activity recognition based on single accelerometer." IEEE, 10 2015, pp. 1488–1492.

[180] T. Zebin, P. J. Scully, and K. B. Ozanyan, "Human activity recognition with inertial sensors using a deep learning approach." IEEE, 10 2016, pp. 1–3.

[181] F. Ordóñez and D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, p. 115, 1 2016.

[182] J. Camps, A. Samà, M. Martín, et al., "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, 1 2018.

[183] R. S. McGinnis, N. Mahadevan, Y. Moon, et al., "A machine learning approach for gait speed estimation using skin-mounted wearable sensors: From healthy controls to individuals with multiple sclerosis," *PLOS ONE*, vol. 12, p. e0178366, 6 2017.

[184] L. Steffan, L. Kaul, and T. Asfour, "Online stability estimation based on inertial sensor data for human and humanoid fall prevention."

[185] K. Hu, Z. Wang, W. Wang, et al., "Graph sequence recurrent neural network for vision-based freezing of gait detection," *IEEE Transactions on Image Processing*, vol. 29, pp. 1890–1901, 2020.

[186] D. Buongiorno, I. Bortone, G. D. Cascarano, et al., "A low-cost vision system based on the analysis of motor features for recognition and severity rating of parkinson's disease," *BMC Medical Informatics and Decision Making*, vol. 19, 2019.

[187] P. Kumari, N. J. Cooney, T. seong Kim, et al., "Gait analysis in spastic hemiplegia and diplegia cerebral palsy using a wearable activity tracking device - a data quality analysis for deep convolutional neural networks." IEEE, 12 2018, pp. 1–4.

[188] J. Juen, Q. Cheng, V. Prieto-Centurion, et al., "Health monitors for chronic disease by gait analysis with mobile phones," *Telemedicine and e-Health*, vol. 20, pp. 1035–1041, 11 2014.

[189] A. M. Saleh and T. Hamoud, "Analysis and best parameters selection for person recognition based on gait model using cnn algorithm and image augmentation," *Journal of Big Data*, vol. 8, 12 2021.

[190] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," Z. Ghahramani, M. Welling, C. Cortes, et al., Eds., vol. 27. Curran Associates, Inc., 2014.

[191] J. Donahue, L. A. Hendricks, M. Rohrbach, et al., "Long-term recurrent convolutional networks for visual recognition and description," 11 2014.

[192] D. Tran, L. Bourdev, R. Fergus, et al., "Learning spatiotemporal features with 3d convolutional networks." IEEE, 12 2015, pp. 4489–4497.

[193] T. Bonci, F. Salis, K. Scott, et al., "An algorithm for accurate marker-based gait event detection in healthy and pathological populations during complex motor tasks," *Frontiers in Bioengineering and Biotechnology*, vol. 10, 6 2022.

[194] R. K. M. de Souza, S. F. B. da Rocha, R. T. Martins, et al., "Gait in normal pressure hydrocephalus: characteristics and effects of the csf tap test," *Arquivos de Neuro-Psiquiatria*, vol. 76, pp. 324–331, 5 2018.

[195] S. Zheng, J. Zhang, K. Huang, et al., "Robust view transformation model for gait recognition."

[196] S. Yu, D. Tan, and T. Tan, "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition." IEEE, 2006.

[197] M. Fayyaz, "Appearance based pedestrian analysis using machine learning," PhD thesis, COMSATS University Islamabad Wah Campus, Pakistan, 2020.

[198] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," Z. Ghahramani, M. Welling, C. Cortes, et al., Eds., vol. 27. Curran Associates, Inc., 2014.

[199] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, pp. 1735–1780, 11 1997.

[200] S.-M. C. Witold Pedrycz, *Studies in computational intelligence-deep learning - algorithms and applications*, 11th ed. Springer International Publishing, 2020.

[201] A. Agga, A. Abbou, M. Labbadi, et al., "Cnn-lstm: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production," *Electric Power Systems Research*, vol. 208, p. 107908, 7 2022.

[202] Y. Wang, Q. Wu, N. Dey, et al., "Deep back propagation–long short-term memory network based upper-limb semg signal classification for automated rehabilitation," *Biocybernetics and Biomedical Engineering*, vol. 40, pp. 987–1001, 7 2020.

[203] C. Feichtenhofer, A. Pinz, and R. P. Wildes, "Spatiotemporal residual networks for video action recognition," 11 2016.

[204] S. Ji, W. Xu, M. Yang, et al., "3d convolutional neural networks for human action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 221–231, 1 2013.

[205] X. Wang, A. Farhadi, and A. Gupta, "Actions transformations." IEEE, 6 2016, pp. 2658–2667.

[206] Łukasz Kidziński, B. Yang, J. L. Hicks, et al., "Deep neural networks enable quantitative movement analysis using single-camera videos." *Nature communications*, vol. 11, p. 4054, 2020.

[207] K. Sato, Y. Nagashima, T. Mano, et al., "Quantifying normal and parkinsonian gait features from home movies: Practical application of a deep learning-based 2d pose estimator." *PloS one*, vol. 14, p. e0223549, 2019.

[208] C. Chambers, G. Kong, K. Wei, et al., "Pose estimates from online videos show that side-by-side walkers synchronize movement under naturalistic conditions." *PloS one*, vol. 14, p. e0217861, 2019.

[209] N. Nakano, T. Sakura, K. Ueda, et al., "Evaluation of 3d markerless motion capture accuracy using openpose with multiple video cameras." *Frontiers in sports and active living*, vol. 2, p. 50, 2020.

[210] M. Ota, H. Tateuchi, T. Hashiguchi, et al., "Verification of reliability and validity of motion analysis systems during bilateral squat using human pose tracking algorithm." *Gait posture*, vol. 80, pp. 62–67, 2020.

[211] M. Zago, M. Luzzago, T. Marangoni, et al., "3d tracking of human motion using visual skeletonization and stereoscopic vision." *Frontiers in bioengineering and biotechnology*, vol. 8, p. 181, 2020.

[212] B. Sijobert, M. Benoussaad, J. Denys, et al., "Implementation and validation of a stride length estimation algorithm, using a single basic inertial sensor on healthy subjects and patients suffering from parkinson's disease," *Health*, vol. 07, pp. 704–714, 2015.

[213] S. R. Simon, "Quantification of human motion: gait analysis—benefits and limitations to its application to clinical problems," *Journal of Biomechanics*, vol. 37, pp. 1869–1880, 12 2004.

[214] S. D. Din, A. Godfrey, and L. Rochester, "Validation of an accelerometer to quantify a comprehensive battery of gait characteristics in healthy older adults and parkinson's disease: Toward clinical and at home use." *IEEE journal of biomedical and health informatics*, vol. 20, pp. 838–847, 2016.

[215] F. Horak, L. King, and M. Mancini, "Role of body-worn movement monitor technology for balance and gait rehabilitation," *Physical Therapy*, vol. 95, pp. 461–470, 3 2015.

[216] A. König, L. Klaming, M. Pijl, et al., "Objective measurement of gait parameters in healthy and cognitively impaired elderly using the dual-task paradigm," *Aging Clinical and Experimental Research*, vol. 29, pp. 1181–1189, 12 2017.

[217] U. D. Croce, A. Cereatti, and M. Mancini, *Gait parameters estimated using inertial measurement units*. Springer International Publishing, 2017.

[218] A. Mirelman, T. Herman, M. Brozgol, et al., "Executive function and falls in older adults: New findings from a five-year prospective study link fall risk to cognition," *PLoS ONE*, vol. 7, p. e40297, 6 2012.

[219] S. Studenski, "Gait speed and survival in older adults," *JAMA*, vol. 305, p. 50, 1 2011.

[220] J. M. Hausdorff, D. A. Rios, and H. K. Edelberg, "Gait variability and fall risk in community-living older adults: A 1-year prospective study," *Archives of Physical Medicine and Rehabilitation*, vol. 82, pp. 1050–1056, 8 2001.

[221] J. G. Nutt, "Classification of gait and balance disorders." *Advances in neurology*, vol. 87, pp. 135–41, 2001.

[222] A. Mirelman, T. Gurevich, N. Giladi, et al., "Gait alterations in healthy carriers of the lrrk2 g2019s mutation," *Annals of Neurology*, vol. 69, pp. 193–197, 1 2011.

[223] D. Kobsar, J. M. Charlton, C. T. Tse, et al., "Validity and reliability of wearable inertial sensors in healthy adult walking: a systematic review and meta-analysis," *Journal of NeuroEngineering and Rehabilitation*, vol. 17, p. 62, 12 2020.

[224] D. Trojaniello, A. Cereatti, E. Pelosin, et al., "Estimation of step-by-step spatio-temporal parameters of normal and impaired gait using shank-mounted magneto-inertial sensors: application to elderly, hemiparetic, parkinsonian and choreic gait," *Journal of NeuroEngineering and Rehabilitation*, vol. 11, p. 152, 2014.

[225] F. A. Storm, C. J. Buckley, and C. Mazzà, "Gait event detection in laboratory and real life settings: Accuracy of ankle and waist sensor based methods," *Gait Posture*, vol. 50, pp. 42–46, 2016.

[226] D. Maquet, F. Lekeu, E. Warzee, et al., "Gait analysis in elderly adult patients with mild cognitive impairment and patients with mild alzheimerâ128;153;s disease: simple versus dual task: a preliminary report," *Clinical Physiology and Functional Imaging*, vol. 30, pp. 51–56, 1 2010.

[227] N. Herssens, T. van Criekinge, W. Saeys, et al., "An investigation of the spatio-temporal parameters of gait and margins of stability throughout adulthood," *Journal of The Royal Society Interface*, vol. 17, p. 20200194, 5 2020.

[228] S. Patel, H. Park, P. Bonato, et al., "A review of wearable sensors and systems with application in rehabilitation," *Journal of NeuroEngineering and Rehabilitation*, vol. 9, p. 21, 12 2012.

[229] M. Montero-Odasso, A. Oteng-Amoako, M. Speechley, et al., "The motor signature of mild cognitive impairment: results from the gait and brain study." *The journals of gerontology. Series A, Biological sciences and medical sciences*, vol. 69, pp. 1415–21, 11 2014.

[230] G. Allali, R. W. Kressig, F. Assal, et al., "Changes in gait while backward counting in demented older adults with frontal lobe dysfunction," *Gait Posture*, vol. 26, pp. 572–576, 10 2007.

[231] J. Verghese, M. Robbins, R. Holtzer, et al., "Gait dysfunction in mild cognitive impairment syndromes." *Journal of the American Geriatrics Society*, vol. 56, pp. 1244–51, 7 2008.

[232] O. Beauchet, G. Allali, C. Launay, et al., "Gait variability at fast-pace walking speed: a biomarker of mild cognitive impairment?" *The journal of nutrition, health aging*, vol. 17, pp. 235–9, 3 2013.

[233] O. Beauchet, E. Freiberger, C. Annweiler, et al., "Test-retest reliability of stride time variability while dual tasking in healthy and demented adults with frontotemporal degeneration." *Journal of neuroengineering and rehabilitation*, vol. 8, p. 37, 7 2011.

[234] C. Annweiler, O. Beauchet, R. Bartha, et al., "Motor cortex and gait in mild cognitive impairment: a magnetic resonance spectroscopy and volumetric imaging study." *Brain : a journal of neurology*, vol. 136, pp. 859–71, 3 2013.

[235] H. H. Dodge, N. C. Mattek, D. Austin, et al., "In-home walking speeds and variability trajectories associated with mild cognitive impairment." *Neurology*, vol. 78, pp. 1946–52, 6 2012.

[236] S.-R. Ke, H. Thuc, Y.-J. Lee, et al., "A review on video-based human activity recognition," *Computers*, vol. 2, pp. 88–131, 6 2013.

[237] S. Shirke, S.S.Pawar, and K. Shah, "Literature review: Model free human gait recognition." IEEE, 4 2014, pp. 891–895.

[238] N. Neverova, C. Wolf, G. Lacey, et al., "Learning human identity from motion patterns," *IEEE Access*, vol. 4, pp. 1810–1820, 2016.

[239] C. Prakash, R. Kumar, and N. Mittal, "Recent developments in human gait research: parameters, approaches, applications, machine learning techniques, datasets and challenges," *Artificial Intelligence Review*, vol. 49, pp. 1–40, 1 2018.

[240] C. Prakash, R. Kumar, N. Mittal, et al., "Vision based identification of joint coordinates for marker-less gait analysis," vol. 132. Elsevier B.V., 2018, pp. 68–75.

[241] N. Kumar, N. Kunju, A. Kumar, et al., "Active marker based kinematic and spatio-temporal gait measurement system using labview vision," pp. 600–605, 2010.

[242] M. Bilal, M. Maqsood, S. Yasmin, et al., "A transfer learning-based efficient spatiotemporal human action recognition framework for long and overlapping action classes," *The Journal of Supercomputing*, vol. 78, pp. 2873–2908, 2 2022.

[243] S. Tyson and L. Connell, "The psychometric properties and clinical utility of measures of walking and mobility in neurological conditions: a systematic review," *Clinical Rehabilitation*, vol. 23, pp. 1018–1033, 11 2009.

[244] H. Fujiyoshi, T. Hirakawa, and T. Yamashita, "Deep learning-based image recognition for autonomous driving," *IATSS Research*, vol. 43, pp. 244–252, 12 2019.

[245] J.-H. Choi and J.-S. Lee, "Embracenet: A robust deep learning architecture for multimodal classification," *Information Fusion*, vol. 51, pp. 259–270, 11 2019.

[246] F. I. Diakogiannis, F. Waldner, P. Caccetta, et al., "Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 162, pp. 94–114, 4 2020.

[247] L. Liu, W. Ouyang, X. Wang, et al., "Deep learning for generic object detection: A survey," *International Journal of Computer Vision*, vol. 128, pp. 261–318, 2 2020.

[248] O. Russakovsky, J. Deng, H. Su, et al., "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, pp. 211–252, 12 2015.

[249] C. Szegedy, S. Ioffe, V. Vanhoucke, et al., "Inception-v4, inception-resnet and the impact of residual connections on learning," 2 2016.

[250] G. Huang, Z. Liu, L. V. D. Maaten, et al., "Densely connected convolutional networks." IEEE, 7 2017, pp. 2261–2269.

[251] K. Homan, K. Yamamoto, K. Kadoya, et al., "Comprehensive validation of a wearable foot sensor system for estimating spatiotemporal gait parameters by simultaneous three-dimensional optical motion analysis," *BMC Sports Science, Medicine and Rehabilitation*, vol. 14, 12 2022.

[252] L. I. Lin, "A concordance correlation coefficient to evaluate reproducibility." *Biometrics*, vol. 45, pp. 255–68, 3 1989.

[253] D. V. Cicchetti, "Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology." *Psychological Assessment*, vol. 6, pp. 284–290, 12 1994.

[254] B. Galna, G. Barry, D. Jackson, et al., "Accuracy of the microsoft kinect sensor for measuring movement in people with parkinson's disease." *Gait posture*, vol. 39, pp. 1062–8, 4 2014.

[255] M. P. Kadaba, H. K. Ramakrishnan, M. E. Wootten, et al., "Repeatability of kinematic, kinetic, and electromyographic data in normal adult gait," *Journal of Orthopaedic Research*, vol. 7, pp. 849–860, 11 1989.

[256] L. C. Benson, C. A. Clermont, E. Bošnjak, et al., "The use of wearable devices for walking and running gait analysis outside of the lab: A systematic review," *Gait Posture*, vol. 63, pp. 124–138, 6 2018.

[257] J. M. Hausdorff, P. L. Purdon, C. K. Peng, et al., "Fractal dynamics of human gait: stability of long-range correlations in stride interval fluctuations." *Journal of applied physiology (Bethesda, Md. : 1985)*, vol. 80, pp. 1448–57, 5 1996.

[258] J. M. Hausdorff, C. K. Peng, Z. Ladin, et al., "Is walking a random walk? evidence for long-range correlations in stride interval of human gait." *Journal of applied physiology (Bethesda, Md. : 1985)*, vol. 78, pp. 349–58, 1 1995.

[259] R. A. de, N. Stergiou, L. Decker, et al., "Complexity and human gait," pp. 2–12, 2010.

[260] S. M. Pincus, "Approximate entropy as a measure of system complexity." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88, pp. 2297–301, 3 1991.

[261] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy." *American journal of physiology. Heart and circulatory physiology*, vol. 278, pp. H2039–49, 6 2000.

[262] Z. M. Nikolic and D. B. Popovic, "Predicting quadriceps muscle activity during gait with an automatic rule determination method." *IEEE transactions on bio-medical engineering*, vol. 45, pp. 1081–5, 8 1998.

[263] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series." *Physical review letters*, vol. 89, p. 068102, 8 2002.

[264] M. J. Kurz and N. Stergiou, "The aging humans neuromuscular system expresses less certainty for selecting joint kinematics during gait," *Neuroscience Letters*, vol. 348, pp. 155–158, 9 2003.

[265] W. Aziz and M. Arif, "Complexity analysis of stride interval time series by threshold dependent symbolic entropy." *European journal of applied physiology*, vol. 98, pp. 30–40, 9 2006.

[266] A. H. Khandoker, M. Palaniswami, and R. K. Begg, "A comparative study on approximate entropy measure and poincaré plot indexes of minimum foot clearance variability in the elderly during walking." *Journal of neuroengineering and rehabilitation*, vol. 5, p. 4, 2 2008.

[267] J. E. Deffeyes, R. T. Harbourne, A. Kyvelidou, et al., "Nonlinear analysis of sitting postural sway indicates developmental delay in infants." *Clinical biomechanics (Bristol, Avon)*, vol. 24, pp. 564–70, 8 2009.

[268] S. J. Mcgregor, M. A. Busa, R. Parshad, et al., "Control entropy of gait: Does running fitness affect complexity of walking?" 2011.

[269] Y. Tochigi, N. A. Segal, T. Vaseenon, et al., "Entropy analysis of tri-axial leg acceleration signal waveforms for measurement of decrease of physiological variability in human gait." *Journal of orthopaedic research : official publication of the Orthopaedic Research Society*, vol. 30, pp. 897–904, 6 2012.

[270] B. D. La, C. Torres, M. D. Sánchez, et al., "Effects of treadmill on gait variability in healthy young subjects influencia del tapiz rodante sobre la variabilidad de la marcha en sujetos jóvenes y sanos," pp. 53–62, 2012.

[271] A. Q. Abbasi and W. A. Loun, "Symbolic time series analysis of temporal gait dynamics," *Journal of Signal Processing Systems*, vol. 74, pp. 417–422, 3 2014.

[272] A. Qumar, W. Aziz, S. Saeed, et al., "Comparative study of multiscale entropy analysis and symbolic time series analysis when applied to human gait dynamics." IEEE, 12 2013, pp. 126–132.

[273] H. Azami, A. Fernández, and J. Escudero, "Refined multiscale fuzzy entropy based on standard deviation for biomedical signal analysis," *Medical Biological Engineering Computing*, vol. 55, pp. 2037–2052, 11 2017.

[274] H. Azami, A. Fernández, and J. Escudero, "Multivariate multiscale dispersion entropy of biomedical times series," *Entropy*, vol. 21, p. 913, 9 2019.

[275] H. Azami, S. E. Arnold, S. Sanei, et al., "Multiscale fluctuation-based dispersion entropy and its applications to neurological diseases," *IEEE Access*, vol. 7, pp. 68 718–68 733, 2019.

[276] C. K. Karmakar, A. H. Khandoker, R. K. Begg, et al., "Understanding ageing effects by approximate entropy analysis of gait variability." IEEE, 8 2007, pp. 1965–1968.

[277] U. R. Acharya, S. V. Sree, C. M. Lim, et al., "Comparison of walking parameters obtained from the young, elderly and adults with support." *Computer methods in biomechanics and biomedical engineering*, vol. 16, pp. 1202–12, 2013.

[278] J. M. Yentes, N. Hunt, K. K. Schmid, et al., "The appropriate use of approximate entropy and sample entropy with short data sets." *Annals of biomedical engineering*, vol. 41, pp. 349–65, 2 2013.

[279] P. C. Raffalt, J. McCamley, W. Denton, et al., "Sampling frequency influences sample entropy of kinematics during walking." *Medical biological engineering computing*, vol. 57, pp. 759–764, 4 2019.

[280] J. D. McCamley, W. Denton, A. Arnold, et al., "On the calculation of sample entropy using continuous and discrete human gait data." *Entropy (Basel, Switzerland)*, vol. 20, 10 2018.

[281] J. M. Yentes and P. C. Raffalt, "Entropy analysis in gait research: Methodological considerations and recommendations," *Annals of Biomedical Engineering*, vol. 49, pp. 979–990, 3 2021.

[282] A. L. Goldberger, L. A. N. Amaral, L. Glass, et al., "Physiobank, physiotoolkit, and physionet," *Circulation*, vol. 101, 6 2000.

[283] J. M. Hausdorff, A. Lertratanakul, M. E. Cudkowicz, et al., "Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis." *Journal of applied physiology (Bethesda, Md. : 1985)*, vol. 88, pp. 2045–53, 6 2000.

[284] W. Chen, J. Zhuang, W. Yu, et al., "Measuring complexity using fuzzyen, apen, and sampen," *Medical Engineering Physics*, vol. 31, pp. 61–68, 1 2009.

[285] J. Kurths, A. Voss, P. Saparin, et al., "Quantitative analysis of heart rate variability." *Chaos (Woodbury, N.Y.)*, vol. 5, pp. 88–94, 3 1995.

[286] P. Grassberger, "Finite sample corrections to entropy and dimension estimates," *Physics Letters A*, vol. 128, pp. 369–373, 4 1988.

[287] M. C. Eguia, M. I. Rabinovich, and H. D. I. Abarbanel, "Information transmission and recovery in neural communications channels," *Physical Review E*, vol. 62, pp. 7111–7122, 11 2000.

[288] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy to distinguish physiologic and synthetic rr time series." *Computers in cardiology*, vol. 29, pp. 137–40, 2002.

[289] M. Costa, C.-K. Peng, A. L. Goldberger, et al., "Multiscale entropy analysis of human gait dynamics," *Physica A: Statistical Mechanics and its Applications*, vol. 330, pp. 53–60, 12 2003.

[290] A.-B. Liu and C.-W. Lin, "Multiscale approximate entropy for gait analysis in patients with neurodegenerative diseases," *Entropy*, vol. 21, p. 934, 9 2019.

[291] Y. Xia, Q. Gao, Y. Lu, et al., "A novel approach for analysis of altered gait variability in amyotrophic lateral sclerosis," *Medical Biological Engineering Computing*, vol. 54, pp. 1399–1408, 9 2016.

[292] L. A. Lipsitz, "Age-related changes in the "complexity" of cardiovascular dynamics: A potential marker of vulnerability to disease," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 5, pp. 102–109, 3 1995.