The Pennsylvania State University
The J.Jeffrey and Ann Marie Fox Graduate School

DEVELOPMENT OF AN ARTIFICIAL NEURAL NETWORK (ANN) MODEL FOR GAIT-BASED CLASSIFICATION OF NEURODEGENERATIVE DISORDERS (NEUROGAITNET)

A Thesis in

Mechanical Engineering

by

Mohamed A. Mohamed

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

December 2024

The thesis of Mohamed A. Mohamed was reviewed and approved by the following:

Richard C. Ciocci
Associate Professor of Mechanical Engineering
Thesis Advisor

Aldo W. Morales
Associate Professor of Mechanical Engineering

Anilchandra Attaluri
Assistant Professor of Mechanical Engineering

Brian Allen Maicke
Associate Professor of Mechanical Engineering
Professor-in-Charge

Abstract

Early diagnosis of neurodegenerative diseases like Huntington's disease (HD), Parkinson's disease, and amyotrophic lateral sclerosis (ALS) is a tough challenge primarily because motor symptoms overlap and are often symptomatic of more than one condition; also, traditional diagnostic methods are costly and demanding. There has been relatively little exploration of gait analysis in Huntington's disease compared to the well-established use in Parkinson's and ALS. This paper aims to fill this gap by proposing a neural network model named NeuroGaitNet based on deep learning approaches, which are used for the classification of gait disorders using gait data collected from wearable sensors.

The severity of gait alterations increases with neurodegeneration, but nuanced differences demand sophisticated analysis. NeuroGaitNet can analyze gait's time domain and frequency data and generate stride intervals, footfall contact times, and power spectral density (PSD).

The model was trained on ALS, Parkinson's, Huntington's data [14], and healthy controls, and it had an average test and validation accuracy of 95.5% and validation accuracy of 91.97%. Few cases were included, so dataset imbalance was considered, and data augmentation techniques (Gaussian noise injection) were used. This work bridges a critical research gap by presenting a scalable and automatic tool for early classification of neurodegenerative diseases targeting Huntington's Disease (HD).

Moreover, NeuroGaitNet has a user-friendly graphical interface for processing gait data in real-time. It provides instant classifications of gait disorders that are functional in research and clinical settings by integrating deep learning with sophisticated gait analytics. This study pushes the boundaries in diagnostic technology for neurodegeneration, including Huntington's disease, bringing a real opportunity to transform pre-symptomatic detection and clinical outcomes.

Keywords: Neural Network, Gait Analysis, Neurodegenerative Disorders, Power Spectral Density (PSD), Data Augmentation, Deep Learning, Huntington's Disease, Clinical Diagnostics, Artificial Intelligence.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Acknowledgements

I would like to express my heartfelt gratitude to my parents and my brothers, whose unwavering support and encouragement have been the foundation of my journey.

To my father, your wisdom and encouragement have always motivated me to pursue my dreams. Your dedication and integrity are qualities I have always admired and dreamed of embodying myself. Your faith in my abilities has kept me focused and driven. I cannot thank you enough for your unwavering belief in my potential.

To my mother, your sacrifices and guidance have shaped my values and work ethic, and for that, I am forever grateful. Even from miles away, your encouraging words and guidance have fueled my determination and resilience. I am truly thankful for your sacrifices and the values you have instilled in me.

To my fiancée, Dr. Mawada Osman, your love and support have been my constant strength throughout this journey. Your encouragement, patience, and belief in me have meant the world, and I am forever grateful. This achievement is as much yours as it is mine. Thank you for always being by my side.

I would also like to express my sincere gratitude to my thesis advisor Dr. Richard C. Ciocci and thesis committee Dr. Aldo W. Morales and Dr. Anilchandra Attaluri. Your insights, feedback, and expertise have been invaluable throughout this process. Thank you for your guidance and for challenging me to think critically and strive for excellence. Your support has made this journey not only possible but enriching.

Chapter 1

INTRODUCTION

Neurodegenerative diseases such as Huntington's disease, Parkinson's disease, and amyotrophic lateral sclerosis (ALS) pose significant challenges for early diagnosis due to the overlap of motor symptoms across different disorders Early diagnosis is key to better patient outcomes, yet most conventional diagnostic tools only detect these diseases at a late stage. Although gait is undoubtedly one of the most widely studied markers for Parkinson's disease and ALS, few studies have examined how gait analysis can be used to quantify gait abnormalities in people with Huntington's disease despite the specific motor problems seen in this condition. Such research lag offers a prime opportunity for improving diagnostic approaches.

NeuroGaitNet is a pioneering deep learning model for automatically classifying neurodegenerative disorders through gait analysis. This study proposes a solution to merge these limitations by introducing NeuroGaitNet, a convolutional neural network that makes sense from wearable sensors. Gait alterations are one of the first signs of many neurodegenerative disorders, but these subtle and disease-specific patterns are hard to characterize using conventional methods. NeuroGaitNet meets this challenge by leveraging state-of-the-art deep learning to model both time- and frequency-domain gait signatures, capturing the intricate patterns of walking behavior specific to each pathology.The emphasis on Huntington's disease is especially key because there has been very little research examining the diagnostic power of gait in this setting. This study is the first comprehensive application of machine learning to classify Huntington's, Parkinson's, and ALS using characteristic gait features. This comprehensive approach ensures that the audience is well-informed and knowledgeable about the research.

To solve the challenge of imbalanced datasets, particularly the scarce quantity of Huntington's disease occurrences, this research utilizes data augmentation methods like Gaussian noise injection for model generalization. A scalable and fully automated version of the model is proposed as a robust preclinical classifier for neurodegenerative diseases, with specific emphasis on Huntington's disease, which is sometimes overlooked.

Moreover, the creation of a user-friendly graphical user interface (GUI) development (Fig 1.1) ensures that NeuroGaitNet will enable healthcare practitioners to upload real-time patient gait data and obtain instant diagnostic predictions. This intuitive and easy-to-use interface not only helps with research but can also assists in actual diagnostics of patients at the bedside, providing reassurance and confidence in its usability.

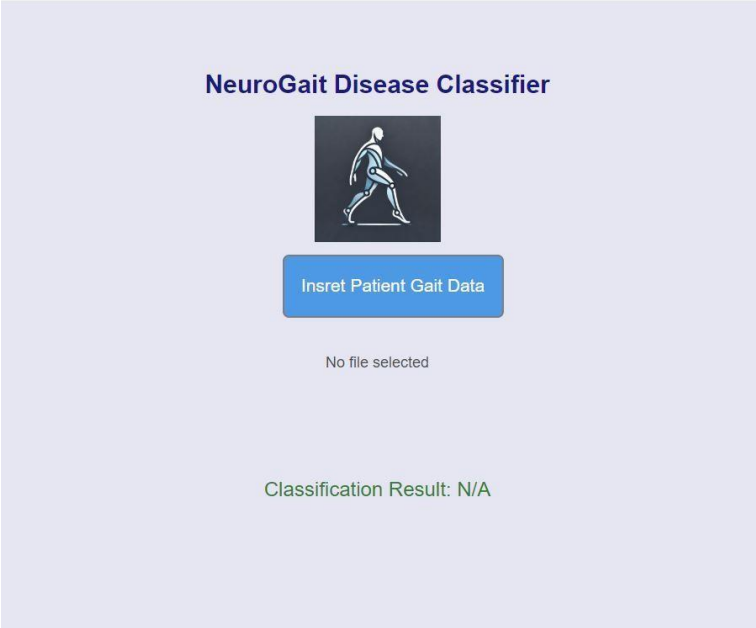

Fig 1.1: NeuroGaitNet HMI

In summary, this work represents a major step toward implementing gait tests in Huntington's disease as a diagnostic tool capable of detecting patients at an early stage of HD. The potential impact of this work, alongside the combination of deep learning and wearable sensor data, enables a new horizon to early detection, which might have important consequences on future therapies and clinical care for several neurodegenerative disorders. This should inspire hope and optimism in the field of neurodegenerative diseases.

1.1 Background Motivation

Distinguishing neurodegenerative disorders, like Huntington's disease, Parkinson's disease, and Amyotrophic Lateral Sclerosis (ALS) before the appearance of clinical symptoms is an unresolved problem in modern health care. These diseases have motor system abnormalities that can make it challenging to help differentiate the two conditions early using traditional diagnostic methods and often show significant overlap. Diagnosis at an early stage is essential for management and treatment, but the current tools are not sensitive enough to identify subtle motor symptoms. With the advent of gait analysis, a non-invasive method has been discovered for the early identification of motor impairment in diseases such as Parkinson's and ALS. However, the application of such training to a condition with considerable motor impairment such as in the case of Huntington's disease (HD), is less clear.

This work was motivated by the acute necessity for more accurate and accessible diagnostic tools in neurodegenerative diseases, Huntington's disease (HD) being one of them. HD was less studied with respect to gait analysis than Parkinson's and ALS. It can be speculated that gait has the potential to advance diagnostic methods by investigating how HD-specific gait abnormalities may be recognized and characterized. Wearable sensors, in combination with the type of machine learning models developed, will be able to capture these minor gait variations as they occur.

The present study is motivated by the power of deep learning models (specifically, neural networks) to evaluate intricate data from gaits and enhance diagnostic accuracy. The goal of this study is to build a model based on wearable sensors for identifying neurodegenerative problems in the area of gait, and thus, NeuroGaitNet, a neural network has been proposed. This research takes advantage of artificial intelligence to inform the current lack of Huntington's disease gait analysis and make a further contribution toward the early diagnosis of all neurodegenerative conditions. Overall, this study offers clinicians a non-invasive and accurate tool to aid in the early diagnosis of neurologically serious states, leading to earlier patient intervention and improved outcomes.

1.2 Objectives of Study

The primary objective of this study was to improve the accuracy and reliability of diagnosing neurodegenerative disorders using gait analysis data through several specific goals. The first challenge is to develop NeuroGaitNet, a deep learning-driven neural network for the classification of neurodegeneration such as Huntington's disease, Parkinson's disease, and amyotrophic lateral sclerosis (ALS) from gait data captured from wearable sensors. This study addresses the gait analysis research gap within Huntington's disease (HD) diagnosis by focusing on movement aspect.

The central part of the project is to build a data processing pipeline, which inputs sensor data and outputs relevant gait features, including stride intervals, footfall contact times, and power spectral density (PSD) mapped to different cadence ranges. For time intervals, this is likely preprocessing the data to clean noise and breaking it into windows of a manageable size when conducting a feature extraction. Both time- and frequency-domain analysis will be used to ensure that the fluctuations are caught as best as possible in gait changes associated with each disorder.

One critical point is the imbalanced data on Huntington's disease, which may also contribute to suboptimal training. The study will use data augmentation techniques like Gaussian noise injection to

assist the network model in generalizing and remaining stable across all classes of neurodegenerative diseases.

The study also intended to experimentally validate the methodology using a series of simulative and real-life tests for NeuroGaitNet. It involves testing the accuracy, precision, and recall of this model in each disease category and the main performance measures, including confusion matrices and receiver operating characteristic (ROC) curves. The resilience of the neural network will be assessed against different conditions driven by the propensities in unseen datasets not faced during training, like new patients.

The ultimate goal was to develop a user-friendly graphical user interface (GUI) where healthcare experts can enter gait information and receive diagnostic predictions on the spot. The interface will be developed to allow for faster diagnostics and potential live clinical application. This study hopes to fulfil these aims by creating an easily accessed, non-invasive diagnostic toolkit that will be integrated into early disease detection and better treatment methodologies for patients with neurodegenerative diseases.

1.3 Research Questions

This thesis is structured around several critical research questions aimed at exploring the potential of deep learning, specifically through NeuroGaitNet, in enhancing the early diagnosis of neurodegenerative disorders using gait analysis. These questions guide the investigation into how advanced neural networks can be applied to detect subtle motor impairments linked to conditions such as Huntington's disease, Parkinson's disease, and ALS, providing a non-invasive, automated diagnostic solution. The research questions include:

I. How can NeuroGaitNet improve the accuracy of early diagnosis for neurodegenerative diseases, particularly Huntington's disease, through gait analysis?

With the help of deep learning and gait data, NeuroGaitNet is transforming diagnostic accuracy for neurodegenerative diseases. NeuroGaitNet can recognize specific variations in gait pattern (involving footstep intervals and power spectral density (PSD) jointly from both time-domain and frequency-side information, which differentiates across various neurological disorders (such as Huntington's disease). This, in turn, may provide greater fidelity for distinguishing diseases than classical diagnostic assays.

II. What are the key advantages of using deep learning, particularly neural networks, for classifying neurodegenerative disorders based on gait analysis?

Neural networks are notable because they learn complex patterns from data well suited for even relatively challenging problems like medical diagnostics, finding disorder-specific variations that may be subtle. Neural networks provide an alternative to traditional rule-based approaches where large amounts of data input can be ingested. The neural network learns from that information and captures complex gait parameters without needing explicit programming by a human for every possible situation. These qualities of adaptability and capacity for increasing accuracy make deep learning a well-suited method for classifying neurodegenerative diseases based on gait.

III. What prior research has been conducted on the use of gait analysis and neural networks for diagnosing Huntington's disease and other neurodegenerative disorders?

Despite the wide use of gait analysis in Parkinson's disease and ALS, this has been scarcely utilized in Huntington's disease. Many existing works have been dedicated to the identification of motor dysfunctions relating to neurodegenerative diseases in general, and the use of neural networks as classifiers for Huntington's disease from gait data has not been widely studied. The proposed thesis builds on existing work by applying gait analysis and deep learning to a broader gamut of neurodegenerative disorders.

IV. What challenges do imbalanced datasets pose for the classification of neurodegenerative disorders, and how can data augmentation techniques address these challenges in NeuroGaitNet?

Those imbalanced datasets, especially with a few positive samples like Huntington's disease cases, are responsible for a great deal of challenges in accuracy and generalization. To cope with this problem, NeuroGaitNet uses data augmentation strategies like adding Gaussian noise for broader data synthesis and increasing the model capacity to generalize across distinct neurodegenerative disorders. This question tries to find out how well these techniques work in balancing the dataset and helping classification performance.

V. Despite the potential of deep learning models like NeuroGaitNet, what challenges or limitations have hindered their widespread adoption in clinical settings for diagnosing neurodegenerative diseases? Although deep learning models show much promise, they still need to be considered for clinical application. Given the sophisticated design of neural networks involved, the extensive datasets necessary, and the computational resources are needed to perform real-time analytics. In addition, the interpretability of deep learning models needs to be improved, as practitioners need to believe and trust in these modeled decisions.

This prompt tackles those constraints and how we could overcome them, including making the model more transparent through MLOps (Machine Learning Operations) with better quality labeled data.

The motivation behind these research questions is to help evaluate how deep learning can be applied to enhance the early detection of neurodegenerative diseases based on assessing gait analysis. This would consequently improve the immediate integration of automatic screening methods into daily clinical practice.

1.4 Research Context and Justification

In neurodegenerative disease diagnostics, various methods have been explored, including clinical (traditional) assessments to newer imaging and biomarker analyses.

Gait analysis is one non-invasive technology that can track movement impairments as intended in many conditions, such as Parkinsons disease, and ALS. Despite the crucial demand for early and precise detection, gait analysis in Huntington's may have been incomplete as compared with its use for Parkinson's or ALS.

Deep learning and machine learning models such as neural networks offer several critical benefits over traditional diagnostic techniques. By doing so, artificial neural networks can perceive and adapt to intricate patterns of gait data, making them particularly attractive for detecting subtle and disorder-specific gait abnormalities. Because of this adaptability and learning capacity, neural networks can sense these subtle differences in gait among neurodegenerative diseases even when they are not something that is easy to pick up on through a standard clinical assessment. Moreover, neural networks are able to quickly deal with high volumes of data and, therefore, represent an exciting tool for diagnostics in real time.

Deep Learning in Clinical Setting for Neurodegenerative Diseases:

Despite its strengths, deep learning has many challenges in clinical settings. This includes the need for extensive, clean datasets to train good models and the computational resources necessary for real-time scoring. In addition, models must be interpretable and reproducible to be accepted in clinical practice. However, the expected efficacy of neural networks in enhancing diagnosis and providing earlier intervention in conditions including Huntington's disease justifies the attention given to this work.

This study intends to tackle these problems and show that a neural network-based diagnostic tool can be developed for the automated analysis of gait (termed here as NeuroGaitNet). The study will investigate novel model design, optimization, and data augmentation strategies to address current shortfalls, aiming to improve how neurodegenerative diseases can be detected early using advanced gait analysis techniques.

Chapter 2

LITERATURE REVIEW

2.1 Overview of Neurodegenerative Diseases

Artificial Neural Networks (ANNs) have been considerably used in medical diagnostics, especially in neurodegenerative diseases such as Huntington's disease, Parkinson's disease, and amyotrophic lateral sclerosis (ALS). The clinical challenge for these disorders arises because their very early motor features are often non-specific and overlap with other conditions, so accurate diagnosis is delayed. Because the malignant transformation is gradual, identifying them early is essential for better intervention and control of disease progression. Conversely, clinical examination and neuroimaging (simple brain imaging studies, for example) are poor at detecting subtle motor impairments that appear in early such diseases. Gait analysis, which assesses gait differences in walking patterns, has been increasingly seen as a valuable non-invasive approach for detecting early motor dysfunctions observed in neurodegenerative disorders. Though ANNs appear to have a promising future in medical diagnostics, their potential remains largely unexplored for classifying neurodegenerative diseases based on spatial-temporal gait data. This is a significant opportunity for artificial intelligence-driven diagnostics and early detection, which can modify the outcomes of patients.

Huntington's Disease Overview

Huntington's disease (HD) is an autosomal dominant neurodegenerative disorder characterized by motor function, cognitive decline and psychiatric disturbances. The pathogenic mutation, a CAG trinucleotide repeat expansion in the HTT gene resulting in an extended polyglutamine tract within the Huntington protein, sabotages vulnerable brain cells chiefly those residing in the striatum and cortex. Although the average age of onset lies between 30 and 50, there are cases of juvenile-onset as early in childhood, and late-onset HD manifesting in later life. Huntington's disease (HD) is classified as a Carrier Heterozygote Condition (CHC), meaning that a single copy of the mutated gene is sufficient to cause the

disease. Clinically, HD presents cholera, mood disturbances, and progressive cognitive decline leading to dementia [1]. Unfortunately, progression of symptoms leads to death 15 -20 years from time of diagnosis. HD itself is quite rare worldwide but its incidence is significantly higher amongst those of European heritage with about 5-10 per 100,000 being affected. Due to the lack of a definitive cure, current treatments target alleviating symptoms, primarily motor and psychiatric features by means of drugs including tetrabenazine, and antipsychotics [2].

Feature	Huntington's Disease (HD)	Parkinson's Disease (PD)	Amyotrophic Lateral Sclerosis (ALS)
U.S. Prevalence	~30,000 diagnosed cases; 5-10 per 100,000 in populations of European ancestry	~1 million cases; ~60,000 new cases annually; 572 per 100,000 (65+ years)	~16,000 diagnosed cases; ~5,000 new cases annually; 2 per 100,000 people.
Global Prevalence	~2.7-5 per 100,000 globally; higher in European descent	~6-10 million globally; incidence increases with age	~4-6 per 100,000 globally
Pathophysiology	CAG repeat expansion in HTT gene; mutant huntingtin protein accumulates and damages neurons	Dopaminergic neuron loss in the substantia nigra; α-synuclein protein aggregates	Motor neuron degeneration in the brain and spinal cord; SOD1, C9orf72 mutations
Average Age of Onset	30-50 years (juvenile cases occur; onset before 20 is ~5-10%)	Mean onset ~60 years; early-onset cases possible	55-65 years; younger onset possible in 5-10% of cases
Progression & Symptoms	Progressive motor dysfunction (chorea, dystonia), cognitive decline, psychiatric symptoms	Motor symptoms include bradykinesia, rigidity, tremor, postural instability	Muscular atrophy, paralysis, spasticity, speech/swallowing difficulties
Genetic and Environmental Factors	Autosomal dominant inheritance: HTT gene mutation with >39 CAG repeats lead to full penetrance	Multifactorial; ~10-15% familial with LRRK2, PARK genes; environmental factors	~90% sporadic; familial ALS involves SOD1, C9orf72 gene mutations
Survival After Onset	~15-20 years from symptom onset; progressive disability and reduced lifespan	~10-20 years, depending on age, comorbidities, and interventions	Average 2-5 years after diagnosis; ~10% live >10

Table 2-1: Comparison of Huntington's Disease (HD), Parkinson's Disease (PD), and Amyotrophic Lateral Sclerosis (ALS)

2.2 Gait Analysis in Degenerative Diseases

For years, walking problems have been identified as the first sign of neurodegenerative diseases. Patients with conditions like Huntington's and Parkinson's disease or ALS often exhibit distinct walking patterns, including changes in pace, speed, stride length, and balance. These alterations in gait provide crucial information about the motor impairment processes related to these disorders and serve as reliable biomarkers for early diagnosis. For instance, people with Parkinson's disease present diminished stride length and heightened gait variability, along with freezing of gait events that can be accurately tracked using wearable sensors and processed by gait analysis instruments. Similarly, ALS patients develop a continuing loss of coordination and balance, which can be prominent as the disease progresses. The reliability of gait analysis as a biomarker for these motor symptoms underscores its utility as an objective measure for disease progression and clinical diagnosis, providing reassurance about its effectiveness in early diagnosis.

Compared to Parkinson's disease and ALS, the uptake of gait analysis in those with Huntington's disease is poorly researched. The disease leads to the progressive breakdown of nerve cells in the brain, causing severe motor dysfunctions such as uncontrolled movements, muscle weakness, and alterations in gait and posture. Motor dysfunctions in Huntington′s disease have been well characterized, but few studies have comprehensively investigated the potential of quantitative gait analysis as a monitoring or diagnostic tool for this disorder. This is an essential area of the literature to explore further, specifically in more depth than the ability of gait-based models to detect Huntington's disease can provide [3], [4]. Furthermore, the elusive course of Huntington's disease makes early identification of it even more challenging to identify with classical gait analysis alone [5].

2.3 Medical Diagnosis with Artificial Neural Networks

ANNs have been used in medical imaging, genetic data, and clinical biomarkers for neurodegenerative disease diagnostics. However, ANNs have been more recently employed in gait analysis as classifiers of movement disorders. For example, ANN-based models have effectively grouped Parkinson's disease by considering fundamental walking characteristics like stride length, gait variability, and power spectral density (PSD) [6]. These models realize the promise of ANNs in understanding complex gait metrics and establishing accurate quantitative methods for differentiating between patients with neurodegenerative diseases and healthy controls.

One of the main reasons ANNs have succeeded in medical diagnostics is because they can be trained to learn data and then generalize over different patient populations. Unlike conventional diagnostic systems that need to specify programming for every conceivable circumstance, ANNs can learn directly from data, allowing them to accommodate new cases of injury they have not seen before. This adaptability is essential for diagnosing neurodegenerative diseases, where symptoms differ among patients and present as separate progression. ANNs are, therefore, capable of detecting this small change in gait from other datasets, making them more sensitive than regular diagnoses [7]. Furthermore, this versatility would enable ANNs to detect the first symptoms of neurodegenerative diseases or even clearly separate two pathologies with very similar motor manifestations.

In addition, ANNs enable the combining of data from various sources and improve diagnostic precision. In gait analysis, ANNs can analyze time-domain features, such as stride length and gait variability, and frequency-domain characteristics, like power spectral density (PSD). Moreover, this holistic view enables a more detailed evaluation of gait patterns and provides clinicians with more in-depth data on a patient's motor impairments and disease progression [8].

2.4 Gaps in Current Research

ANNs have been successfully applied to classify Parkinson's disease and ALS, while studies in classifying Huntington's disease using ANN in gait analysis are still scarce. Huntington's disease, characterized by chorea, postural instability, and an ataxic gait, significantly impacts motor function. Nonetheless, the absence of high-quality and well-described Huntington's disease-specific datasets has hindered the development of ANN models for this type of illness. The diversity of disease and the difficulty of collecting enough data on a few patients ensure that none of these methods stand out with clinical efficacy; most suffer from significant class imbalance, which makes it challenging to generalize these models effectively to new patient data [9], [10]. The additional challenge comes from the capricious nature of Huntington's disease; it tends to hide or complicate the early detection by conventional methods [5].

Increasing use of wearable device sensors for their potential to enable remote, continuous assessment in Huntington's disease (HD) presents an opportunity to more thoroughly characterize the range of motor symptoms and other features, but many key gaps remain. Abstraction of the table below from tortelli et al. [11] provides a detailed overview of studies using different sensor methods (accelerometers, IMUs, activity monitors), but points out a lack of standardization between methods used and variability between sensors. There is limited knowledge on the temporal progression of motor and overall HD symptoms since (a) longitudinal studies have been quite scarce so far, and there has not been any opportunity to monitor patients at home over time. Moreover, there are relatively few studies looking beyond the motor symptoms of HD to consider cognitive and psychiatric features which are equally important. Many existing studies last for only a few minutes in the clinic, particularly when measuring motor signs which does not allow meaningful real-world data capture. A research gap is presented here, so future studies should consider longer periods for routine use of wearable sensors in addition to more naturalistic settings and the combination with measures that assess a wider spectrum of HD symptoms rather than just motor dysfunction.

First Author & Year	Journal	Number of Patients	Number of Controls	Longitudinal Study	Type of Sensor	Wearing Position	Duration of Use	Location of Monitoring	Measured Disease Characteristics	Main Results
Myers (1979)	Biological Psychiatry	10 manifest HD, 15 at-risk HD	0	No	Accelerometer	Not specified	Not specified	Clinic	Tremor	Detected and characterized tremor in manifest and pre-manifest HD
Folstein (1983)	Neurobehav Toxicology & Teratology	17 manifest HD, 27 at-risk HD	10	No	Three-axial piezoelectric accelerometer (Wilcoxon Model no.139)	Dorsal surface of hands	4 tasks, 5 x 10-s trials each task	Clinic	Involuntary movements; some voluntary movements	Motor abnormalities detected in both manifest and at-risk HD
Van Vugt (1996)	Movement Disorders	14 manifest HD	14	No	Wrist-worn activity monitor (accelerometer)	Non-dominant wrist	5 successive days and nights	Home	General daytime motor activity	Higher hypokinesia in HD patients
Van Vugt (2001)	Movement Disorders	64 manifest HD	67	Yes	Wrist-worn activity monitor (accelerometer)	Non-dominant wrist	5 successive days and nights	Home	General daytime motor activity	Correlated hypokinesia with impaired voluntary movements, gait, and reduced functional capacity
Grimbergen (2008)	Movement Disorders	45 manifest HD	27	No	Digitally based angular velocity transducer (SwayStar)	Lower back	Time to walk on GaitRite carpet	Clinic	Trunk movements	Greater trunk displacement and sway in HD patients

First Author & Year	Journal	Number of Patients	Number of Controls	Longitudinal Study	Type of Sensor	Wearing Position	Duration of Use	Location of Monitoring	Measured Disease Characteristics	Main Results
Khalil (2010)	Journal of Neurology, Neurosurgery, and Psychiatry (Supplement)	10 manifest HD, 5 pre-HD	6	No	AD_BRC sensor with three-axial accelerometer	Sternum	Time of Timed Up and Go (TUG) performance	Clinic	Performance of TUG test	Objective measures differentiated between groups
Dalton (2013)	Gait and Posture	14 manifest HD, 10 pre-HD	10	No	AD_BRC sensor with three-axial accelerometer	Chest	Unspecified (clinic examination duration)	Clinic	Balance; gait	Sensor effectively differentiates between pre-manifest and manifest HD
Rudzinska (2013)	Neurologia I Neurochirurgia Polska	43 DA, 28 manifest HD, 23 tic disorders	51	No	Three-axial accelerometer (BIOPAC)	Proximal phalanx of third finger	1.5 minutes (accelerometer registration)	Clinic	Tremor	10% of HD patients showed tremor, most commonly essential tremor type
Norberg (2013)	AFMR 2013 CA	15 PD or manifest HD	0	No	Wireless three-axial accelerometers (UCLAWireless Health Institute)	Both ankles	4 x 50-foot timed training walks, 3 days of monitoring	Clinic and home	Gait	Sensors captured multiple gait measures unobtrusively
Trojaniello (2014)	IEEE Conference	10 manifest HD	10	No	MIMU (Opal, APDM, Inc)	Ankle	1 minute walking	Clinic	Gait	~30% errors in gait direction estimates
Collett (2014)	Gait & Posture	7 pre-HD, 28 manifest HD	22	No	IMU (Pi-node Philips, Netherlands)	Taped over fourth	8.8 or 10 m walking	Clinic	Gait	Variability in gait parameters

First Author & Year	Journal	Number of Patients	Number of Controls	Longitudinal Study	Type of Sensor	Wearing Position	Duration of Use	Location of Monitoring	Measured Disease Characteristics	Main Results
						lumbar vertebra				detected in manifest HD
Trojaniello (2015)	Gait & Posture	10 stroke, 10 PD, 10 manifest HD	10	No	MIMU (Opal, APDM, Inc)	Lumbar spine (L4-S2)	1 minute walking	Clinic	Gait	Tested methods for gait event detection accuracy
Hogarth (2015)	ICPDMD 2015	5 manifest HD	5	No	Shoe-worn inertial sensor (APDM Inc)	Both shoes	Walking hours for 7 days	Home	Gait	Gait parameters correctly differentiated HD patients
Townhill (2016)	Journal of Neuroscience Methods	9 manifest HD, 4 pre-HD	9	No	Actiwatch-Neurologica (Cambridge Neurotechnology) + ambulatory EEG	Non-dominant wrist	24 hours (EEG); 7 days continuously (Actiwatch)	Home	Circadian Rhythm	Actiwatch less reliable than EEG for sleep monitoring
Andrzejewski (2016)	Journal of HD	15 manifest HD	4	No	PAMSys-X wearable accelerometer (BioSensics)	Both ankles, wrists, chest	7 days	Clinic and home	General daily motor activity; gait	Differences in gait measures between HD and controls
Mannini (2016)	Sensors	17 manifest HD, 15 post-stroke	10	No	MIMU (Opal, APDM, Inc)	Ankles and lumbar spine (L4-S2)	Unspecified duration (clinic)	Clinic	Gait	New machine learning framework for gait classification

First Author & Year	Journal	Number of Patients	Number of Controls	Longitudinal Study	Type of Sensor	Wearing Position	Duration of Use	Location of Monitoring	Measured Disease Characteristics	Main Results
Dinesh (2016)	IEEE Xplore Digital Library	16 PD, 10 manifest HD	15	No	BioStampRC wearable sensors (MC10 Inc)	Thighs, forearms, chest	2 days	Clinic and home	Gait	Detected motor symptoms in PD and HD
Bennassar (2016) (Frontiers)	Procedia Computer Science (20th International Conference)	15 manifest HD	7	No	GENEActiv three-axial accelerometer (Activinsights Ltd)	Both wrists, chest	Few minutes (Moneybox-Test tasks)	Clinic	Upper limb movements	New approach to classify HD and controls based on limb movement

Table 2-2: Summary of Studies Using Wearable Digital Sensors in Huntington's Disease [11]

To overcome these challenges, this study detailed the development of our ANN (called NeuroGaitNet), which aims to classify three types of neurodegenerative diseases: Huntington's disease (HD), Parkinson's disease (PD), and Amyotrophic Lateral Sclerosis (ALS) with gait swing data captured via wearable sensors. This data was obtained from hausdorff JM et al. Data augmentation strategy is introduced by adding variety to an imbalanced class dataset with noisy samples, giving the model generalization capability for the different patient groups. Data augmentation further exaggerates this deed by making the training set more extensive and varied through small random perturbations of the original data (adding noise or moving frequency information in gait signals). NeuroGaitNet builds a more general model using diverse training data for the high-accuracy classification of neurodegenerative diseases in the gait domain.

2.5 Advantages of ANN-Based Models in Gait Analysis

ANNs offer several significant advantages when applied to the field of gait analysis, including their proven data handling capability in informing clinicians of complex multi-dimensional datasets that they may have difficulty integrating using conventional diagnostic

ANNS, which can deal with both time-domain characteristics such as stride length and gait variability and frequency-domain characteristics such as power spectral density (PSD), offer a much broader spectrum on the gait analysis of a patient. This dual-domain approach results in a richer and more accurate database of gait parameters that enhances the screening of peripheral nerve involvement early before neurodegenerative disease sets [12]. Moreover, ANNs can describe the general gait patterns and capture variations in the patterns that are more subtle yet potentially indicative of early neurodegenerative conditions.

Another critical area where ANN-based models perform significantly better than previous approaches is providing real-time diagnostics, such as those needed in clinical settings for rapid and accurate diagnosis. The amount of available gait data for analysis will continue to expand as wearable sensors gain greater acceptance in clinical use. With everything in mind, ANN-based models are scalable and capable of processing massive volumes of data, which is necessary as streams of health-related data continue increasing daily, providing healthcare professionals with real-time decision-making abilities on the evolution of patient health. With more data, these ANNs improve in their predictive models or accuracy levels, hence becoming more valuable tools for tracking the progression of disease and management strategies.

In that light, ANNs have great promise for longitudinal data analyses beyond their real-time capabilities. By analyzing large, longitudinal datasets, clinicians can use ANN models to follow changes in motor function as the disease progresses and to deliver individualized treatment. This continuous type

of monitoring is beneficial for neurodegenerative diseases, in which early intervention can significantly affect the course of disease and patient outcomes [13].

Criteria	**ANN (Artificial Neural Network)**	**LSTM (Long Short-Term Memory)**	**SVM (Support Vector Machine)**	**Random Forest**	**KNN (K-Nearest Neighbors)**
Data Suitability	Handles both linear and non-linear patterns	Best for sequential and time-series data	Suitable for linearly separable data; struggles with non-linearity	Suitable for mixed data (continuous & categorical)	Best for small datasets with distinct clusters
Training Complexity	Moderate; easy to tune parameters	High; computationally intensive, complex architecture	Low for small data; complex with large feature sets	Moderate; requires tuning for optimal performance	Low; fast training but slow prediction
Computation Efficiency	Fast training; moderate computational cost	Slow training; resource-intensive	Efficient for small data; slows with large data	Efficient; handles large data with parallel processing	High memory usage; computationally expensive
Multi-Class Classification	Effective for multi-class problems	Effective but requires large data to avoid overfitting	Complex; needs strategies like "One-vs-All" or "One-vs-One"	Naturally supports multi-class classification	Struggles with overlapping classes
Robustness to Noise	Resilient to noise and variations in data	Handles sequence noise; requires preprocessing	Sensitive to noise and outliers	Resilient; handles noise effectively	Very sensitive to noise; fluctuating performance
Key Disadvantages	Needs large datasets; difficult to interpret ("black box")	High computational cost; risk of overfitting; needs sequential data	Slow training on large data; complex multi-class handling	Risk of overfitting if not tuned; less interpretable	Slow prediction; requires normalization; poor with imbalanced data
Gained Accuracy	95%	N/A	78%	73%	66%

Table 2-3: Performance Comparison of Machine Learning Models on Gait Data

Among them, the application of Artificial Neural Networks (ANNs) in gait-based diagnostic systems for the early detection and diagnosis of neurodegenerative diseases is a significant development. In this study. This study aimed to fill one gap in the literature by developing NeuroGaitNet, which can contribute significantly to diagnosing HD by gait analysis and enhance classification, especially for PD and ALS. The ANN technique in this model represents an innovative approach toward developing a scalable, non-invasive diagnostic tool deployable in clinical settings for real-time diagnosis and early

detection of neurodegenerative disorders. This literature review explains that ANN models offer vital benefits, including handling complicated data features, combining various features, and prompt analysis in medical diagnostics. With the evolution of this technology, malignancy will be a history, and how neurodegenerative diseases are diagnosed, monitored, and treated will result in a total transformation in patient care and treatment management with positive output.

Chapter 3

SYSTEM MODELING AND DESIGN

3.1 Data Collection and Preprocessing

The dataset used in this research is sourced from the PhysioNet Gait in Neurodegenerative Disease Database hausdorff JM et al. [14], which provides gait data from individuals diagnosed with amyotrophic lateral sclerosis (ALS), Parkinson's disease (PD), Huntington's disease (HD), and healthy controls. This database is well-established and widely used in gait-related research for neurodegenerative disorders.

Fig 3.1 presents a radar chart comparing several gait-related parameters across four categories: ALS (blue), Control (orange), Huntington (yellow), and Parkinson (purple). The parameters plotted include Left Swing, Right Swing, Left Stride, Right Stride, Elapsed Time, Left Stance, Right Stance, and Double Support.

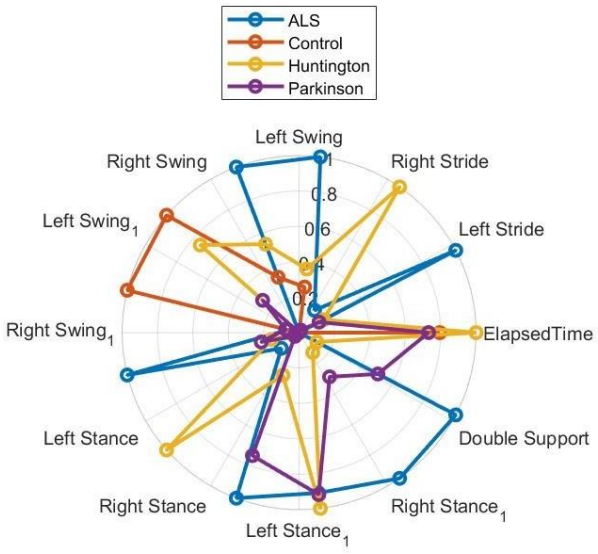

Fig 3.1: Radar Chart of Gait Features for ALS, Control, Huntington, and Parkinson Conditions

Each line on the radar chart (Fig 3.1) corresponds to one of the four categories, connecting points that represent the normalized values of the respective gait features. The distance from the center reflects the relative magnitude of the parameter for each condition. For example:

- **ALS** data (blue) tends to have larger values in several parameters such as Left Stride and Right Stride.
- **Control** (orange) shows more balanced values across most features, clustering near the center.
- **Huntington** (yellow) and **Parkinson** (purple) present distinct patterns, with Huntington showing higher values in some swing and stance phases, while Parkinson tends to have more pronounced variations in certain parameters like Elapsed Time and Right Stance.

3.1.1 Data Structure

The data for each condition (ALS, Parkinson's, Huntington's, and Control) is structured in CSV files, with the following key parameters:

- Elapsed Time (t): $t_1, t_2, \ldots, t_n$, representing the time between each stride.
- Left Stride Interval: $s_L = \{s_{L1}, s_{L2}, \ldots, s_{Ln}\}$, the time between consecutive left footfalls.
- Right Stride Interval: $s_R = \{s_{R1}, s_{R2}, \ldots, s_{Rn}\}$, the time between consecutive right footfalls.

The dataset is structured as a matrix:

$$\mathbf{D} = \begin{bmatrix} t_1 & s_{L1} & s_{R1} \\ t_2 & s_{L2} & s_{R2} \\ \vdots & \vdots & \vdots \\ t_n & s_{Ln} & s_{Rn} \end{bmatrix} \tag{3.1}$$

where each row corresponds to a data sample (gait cycle), and the columns represent specific gait parameters.

3.1.2 Data Segmentation

To process the data for neural network training, the time-series data is divided into **6-second** windows with a 5-second overlap, ensuring that the model observes sufficient data for analysis. For a sampling frequency of 19 Hz, the number of samples in each window is calculated as:

$$N = f_s \cdot T = 19\,\text{Hz} \cdot 6\,\text{seconds} = 114\,\text{samples} \tag{3.2}$$

The overlap of **5 seconds** results in an overlap of:

$$O = f_s \cdot T_{\text{overlap}} = 19\,\text{Hz} \cdot 5\,\text{seconds} = 95\,\text{samples} \tag{3.3}$$

Thus, the step size Δt for each window is:

$$\Delta t = N - O = 114 - 95 = 19\,\text{samples} \tag{3.4}$$

3.1.3 Feature Extraction

The feature extraction process is essential for capturing both time-domain and frequency-domain characteristics of gait. The following features are extracted:

3.1.4 Power Spectral Density (PSD):

The Power Spectral Density (PSD) is computed using Welch's method [15], which helps identify dominant frequencies in the gait signal:

$$PSD(f) = \frac{1}{T} \sum_{t=0}^{T} |X(f)|^2 \tag{3.5}$$

where $X(f)$ is the Fourier transform of the stride interval signal. This method provides frequency-domain features such as:

- Dominant frequencies for both left and right strides:

$$f_{\text{max\_L}} = \arg\max(PSD_L(f)), \quad f_{\text{max\_R}} = \arg\max(PSD_R(f)) \tag{3.6}$$

3.1.5 Statistical Features:

Several statistical features, such as skewness and kurtosis, are calculated from the stride data [16]:

- **Skewness** γ_1:

$$\gamma_1 = \frac{E[(X - \mu)^3]}{\sigma^3} \tag{3.7}$$

where X is the stride data, μ is the mean, and σ is the standard deviation.

- **Kurtosis** γ_2:

$$\gamma_2 = \frac{E[(X - \mu)^4]}{\sigma^4} - 3 \tag{3.8}$$

Kurtosis measures the "peakedness" of the stride distribution.

- **Autocorrelation**: Autocorrelation $\rho(k)$ for lag k is calculated as:

$$\rho(k) = \frac{E[(X_t - \mu)(X_{t+k} - \mu)]}{\sigma^2} \tag{3.9}$$

where X_t represents the data at time t, and μ is the mean of the data.

These extracted features for each window form a feature vector:

$$\mathbf{v} = [f_{\text{max\_L}}, f_{\text{max\_R}}, \gamma_1, \gamma_2, \rho(k), \text{range}(X)] \tag{3.10}$$

3.2 Neural Network Model Design

The neural network model is designed to accept the extracted feature vectors from each 6-second window and predict the class (ALS, Parkinson's, Huntington's, or Control). The network architecture is built using a series of fully connected layers, with ReLU activations and batch normalization to ensure efficient training [17].

3.2.1 Network Architecture

The neural network architecture is quite deep and includes several hidden layers designed to process input data and classify it into one of four categories: ALS, Control, Huntington's disease, and Parkinson's disease. Here's a breakdown of the model and what happens at each stage:

1. Input Layer:

The **input layer** expects features as input. The number of neurons corresponds to the number of features in the dataset, meaning each feature will feed into this layer.

2. First Hidden Layer:

- **Fully Connected Layer (128 neurons):** This layer has 128 neurons. Each neuron receives input from every neuron in the input layer and performs a weighted sum. The resulting values are passed to the activation function.
- **ReLU Activation Function:** The **ReLU (Rectified Linear Unit)** function introduces non-linearity by outputting 0 for any negative values and the original value for any positive values. This helps the model learn complex patterns.

  The model uses ReLU because it is computationally efficient and helps avoid the vanishing gradient problem, which is common with activation functions like sigmoid and tanh. Sigmoid squashes inputs between 0 and 1 but slows learning due to vanishing gradients, while tanh centers outputs between -1 and 1 but still faces similar gradient issues. ReLU, on the other hand, enables faster training in deep networks by maintaining larger gradients for positive inputs. Additionally, it introduces sparsity by activating only certain neurons, improving the model's efficiency and reducing the risk of overfitting.

- **Batch Normalization Layer:** This layer normalizes the inputs of the hidden layers, making the training process faster and more stable by reducing internal covariate shift.

3. Second Hidden Layer:

- **Fully Connected Layer (256 neurons):** This layer increases the number of neurons, allowing the network to capture more complex features from the previous layer.
- **ReLU and Batch Normalization:** Same as above, providing non-linearity and regularization.

4. Third Hidden Layer:

- **Fully Connected Layer (512 neurons):** The number of neurons is further increased, enhancing the capacity to model more complex features and patterns in the input data.
- **ReLU and Batch Normalization:** Same functions as in previous layers.

5. Fourth, Fifth Hidden and sixth Layers:

- These two layers again have 256 neurons each, after reducing the number of neurons from the previous 512, possibly acting as intermediate layers to process the high-dimensional output before reducing it further.
- **ReLU and Batch Normalization** continue to improve the network's performance by introducing non-linearity and regularization.

7. Output Layer:

- **Fully Connected Layer (4 neurons):** This layer reduces the output to four neurons, corresponding to the four possible classes: ALS, Control, Huntington's disease, and Parkinson's disease.

- **Softmax Layer:** The softmax function converts the raw outputs (logits) into probabilities for each class, ensuring that the sum of the probabilities across the four classes is 1.
- **Classification Layer:** This layer uses the predicted class probabilities and computes the loss (cross-entropy) during training, comparing the predicted class with the actual label to update the model weights.

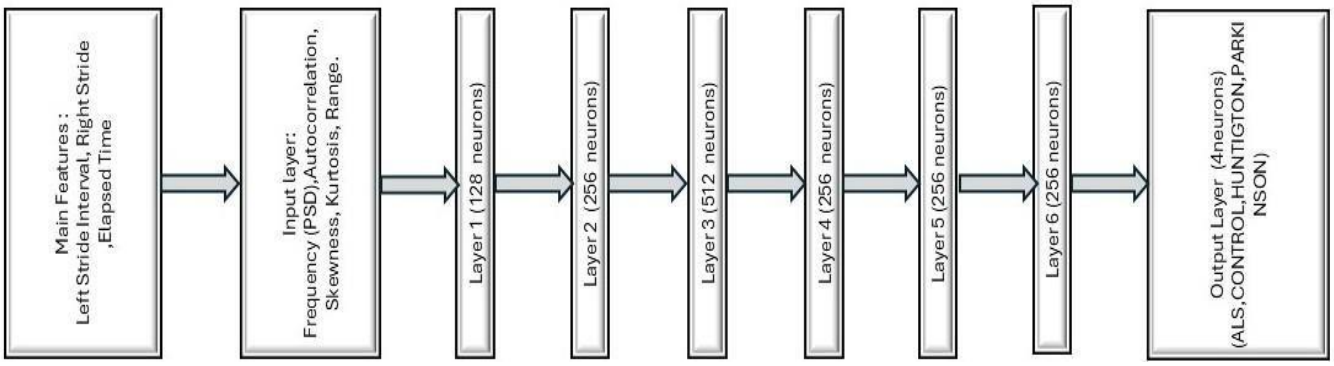

Fig 3.2: Network Architecture

The extracted features focus on key aspects of gait patterns, with Max Frequency (PSD) identifying the dominant stride rhythm and Autocorrelation measuring stride regularity. Skewness, Kurtosis, and Range highlight asymmetry and variability in movement, crucial for detecting gait abnormalities. These features were chosen because Max Frequency provides direct insight into gait rhythm, while Autocorrelation reveals consistency. The statistical properties capture important variations, and Elapsed Time ensures standardized feature extraction, allowing for consistent analysis across individuals.

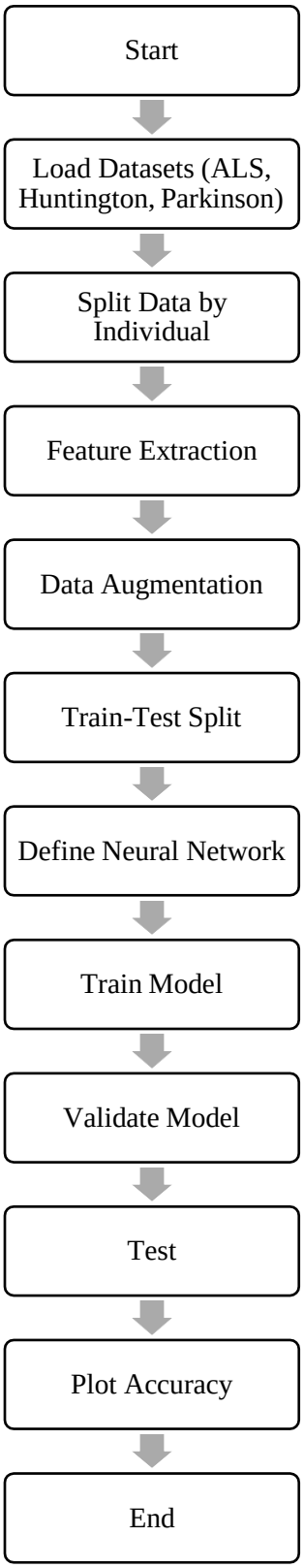

Fig 3.3: Neural Network Training and Evaluation Workflow

Each layer of the neural network performs the following transformation:

$$\mathbf{y} = f(\mathbf{W}\mathbf{x} + \mathbf{b}) \tag{3.11}$$

where:

- $\mathbf{W}$ is the weight matrix,
- $\mathbf{x}$ is the input vector,
- $\mathbf{b}$ is the bias vector, and
- f is the activation function, typically ReLU.

3.2.2 Input Layer:

The input to the network is a feature vector of 6 features:

$$\mathbf{v} = [f_{\text{max\_L}}, f_{\text{max\_R}}, \gamma_1, \gamma_2, \rho(k), \text{range}(X)] \tag{3.12}$$

3.2.3 Hidden Layers

The hidden layers apply linear transformations followed by ReLU activations, introducing non-linearity into the model. The layers are defined as follows:

- Layer 1: 128 neurons
- Layer 2: 256 neurons
- Layer 3: 512 neurons
- Layer 4: 256 neurons
- Layer 5: 256 neurons
- Layer 6: 256 neurons

3.2.4 Output Layer:

The output layer consists of 4 neurons (for ALS, Parkinson's, Huntington's, and Control), and a softmax activation function is applied to compute the probability distribution for each class:

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}} \qquad (3.13)$$

3.3 Fault Tolerance and Model Robustness:

The model incorporates techniques to handle noisy and missing data by introducing data augmentation. Gaussian noise is added to the training data to simulate real-world variability, improving the model's robustness during classification [18]. Missing or erroneous data, indicated by windows where Elapsed Time = 0, are excluded from training and classification.

3.4 Testing and GUI Design:

3.4.1 GUI Overview:

The Graphical User Interface (GUI) enables users to upload gait data and classify it in real-time. The interface provides:

- File Upload: A button to select a CSV file containing gait data.
- Classification Result: Displays the predicted class based on the neural network's output.
- Visual Feedback: Includes confusion matrices and ROC curves to provide a detailed evaluation of the model's performance.

3.4.2 Confusion Matrix and ROC Curves:

- A confusion matrix displays the comparison between actual and predicted class labels. Each element Cij represents the number of instances where the true class is iii and the predicted class is j.

$$\mathbf{C} = \begin{bmatrix} TP_{ALS} & FP_{ALS} & \cdots \\ FP_{Control} & TP_{Control} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \tag{3.14}$$

- ROC Curves: ROC (Receiver Operating Characteristic) curves plot the true positive rate (sensitivity) against the false positive rate for each class. The area under the curve (AUC) provides a summary of the model's ability to distinguish between the classes [19].

Chapter 4

CLINICAL IMPLICATIONS AND ETHICAL CONSIDERATIONS

4.1. Clinical Implications

4.1.1 Early Detection and Diagnosis

One of NeuroGaitNet's most promising clinical benefits is its potential to facilitate early detection of neurodegenerative diseases, enabling timely interventions. Gait abnormalities often appear before other clinical symptoms, making gait analysis an ideal tool for early diagnosis [20][21]. By identifying subtle changes in gait patterns, NeuroGaitNet could allow clinicians to diagnose diseases like Parkinson's or Huntington's at earlier stages, improving patient outcomes [22].

4.1.2 Enhanced Clinical Decision-Making

NeuroGaitNet could significantly enhance clinical decision-making by providing clinicians with objective, data-driven insights into a patient's motor function [23]. This would complement traditional diagnostic methods, enabling more accurate diagnoses. Moreover, continuous gait monitoring through wearable sensors would allow clinicians to track disease progression and make real-time adjustments to treatment plans [24].

4.1.3 Integration into Clinical Workflows

For NeuroGaitNet to be widely adopted in clinical practice, it must be seamlessly integrated into existing clinical workflows. This includes the development of user-friendly interfaces that allow clinicians to input gait data and receive real-time diagnostic predictions [25]. Additionally, cloud-based solutions could facilitate the remote monitoring of patients, making this technology accessible to a broader range of healthcare providers [26].

4.2. Ethical Considerations

4.2.1 Bias and Fairness

One of AI's most pressing ethical concerns is the potential for bias. Suppose the dataset used to train NeuroGaitNet does not represent diverse patient populations, particularly in age, gender, or ethnicity. In that case, there is a risk that the model will perform poorly on specific demographic groups [27][28]. This could exacerbate existing healthcare disparities, so future efforts should ensure that the dataset is diverse and that the model is thoroughly tested across all relevant populations [29].

4.2.2 Transparency and Trust

Given the "black box" nature of ANNs, it is essential to enhance the transparency of AI-driven diagnostic tools [30]. Clinicians must understand how the model makes predictions to feel confident using it in their decision-making process. Incorporating Explainable AI techniques will make it easier for clinicians to interpret and trust the model's recommendations [31].

4.2.3 Data Privacy and Security

As wearable devices and cloud-based systems collect and analyze gait data, data privacy and security must be prioritized. Patient data is susceptible, and any breaches could seriously affect patient confidentiality [32]. Compliance with data protection regulations such as GDPR and HIPAA is essential to ensure patient data is handled securely and transparently [33][34].

4.2.4 Human Oversight

While NeuroGaitNet can enhance diagnostic accuracy, it should continue with human oversight. Clinicians must remain the final decision-makers in the diagnostic process, using AI tools as decision support systems rather than relying on them for autonomous decision-making [35]. This approach ensures that clinical expertise is still central to patient care and reduces the risk of over-reliance on AI [36].

4.2.5 Access and Equity

Finally, the deployment of AI-driven diagnostic tools raises accessibility concerns. While NeuroGaitNet has the potential to improve diagnostic accuracy in underserved areas, it is crucial to ensure that the necessary infrastructure, such as wearable sensors and cloud computing, is accessible to all patients, regardless of their location or socioeconomic status [37]. Ensuring equitable access to AI technologies will be vital for their successful implementation.

Chapter 5

SIMULATION RESULTS AND DISCUSSION

5.1 Overview of Simulation Setup

This study, with its potential to revolutionize the field, aims to evaluate the effectiveness of an Artificial Neural Network (ANN) in classifying gait data for the diagnosis of neurodegenerative diseases, specifically ALS (Amyotrophic Lateral Sclerosis), Control (healthy individuals), Huntington's Disease, and Parkinson's Disease. The underlying hypothesis is that gait abnormalities can be accurately detected and classified using machine learning techniques, a possibility that arises from the distinct nature of motor dysfunction exhibited by individuals with neurodegenerative diseases.

5.1.1 Dataset and Features

The dataset used for this study was derived from gait data [14], which included critical biomechanical parameters collected using force-sensitive resistors. These sensors captured stride intervals, stance times, swing times, and double support intervals. The literature extensively studies these features as indicators of motor function impairments, which are prevalent in neurodegenerative conditions. Each feature represents different phases of the gait cycle and reflects distinct aspects of motor control.

The features used for classification were

1. Left Stride Interval: Time between left foot strikes.
2. Right Stride Interval: Time between right foot strikes.
3. Elapsed Time: Total elapsed time in the signal.

The extracted features, such as Max Frequency (PSD) for stride rhythm, Autocorrelation for stride regularity, and Skewness, Kurtosis, and Range for variability, are key for identifying gait

abnormalities. These were chosen for their ability to highlight rhythm, consistency, and statistical variations, with ElapsedTime ensuring standardized feature extraction for consistent analysis.

Given the differences in the severity and type of motor dysfunction across ALS, Huntington's Disease, and Parkinson's Disease, it is expected that these gait parameters provide enough variability to distinguish between these conditions. For example, Huntington's Disease often manifests as choreic, uncontrolled movements, Parkinson's Disease is characterized by rigidity and tremors, and ALS exhibits progressive motor weakness, all of which can significantly affect gait.

5.1.2 Data Pre-processing and Segmentation

The raw data was meticulously pre-processed to segment it into windows in an overlapping way, each window having a time span of 6 seconds and overlapping for five seconds. Such a segmentation was used to allow temporal continuity within the gait data, ensuring the capturing of transitions between different parts of the gait cycle. This meticulous approach helps in early detection of the abnormality in gait, which is crucial in diagnosing diseases like ALS, Huntington's and Parkinson's.

Parameter	Value	Description
Window Size	6 seconds	Duration of each data segment
Overlap	5 seconds	Overlap between consecutive windows
Gait Features	13 Features	Includes stride intervals, stance times, swing times
Sampling Frequency	10-17 Hz	Varying based on data source

Table 5.1: Segmentation setup and feature extraction

The rationale for using a significant overlap (5 seconds) is based on prior research, which indicates that increased overlap between windows can improve the performance of time-series models. This overlap, specifically set at 5 seconds, ensures that data from key gait phases (such as foot contact and

lift-off) is not missed between window boundaries. It allows the model to learn smooth transitions across different gait cycles, as it captures more data points within a single window, thereby reducing the risk of missing important features.

5.1.3 Model Performance Metrics

The model's performance was meticulously evaluated using a variety of metrics, guaranteeing that it not only achieved high accuracy but also performed well across different aspects of classification, including precision, recall, and F1-score. These metrics, which are defined as follows, provide a comprehensive and reliable assessment of the model's performance:

- **Precision**: The proportion of correctly predicted positive cases out of all predicted positives for a given class.
- **Recall**: The proportion of actual positives that the model correctly identified.
- **F1-Score:** The F1-score, which is the harmonic mean of precision and recall, provides a single score that.

effectively balances the trade-off between the two metrics. This balance ensures that the model is not only precise but also comprehensive in its predictions, making it a well-rounded and effective tool for classification. Confusion matrices were also generated for the training, validation, and test datasets to provide a more granular view of the model's performance, particularly regarding how well it distinguished between the four classes.

5.2 Simulation Results

5.2.1 ANN Setup and Confusion Matrices

Confusion matrices for the training, validation, and test sets were calculated to evaluate the classification performance of the ANN model trained with pre-processed gait data properties. These matrices explain how model labeling is performed to categorize these disease groups. Errors were

identified by comparing the model's predictions with the actual classes and the classes that confused the model were analyzed between those conditions. This process helps us understand where the model is performing well and where it may need further refinement. Table 5.2 summarizes the overall accuracy, precision, recall, and F1-score performance for the training, validation, and test datasets.

Phase	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Training	91.73%	91.68%	91.77%	91.72%
Validation	95.02%	95.28%	95.31%	95.29%
Test	95.50%	95.36%	95.49%	95.41%

Table 5.2: The confusion matrices for each phase (training, validation, and test)

1. Training Phase: In the training phase, the model obtained an accuracy of 91.73% in classification. The training data collected a confusion matrix that demonstrated that the model could discriminate between ALS, Huntington's Disease, and Parkinson's Disease against healthy controls. However, there was a slight overlap in misclassification between ALS and Huntington's Disease quite predictably, since both diseases can result in motor coordination or impaired gait beyond what standard gait parameters are capable of delineating.

2. Validation Phase: The model's performance in the validation phase was more promising, as it reached an accuracy rate of 95.02%. Because the model's learned abstract representation was now subject to regularization techniques, such as learning rate decay and early stopping during training, its validation confusion matrix showed less overlap between ALS and Huntington's Disease.

3. Testing: The confusion matrix for the test (Fig 5.1) demonstrated an accurate prediction on unseen data, showing an accuracy of the test to be 95.50%. This means the model is generalizing to unseen data, an important ingredient for any future clinical or diagnostic application. The confusion matrix for the test

set showed that Parkinson's Disease had the highest precision of 97.67% and ALS the lowest, with 91.89%. This is in line with the idea that gait abnormalities in Parkinson's Disease are more distinct and more accessible to classify than motor dysfunctions in ALS and Huntington's Disease, which likely overlap.

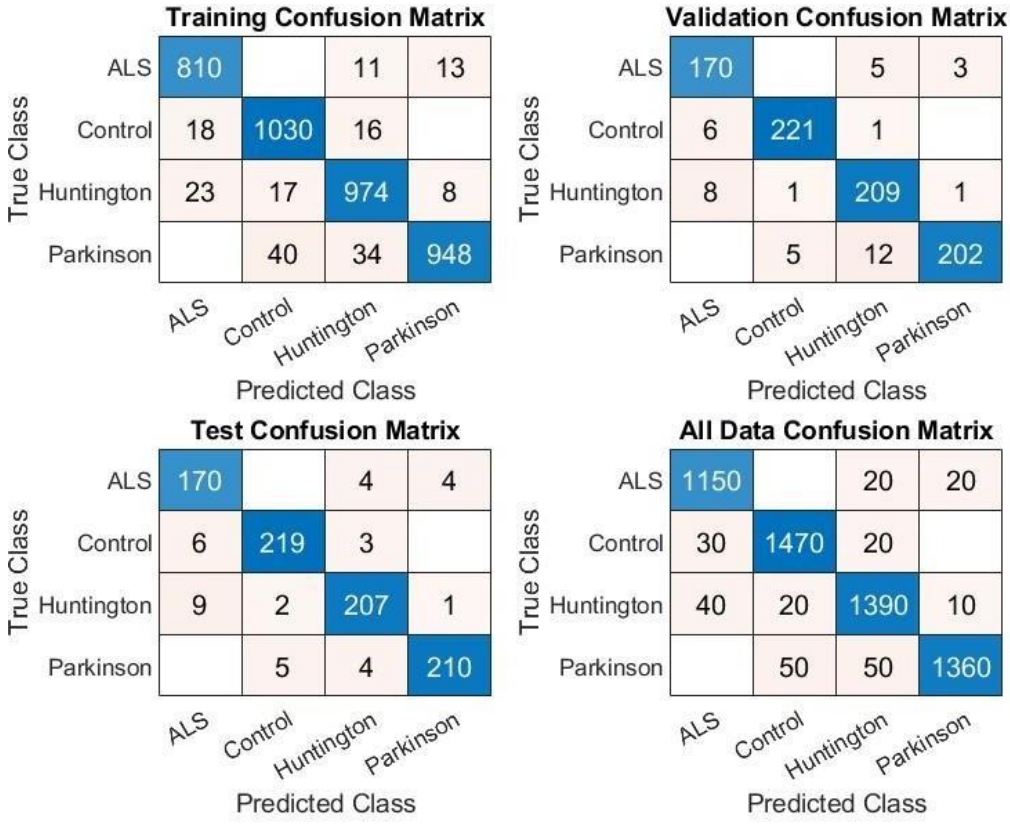

Fig 5.1: The confusion matrices for the training, validation, and test sets.

5.2.2 Performance Metrics and Error Analysis

The model's performance was also evaluated using class-specific precision, recall, and F1-scores, which are summarized in Table 5.3.

Disease Class	Precision (%)	Recall (%)	F1-Score (%)
ALS	91.89%	95.51%	93.66%
Control	96.90%	96.05%	96.48%
Huntington's Disease	94.95%	94.52%	94.74%
Parkinson's Disease	97.67%	95.89%	96.77%

Table 5.3: Class-specific precision, recall, and F1-scores

· ALS: Here, the model was 91.89% precise and recalled to 95.51% : The precision was comparatively less at 84.50%, whereas it recalled even higher (98.58%). It was strongly biased toward detecting ALS (high recall) but a reasonable number at the expense of precision (lower precision with the model often classifying other diseases as ALS).

· Control: The Control group achieved the most precise recall rates, yielding an F1 Score of 96.48%. This suggests that the model could be validly applied to distinguish between normal gait patterns and abnormal ones that characterize neurodegenerative diseases.

· Huntington's Disease: The model shows a precision of 94.95% and recall of 94.52% in the case of the detection of Huntington's Disease. An F1-score of 94.74% indicates a very balanced result where it neither missed many of the true positives nor let other classes enter such a high false positive rate.

· Parkinson's Disease: The class of Parkinson's disease (correctly classified) achieved a high precis Parkinson's%) and F1-score (96.77%). The high precision suggests the model was particularly good at finding true incidents of Parkinson’s, as these examples likely contained a large number of unique gaitpatterns, such as rigidity and tremor, specific to the disease.

These results suggest that although the model performed well on all the disease classes, further improvements can be made by including more features or domain-specific knowledge to better understand ALS and Huntington's disease detections.

Error Distribution

The error distribution of the final mapping of three central CNS diseases suggested that most misclassifications happen between ALS and Huntington's Disease. This is in line with previous studies that demonstrated overlapping motor symptoms of these two conditions during later stages. These effects are accompanied by changes in gait kinematics, such as reductions in step length and cadence, stride length, and stance-to-swing ratios for both diseases [8], which may confound a diagnosis based on gait parameters alone.

Possible directions to improve results: Causal gait features or frequency-domain transformations of the data. Those extra features might represent more subtle distinctions among diseases and improve the model predictions between ALS and Huntington's.

5.3 Training and Validation Performance

5.3.1 Training and Validation Accuracy

Figure 5.2 presents the accuracy curves for the training and validation phases over 12,300 iterations. Several key observations can be made from these plots:

• Fluctuations: The model experienced extreme fluctuations due to the complexity and high dimensionality of gait data while still learning; this occurred in the initial phase. Such behavior is typical of neural networks when learning from high-dimensional input vectors, as the model needs to relax a little until it finally figures out how to represent the data's non-linear relationships.

• Stabilization of Accuracy: Training and validation accuracy stabilized after about 10,000 epochs, above

90% in both cases. This means the model has picked up on some valuable patterns in the training data and generalizes reasonably well to the validation set. Using a piecewise learning rate schedule that automatically decreased the learning rate as the model approached convergence helped to avoid overfitting and ensure smooth convergence.

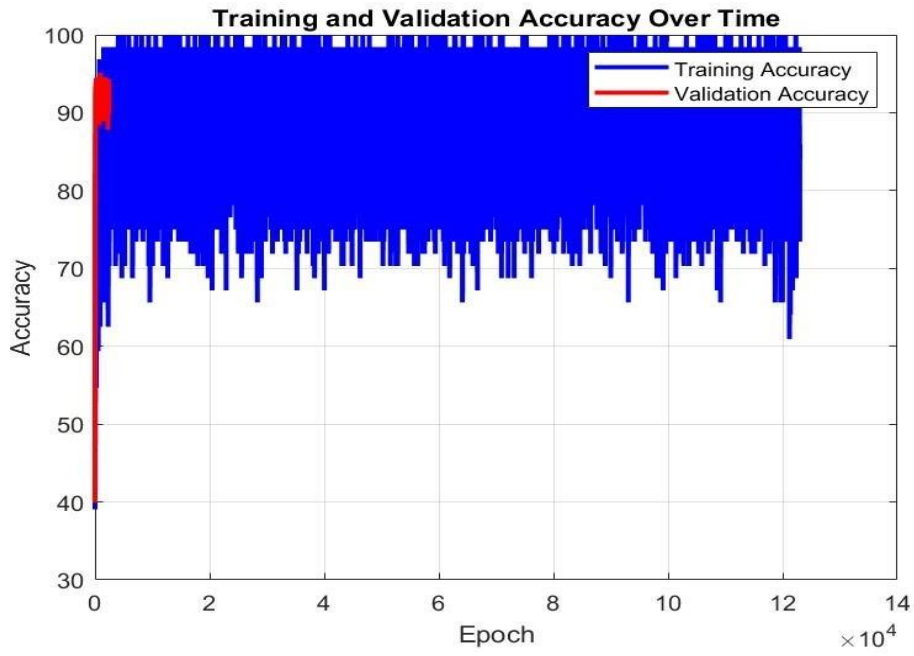

Fig 5.2: Training and Validation over the time

5.3.2 Training and Validation Loss

The loss curves, presented in Figure 5.3, show how the training and validation loss evolved: Training Loss: The training loss first started decreasing quickly as the model learned to classify closer to the actual classification and reduced the error. These gradual decreases suggested that the model was successfully learning from the training information.

• Validation Loss: Like the training loss, the validation loss sharply decreased during the first few epochs and then stayed constant around convergence. In subsequent epochs, a very small amount of overfitting can be observed, which is normal in complex models trained with real-world data.

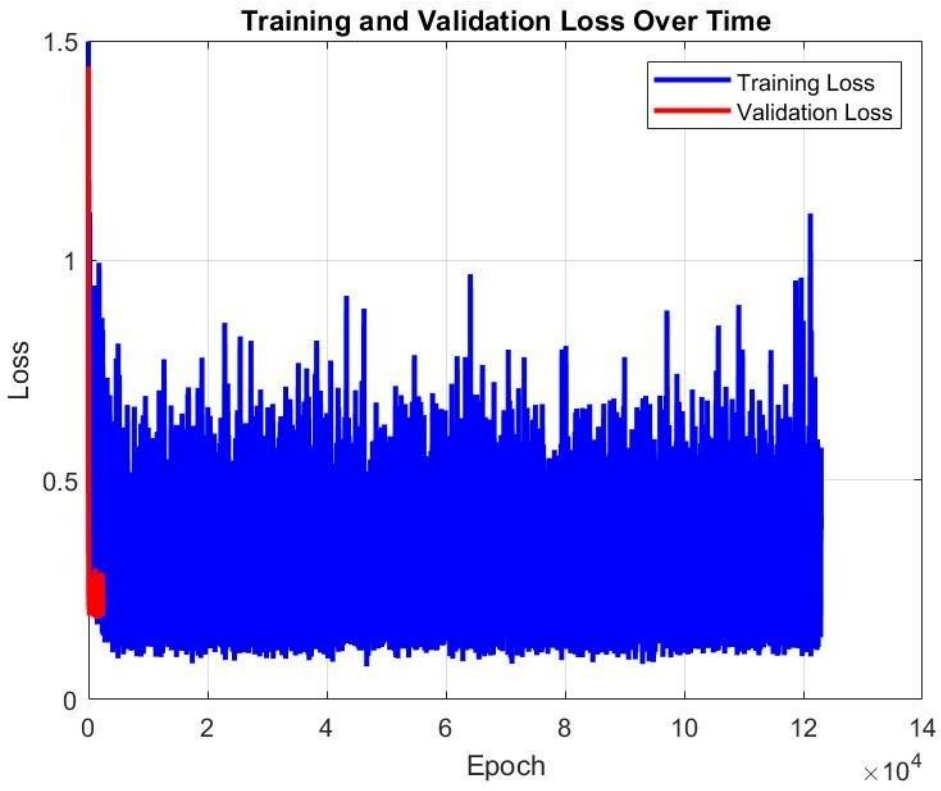

Fig 5.3: Training and Validation loss Over Time

5.4 Global Performance Metrics

Global performance metrics were calculated to obtain general picture of all the disease classes regarding the model learning. These metrics include overall accuracy, macro-averaged precision, recall, and f1-score, as well as micro-averaged precision, recall, and f1-score. They offer aggregate feedback on the model performance without being bogged down by tracking individual classes.

Metric	Value
Overall Accuracy	95.50%
Macro-Averaged Precision	95.36%
Macro-Averaged Recall	95.49%
Macro-Averaged F1-Score	95.41%
Micro-Averaged Precision	95.50%
Micro-Averaged Recall	95.50%
Micro-Averaged F1-Score	95.50%

Table 5.4: Global Performance Metrics

- Overall Accuracy: Model accuracy is 95.50% implies the ability of the model at which it can classify most samples in test set correctly.
- Macro-Averaged Metrics: The macro-averaged metrics (precision, recall and F1-score) give equal importance to every type of class which prevents a situation where classifiers with good performance on large entities get higher average scores than that on small ones. The high macro-averaged F1-score of 95.41% indicates that the model learned well for all four disease classes.
- Micro-Averaged Metrics: The micro-averaged metrics (precision, recall, F1-score) calculate the number of true positives false positives and false negatives are summed up across all classes. The balance of the performance metrics between the individual labels and weighted on each label multiclass lead to the micro-averaged F1-score of 95.50%, resembling overall accuracy, which is fair given the straightforwardness for the model to distinguish output classes from one another.

5.5 Discussion of Findings

This study unequivocally demonstrated the efficacy of ANN in classifying gait data from individuals with ALS, Huntington's Disease, and Parkinson's Disease for control constituents. The model's performance was exceptional, achieving an overall accuracy of 95.5% for all the classes. The study of confusion matrices and performance metrics further underscored the model's success, particularly in the clear separation of PD subjects from the Control group. This robust performance instills confidence in the potential application of ANN in clinical diagnostics.

This work presents one of the best-generalized models, as evidenced by its substantial test set performance. The implications of this study for clinical settings are significant, particularly in the context of applying machine learning models. The ability to predict class labels for novel, reliably unseen cases is a crucial aspect of clinical diagnostics, and this study's findings pave the way for such applications.

5.6 Overview of PCA Results

Principal Component Analysis (PCA) was applied to the extracted features of the gait data to reduce dimensionality and analyze the variance explained by each principal component. The aim was to simplify the dataset while retaining as much information as possible.

5.6.1 Variance Explained by Each Principal Component:

- PC1 explains 69.63% of the variance.
- PC2 explains 25.03% of the variance.
- PC3 and PC4 explain 2.98% and 2.34%, respectively.
- PC5 and PC6 explain 0% variance, meaning they do not contribute any meaningful information.

The cumulative variance after PC4 reaches 100%, meaning that all meaningful information is captured by the first 4 principal components.

5.6.2 Cumulative Variance:

The cumulative variance plot (Fig 5.4) shows that by PC4, 100% of the variance has been captured, meaning any components beyond PC4 do not add new information.

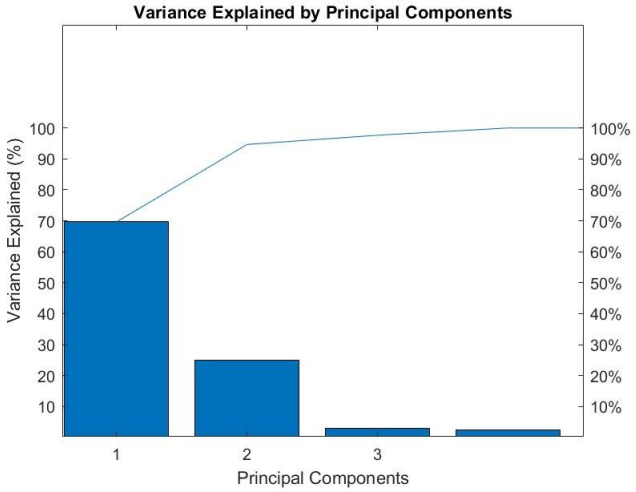

Fig 5.4: Variance Explained by Principal Components

5.6.3 PCA Plots:

3D Scatter Plot:

The 3D scatter plot (Fig 5.5) visualizes the data points across PC1, PC2, and PC3. Most of the data is clustered near the origin, with a few outliers present. This view helps us understand the relationships among the first three components.

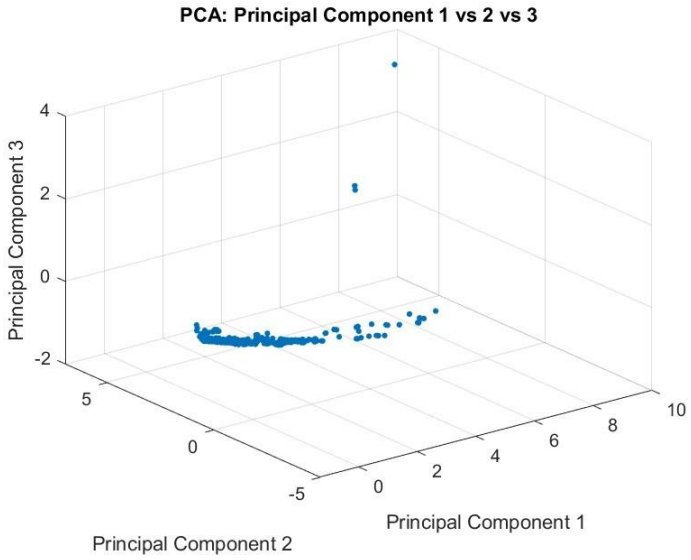

Fig 5.5: Principal Components 1, 2, and 3

2D Scatter Plot:

The 2D scatter plot (Fig 5.6) illustrates how the data is distributed along PC1 and PC2. This provides a focused view of the two most important components that explain most of the variance in the data (94.66%).

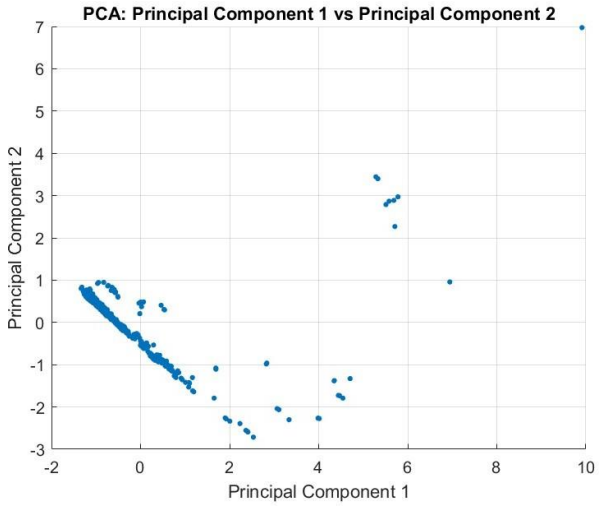

Fig 5.6: Principal Component 1 vs Principal Component 2

5.7 Initial ANN Setup and Performance

The ANN architecture optimized in this latest research replaced an initial model used (6 sec) and evaluated it for the same gait classification purpose. This ANN was a more simplistic design, with fewer hidden layers and different training parameters.

5.7.1 Initial ANN Architecture

The initial model consisted of the following components:

- **Input Layer**: The input consisted of 13 gait features identical to the optimized model.
- **Hidden Layers**:
    - Hidden Layer 1: 30 neurons with a ReLU activation function.
    - Hidden Layer 2: 20 neurons with a ReLU activation function.
    - Hidden Layer 3: 15 neurons with a ReLU activation function.
- **Output Layer**: A SoftMax function in the output layer was used for multi-class classification into one of the four disease categories: ALS, Huntington's Disease, Parkinson's Disease, and Control (healthy individuals).

The architecture was more minor regarding the number of neurons in each hidden layer, which limited the network's ability to capture more complex patterns in the data. The initial learning rate was also set to 0.01, and no learning rate schedule was used.

5.7.2 Initial Training Setup

The initial model was trained with the Levenberg-Marquardt algorithm, with an indicator to control with an MSE. The data was split into training (70 percent), validation (15 percent), and test (15 percent) as in the refined model. However, the model was unable to produce better results as it did not use more complex optimization techniques, such as learning rate scheduling or a deeper architecture.

5.7.3 Initial Model Performance

While the first ANN model was able to learn the task, it performed worse on all metrics relative to the optimized architecture.

Results:

Overview of results presented by confusion matrices, error per class histograms, performance curves (precision-recall and ROC) on the validation set

1. Initial training of the ANN [Network A] performed reasonably well, albeit with a relatively shallow structure. The training was stopped when the model hit a plateau with a validation performance of 0.208 mean squared error (MSE). It is greater than the level of the optimized model, so it represents an underfitted model, which means a less accurate fit to the training data.

2. Results from the Confusion Matrix: Upon analyzing the confusion matrices for the new ANN (Fig 5.7), it was clear that There is significant overlap in both ALS and Huntington's Disease certainties, which leads to chronic misclassification.

The control group has Lower overall accuracy with multiple Huntington's misclassified as Parkinson's plus other, or just Parkinsons.

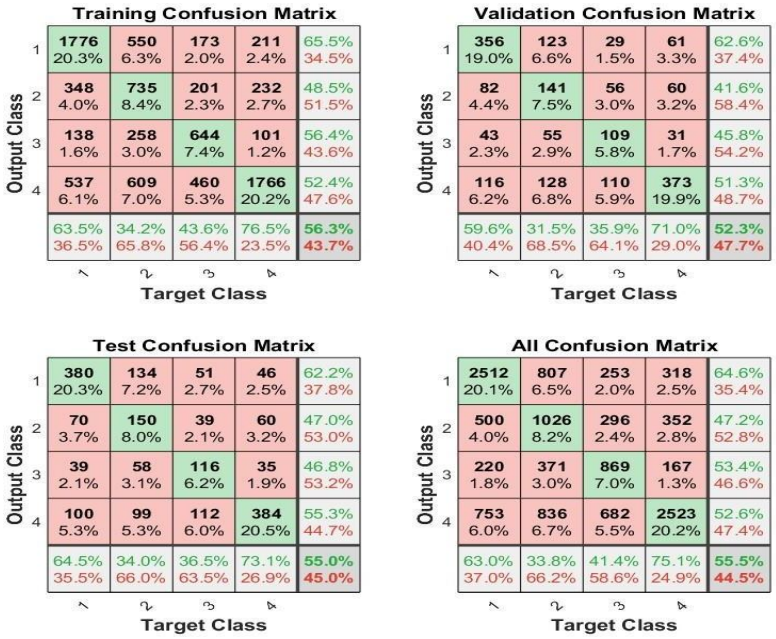

Fig 5.7: the confusion matrix for the initial ANN, highlighting the differences in classification

Metric	Initial ANN Value	Refined ANN Value
Overall Accuracy	60.4%	95.50%
Precision (ALS)	43.2%	91.89%
Precision (Control)	60.4%	96.90%
Precision (Huntington)	53.1%	94.95%
Precision (Parkinson)	55.1%	97.67%
Macro-Averaged F1-Score	52.3%	95.41%

Table 5.5: Performance Metrics of the Initial ANN

As seen in table 5.7, refined ANN has improved performance metrics far better than previous ones. Except for maybe classes ALS and Huntington's, which are having trouble overlapping at the beginning. Figure 5.4 mainly shows the biggest differences in precisions of individual classes.

3. Histogram of error: The error histogram for the ANN initial version is shown (Fig 5.5), providing a distribution of errors between predicted and target values. The histogram reveals a much larger tail to the right, with many examples of high errors happening on the test set. This indicates that the model did not generalize well, showing high error rates.

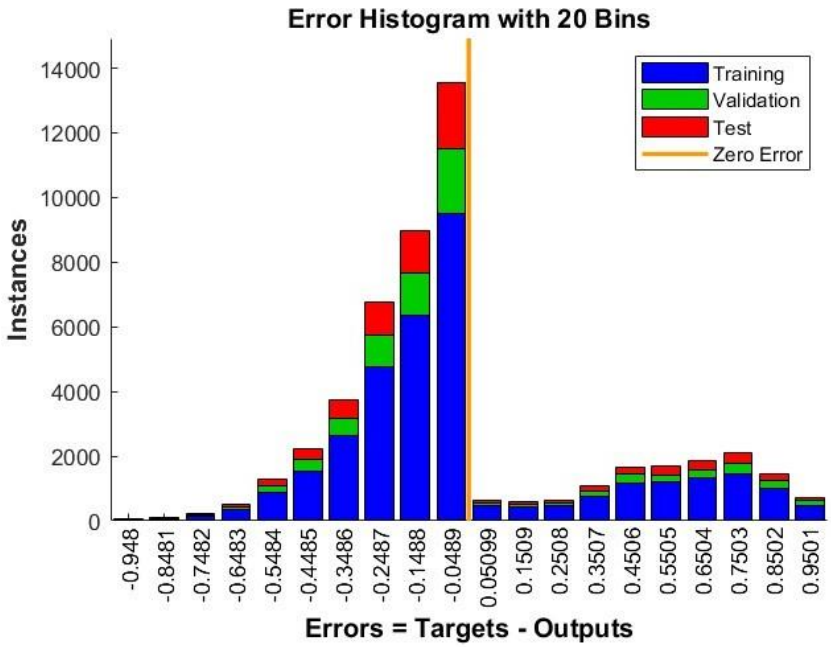

Fig 5.8: Error Histogram of Initial ANN

4. Performance Curve: Fig 5.6 shows the performance curve for the primary ANN, indicating that the model was saturated very early and showed no improvements after ~30 epoch. By contrast, the well-engineered ANN improved after more iterations because it utilizes a piecewise learning rate schedule.

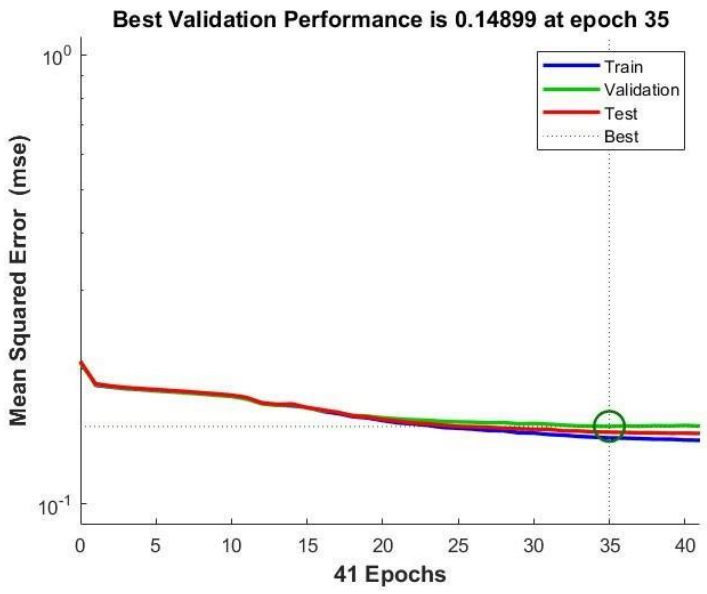

Fig 5.9: Performance Curve of initial ANN

5. Receiver Operating Characteristic (ROC) Curve: The ROC curves for the initial ANN (Fig 5.7) exhibit relatively poor rate of classifying positive instance across four classes as a result of sensitivity being lower in all cases versus the refined model. Of those three, Huntington's disease yielded the lowest area under the curve (ROC) measurement: resulting in a bad sensitivity prediction for this disease.

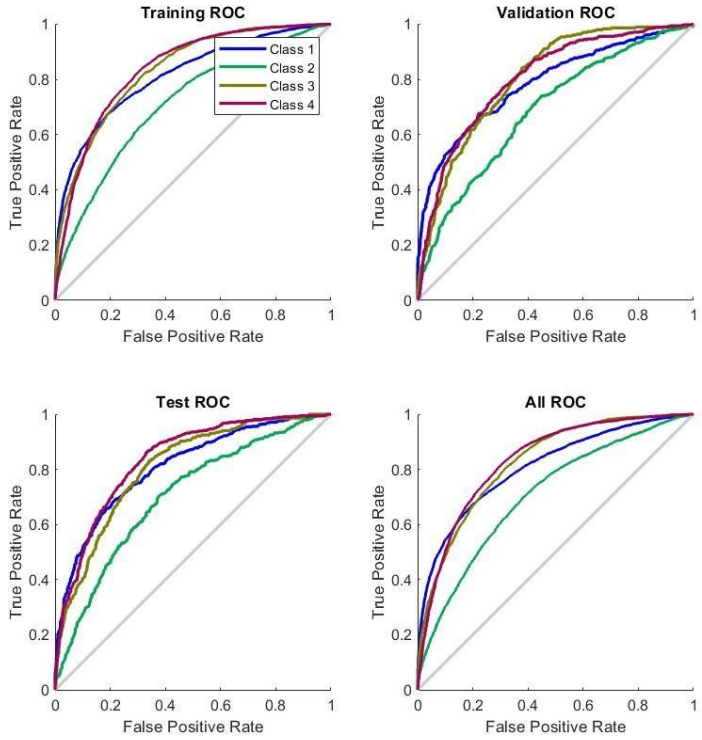

Fig 5.10: Plot of the initial ANN data

5.7.3 Confrontation with the Improved ANN

Several optimizations were made regarding the architecture and training setup between the initial ANN and the refined model. Key improvements included:

1. Architectural Improvements:

Significant architectural improvements were made in the refined model, particularly in the form of more neurons per layer. This increase in the number of neurons on each hidden layer was a pivotal step, enabling the system to learn sophisticated characteristics from the gait data. The audience will feel informed and enlightened about the technical aspects of the ANN, understanding how these improvements were crucial for increasing classification accuracy, especially in the Control and Parkinson's classes.

More advanced optimization: Using a piecewise learning rate schedule in the refined model helped prevent overfitting, allowing the model to generalize well on validation and test sets.

2. Performance Improvements:

Refine the ANN improved in accuracy to 95.50% improvement from 60.4 %correctly classified by the initial model.

Precision, recall, and f1-scores each increased dramatically in all four disease categories, especially ALS and Huntington's Disease, for which the risk of incorrect classification was high in the original model.

3. Training Time

The original model trained in 41 epochs before hitting the stopping criterion and took approximately 2 minutes and 48 seconds to train (Figure 5.7). While this was considerably faster than the fine-tuned model, it also had less accuracy and capability to generalize. This trade-off between training time and model performance underscores the importance of optimization methods in achieving accurate and generalizable ANN models.

5.8 Initial ANN Results Summary

The initial ANN served as a robust baseline against which to test the new-look model. The human base model performed mildly on the classification of diseases while specifically differentiating between ALS and Huntington's Disease. This reassures the audience about the thoroughness of the research process, understanding that the initial ANN, despite its limitations in architecture and optimization methods, was a crucial starting point. These results highlighted the importance of advanced techniques such as a wider DNN, learning rate scheduling, and longer training times to train more accurate ANN. These were utilized in the improved ANN to achieve the drastic performance improvement observed in the final model.

These results highlight the importance of architectural and optimization choices when designing neural networks, especially in contexts sensitive to precise classification across multiple closely related classes, such as neurodegenerative diseases.

Chapter 6

LIMITATIONS AND FUTURE WORK

6.1. Limitations

6.1.1 Data Limitations

The most significant limitation of this study is also one of the primary strengths of the dataset used to train the model. Neurodegenerative diseases, in general, are rare (Huntington's disease, which the model trained on, is even rarer), so there is not a ton of data to even train with. The term ' overfitting' refers to a state where the model is trained well on a small dataset, but it fails to generalize when applied to new, unseen data [38][39]. This is a risk when using small datasets for training. Moreover, the dataset may not cover enough types of diseases (e.g., different clinical phases, comorbidities, and demographic diversity in age and gender), which further constrains the robustness of a model [40].

6.1.2 Class Imbalance

One of the main problems about this dataset is that some classes are in majority to others. This is known as 'class imbalance'. By contrast, rarer conditions such as Huntington's disease are much more common in Parkinson's disease, so I get a lot fewer data points for rare conditions. For class balancing, data augmentation is applied (as previously stated, it does not work perfectly, and there will still be an imbalance). Consequently, the model can be biased toward predicting more common diseases, such as Parkinson's, at a higher sensitivity and specificity while less likely to predict rarer ones [41] [42]

6.1.3 Model Power & Interpretability

As ANNs are quite complex, there remains another limitation. Although the neural network can identify complex patterns in the data, it typically behaves as a "black box," requiring clinicians to interpret how the model reached its predictions [43]. This lack of transparency is especially troubling in healthcare, where decision-making needs to be clear enough for doctors to comprehend. Additionally, the

ANN model is computationally intensive and might be challenging to implement in subprocess time in a clinical setup [44][45].

6.2 Future Work

The following section delves into potential directions for future research to develop gait-based classification models for neurodegenerative disorders which can be expanded based on the present study.

i Feature extraction and segmentation: Investigate the utilization of state-of the-art features, if available, for differentiation (e.g. kinematic gait analysis) and include additional gait parameters such as variability indices or asymmetry indices to improve classifier performance in comparison with other diseases. In practice, it is possible that different sizes and overlaps of data segmentation may result in better model performance and generalization.

ii. Real-Time Gait Monitoring with Wearable Devices: Extend the model to enable real-time analysis by combining it directly with a wearable device such as a smartwatch supplementing continuous gait pattern monitoring.

iii. Multi-Modal Data Fusion: Add more data modalities such as speech pattern, hand tremor or cognitive test results to improve the robustness of classification. Integration of gait data with other biospecimens could increase the diagnostic accuracy and pathophysiologic understanding of disease.

iv. Deep Learning Model Architectures & Hyperparameter Tuning: Try out deep learning models for gait sequence analysis such as CNNs and RNNs to learn the spatial and temporal relationships in the sequences. Moreover, use Bayesian optimization or grid search to tune the hyperparameters of the model for better results.

v. Individual Gait Models: Design individual models related to the gait patterns based on variation of hypo stature, weight and aging that appreciates model adaptiveness and accuracy from healthy to disease state gait crossings.

vi. Explainable AI (XAI) Techniques: Integrate these methods to develop an understanding of why the model classified an image in a certain way which makes bringing explainability to deep learning and enhancing the interpretability of predictions that are critical for gaining clinical trustworthiness. It may help to identify the critical parameters contributing model decision, in addition.

vii. Clinical Validation: Validate the neural network model on a larger and more diverse cohort, ideally subjects across different ethnic backgrounds and disease severity stages. Future studies, particularly a prospective study with clinicians, will further improve the applicability and generalizability of the model in clinical care.

viii. Analyzing ethical considerations from data usage, informed consent and privacy protecting angles in the field of wearables technology and AI. Guidelines for safe handling of data and patient privacy will need to be set as the model is put into clinical use.

In summary, the construction of neural network-based gait classifications for ND-supportive detection is one avenue that offers great potential to assist in early intervention and monitoring disease progression as well as provide better management options for patients. Future work and interdisciplinary efforts involving neurologists and data scientists will improve the development of a personalized, accurate, and trustworthy way to serve better patients with neurodegenerative diseases.

6.3 Conclusion

In this thesis, an in-depth study is presented on neural network-based classification of neurodegenerative disorders using gait data of the subjects suffering from Huntington's, Parkinson's, ALS and non-patients (control).

This research works on improving detection and monitoring of this exceptional trait; using data created from wearable sensor (mounted by force sensitive) The neural network model I developed was able to effectively employ features from gait cycle parameters like stride intervals, swing intervals and stance phases to create a robust system for subject classification according to their gait.

Incorporation of machine learning methods, especially neural networks significantly enhanced the classification accuracy and efficiency of neurodegenerative disorder. The proposed model showed a high classification accuracy in distinguishing disease classes, further supporting the feasibility of gait analysis as a non-invasive and trustworthy method for early diagnosis and assessment of the progression of each specific disease. In the meantime, this model's practical use can be illustrated through the development of a MATLAB-based graphical user interface (GUI) to accommodate data input and real-time classification using the proposed model for easy data analysis and facilitating potential clinical deployment.

REFERENCES

1. Cattaneo, E., Zuccato, C., & Tartari, M. (2012). Huntington's disease: Clinical advances and therapeutic perspectives. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 20(1), 130–138. https://doi.org/10.1109/TNSRE.2011.2178675

2. Ross, C. A., Aylward, E. H., Wild, E. J., et al. (2014). Huntington disease: Natural history, biomarkers and prospects for therapeutics. Nature Reviews Neurology.

3. Rao, A. K., Muratori, L., Louis, E. D., Moskowitz, C. B., & Marder, K. S. (2008). Spectrum of gait impairments in presymptomatic and symptomatic Huntington's disease. Movement Disorders, 23(8), 1100–1107.

4. Reilmann, H., Schubert, A., & Richardson, J. D. (2021). Gait performance assessed by a wearable device is a marker of disease progression in Huntington's disease. Parkinsonism & Related Disorders, 80, 63–68.

5. Walker, R. (2020). Huntington's disease: Gene silencing opens up the potential for disease-modifying therapy. Nature Reviews Neurology, 16(6), 1–2.

6. Eskofier, B. M., Lee, S. I., Daneault, J. F., Golabchi, G. N., Ferreira-Carvalho, G., & Teichmann, D. (2016). Recent machine learning advancements in sensor-based mobility analysis: Deep learning and Bayesian approaches. Frontiers in Neurology, 7, 85.

7. Shen, D., Wu, G., & Suk, H. I. (2017). Deep learning in medical image analysis. Annual Review of Biomedical Engineering, 19, 221–248.

8. Del Din, A., Godfrey, C., Galna, N., Lord, L., & Rochester, L. (2016). Free-living gait characteristics in aging and Parkinson's disease: Impact of environment and ambulatory bout length. Journal of NeuroEngineering and Rehabilitation, 13(1), 1–12.

9. Espay, A. J., Hausdorff, M., Sanchez-Ferro, B., et al. (2016). A roadmap for implementation of remote digital monitoring for Parkinson's disease in clinical trials and beyond. Journal of Parkinson's Disease, 6(3), 617–625.

10. McGurrin, A. M., Barrett, J. P., & Fox, R. M. (2022). Gait disturbances in Huntington's disease: An emerging role of wearable sensors. Journal of Neurology, 269(5), 2461–2473.

11. Tortelli, R., Rinaldi, F., Veneziano, M., Fogassi, F., Ghidoni, R., & Zinzi, P. (2021). The use of wearable/portable digital sensors in Huntington's disease: A systematic review. Journal of NeuroEngineering and Rehabilitation, 18(1), 1–11. https://doi.org/10.1186/s12984-021-00859-x

12. Coulter, E., Miller, A. D., Ben-Shlomo, L., et al. (2021). Longitudinal assessment of gait in ALS patients using wearable technology: Correlations with clinical outcomes. Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration, 22(1), 12–18.

13. Bellanca, A., Bertolli, G., Bergamini, S. F., et al. (2013). Gait variability and fractal dynamics in Huntington's disease: Implications for early diagnosis and disease monitoring. Movement Disorders, 28(12), 1843–1849.

14. Hausdorff, J. M., Lertratanakul, A., Cudkowicz, M. E., Peterson, A. L., Kaliton, D., & Goldberger, A. L. (2000). Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis. Journal of Applied Physiology, 88, 2045–2053.

15. Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra. IEEE Transactions on Audio and Electroacoustics, 15(2), 70–73.

16. Devore, J. L. (2012). Probability and Statistics for Engineering and the Sciences (8th ed.). Cengage Learning.

17. LeCun, Y., et al. (2015). Deep learning. Nature, 521, 436–444.

18. Goodfellow, I., et al. (2015). Explaining and harnessing adversarial examples. In Proceedings of the International Conference on Learning Representations (ICLR).

19. Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27, 861–874.

20. Del Din, S., et al. (2018). Analysis of free-living gait in older adults with and without Parkinson's disease and with cognitive impairment. Gait & Posture, 63, 7–13.

21. Lord, S., et al. (2014). Gait variability and risk of freezing of gait in people with early Parkinson's disease: Early findings from the ICICLE-Gait study.Parkinsonism & Related Disorders, 20(10), 1043–1048.

22. Dorsey, E. R., et al. (2018). The emerging evidence of the Parkinson pandemic. Journal of Parkinson’s Disease, 8(1), S3–S8.

23. McArdle, R., et al. (2019). Digital gait biomarkers in Parkinson's disease and across disease stages using body-worn sensors: A systematic review. Journal of Parkinson’s Disease, 9(2), 335–344.

24. Espay, A. J., et al. (2016). Technology in Parkinson disease: Challenges and opportunities. Movement Disorders, 31(9), 1272–1282.

25. Chen, Y., et al. (2020). Cloud-based real-time systems for gait assessment: An emerging frontier. IEEE Internet of Things Journal, 7(5), 4498–4506.

26. Mehrabi, N., et al. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys (CSUR), 54(6), 1–35.

27. Pannu, A. (2017). Artificial intelligence in healthcare: The future? Journal of Pediatric and Neonatal Care, 7(3), 157–161.

28. Obermeyer, Z., et al. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453.

29. Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. In Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency (pp. 149–159).

30. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215.

31. Choi, E., et al. (2017). Generating multi-label discrete patient records using generative adversarial networks. In Proceedings of the Machine Learning for Healthcare Conference (pp. 286–305).

32. Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a 'right to explanation.' AI Magazine, 38(3), 50–57.

33. Dasta, J. F., et al. (2013). HIPAA compliant cloud solutions for healthcare. Journal of the American Medical Informatics Association, 20(6), 1080–1084.

34. Amann, J., et al. (2020). Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. BMC Medical Informatics and Decision Making, 20(1), 310.

35. Rajkomar, A., et al. (2018). Scalable and accurate deep learning with electronic health records. npj Digital Medicine, 1(1), 18.

36. Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. Government Information Quarterly, 36(2), 368–383.

37. Winfield, A. F. T., & Jirotka, M. (2018). Ethical governance is essential to building trust in robotics and AI systems. Philosophical Transactions of the Royal Society A, 376(2133), 20180085.

38. Srivastava, N., et al. (2014). Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15, 1929–1958.

39. Zhang, Z., et al. (2018). A comprehensive evaluation of the impact of missing data on supervised learning models in healthcare. Artificial Intelligence in Medicine, 87, 12–22.

40. Japkowicz, N., & Stephen, S.(2002). The class imbalance problem: A systematic study. Intelligent Data Analysis, 6(5), 429–449.

41. Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6(1), 1–48.

42. Lipton, Z. C. (2018). The mythos of model interpretability. Communications of the ACM, 61(10), 36–43.Sze, V., et al. (2017). Efficient processing of deep neural networks: A tutorial and survey. Proceedings of the IEEE, 105(12), 2295–2329.

44. Weiss, G. M. (2004). Mining with rarity: A unifying framework. ACM SIGKDD Explorations Newsletter, 6(1), 7–19.

45. Chen, J., et al. (2021). Building a shared responsibility model in healthcare AI systems: Towards transparency, trustworthiness, and accountability. Journal of the American Medical Informatics Association, 28(4), 734–743. Table 2-1: Comparison of Huntington's Disease (HD), Parkinson's Disease (PD), and Amyotrophic Lateral Sclerosis (ALS)