Title
Applied Machine Learning in Healthcare Systems: Classical and Deep Learning Approach for Gait Analysis and Activity Recognition

Permalink
https://escholarship.org/uc/item/2qc8n2h4

Author
Ramli, Albara

Publication Date
2023

Peer reviewed|Thesis/dissertation

Applied Machine Learning in Healthcare Systems: Classical and Deep Learning Approach for Gait Analysis and Activity Recognition

By

ALBARA AH RAMLI
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Computer Science

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

Xin Liu, Chair

_____

Erik K. Henricson, Co-Chair

_____

Chen-Nee Chuah

Committee in Charge

2023

i

To my beloved parents, whose unwavering love and unwavering support have been my constant source of strength and inspiration, to my dear wife and children, whose presence in my life fills me with boundless joy and gratitude, and to my siblings who have always been there for me, I am blessed to have such an amazing family.

Contents

Abstract

Advances in communication technology and hardware performance have facilitated the widespread adoption of new technologies on smart devices. In return, this has led to greater availability, lower power consumption, and lower prices. With the swift progress of both artificial intelligence (AI) and the Internet of Things (IoT) technologies, the application of human activity recognition (HAR) has become increasingly widespread across a range of domains, including security and surveillance, human-robot interaction, entertainment, and healthcare. There is a trend of applying AI to HAR systems in order to achieve superior performance and accurate medical outcome results in real time. There are different approaches involving different implementation scenarios and challenges. An enormous number of proposals have been put forth to use AI to tackle these challenges while competing to be faster and with low power consumption. In this work, we focus on an AI-based solution for healthcare systems capable of identifying patients with muscular disorders, defining the degree of mobility limitation with a proper chart and scale, and identifying and monitoring characteristics that change over time with disease progression. Our work is a novel application of machine learning (ML) on mobile devices targeting rural healthcare. Our work demonstrates how the recent advancements in IoT devices and ML technology can be adapted to measure clinical outcomes, regardless of the point of care. This application can also be used for early clinical diagnosis and planning the course of treatment, as well as monitoring the disease progression.

Acknowledgments

Indeed all praise is due to Allah alone for His unending benevolence and favor, which He has bestowed upon me. I would like to express the utmost gratitude for the blessings of family, wealth, health, and knowledge that He has generously bestowed upon me.

Then, I am incredibly grateful and fortunate to have had Professor Xin as my advisor. Her passion for research, inspirational leadership, attention to detail, continued support, and guidance have made this thesis possible. I will always be grateful for the countless hours she spent reviewing my drafts, providing constructive feedback, and challenging me to push beyond my limits. Beyond her role as an advisor, Professor Xin has been an invaluable mentor, and I am also deeply grateful for the many wonderful colleagues and collaborators I have had the pleasure of working with during my time in her lab. The relationships I have formed will continue to shape my career, and I feel incredibly fortunate to be part of such a supportive and inspiring academic community.

Also, I would like to express my gratitude to my co-advisor, Professor Erik Henricson, for his invaluable guidance, support, and encouragement throughout my Ph.D. journey. His expertise, insight, and dedication have been instrumental in shaping the direction of my research and helping me to navigate the challenges that arise in any scientific pursuit. I am deeply grateful for the opportunity to work with him and for the many valuable lessons and insights he has shared with me along the way. Thank you for your unwavering support and for always pushing me to strive for excellence

I would like to express my appreciation to Professor Chen-Nee Chuah, a member of my dissertation committee, for her invaluable contributions to my research. Her insightful comments, constructive feedback, and critical evaluation have been instrumental in shaping the final outcome of this work. I am also grateful for her valuable contribution as a member of my qualifying exam committee. Thank you for sharing your expertise, devoting your time, and providing unwavering support in helping me achieve my academic goals.

I would like to extend my deepest thanks to my parents, who have been perfect role models for me throughout my academic journey. Their unwavering support, encouragement, and guidance have been instrumental in shaping the person I am today. They have provided me with the freedom to pursue my interests and passions, and I am grateful for every opportunity they have given me.

Thank you for your unconditional love and support and for the countless sacrifices you have made to help me achieve my goals.

I am also grateful for the love and support of my wife and children. Your unwavering support, patience, and understanding have been a constant source of inspiration and motivation for me throughout this process. Your presence has brought balance and joy to my life, and I am grateful for the sacrifices you have made to support me on this journey. Thank you for being my rock and my safe haven during the challenging moments and for celebrating my successes with me. I am blessed to have you in my life.

I would also like to express my heartfelt gratitude to my brothers and sisters for their unwavering love, support, and encouragement throughout my academic journey. Their belief in me and constant support have been a source of strength and inspiration, and I am grateful for every moment we have shared. Thank you for always being there for me and for your unwavering support in all of my endeavors.

CHAPTER 1

Introduction

In the field of Artificial Intelligence (AI), especially Deep Learning (DL), there have been significant advances in recent years, particularly in healthcare applications. While AI is slowly finding its way into real-world healthcare applications, there are still many challenges that need to be addressed. We propose to bridge the gap between computer science and healthcare by addressing some of these challenges with a robust, reliable solution infrastructure capable of using structured medical datasets for many real-world healthcare problems.

Specifically, in the field of human activity recognition (HAR) research, most of the studies explore and improve models that focus on daily activity recognition. There are publicly available datasets modeling daily living activity recognition. However, these datasets are not enough for our research. We need more precise and medical-grade datasets in order to answer questions related to muscle disorders. Data collection comes with its own unique challenges:

- There is a need for extensive cooperation between the Medical and Computer Science fields in order to solve existing real-world healthcare problems.
- Collecting data is expensive and time-consuming in medical experiments., which led researchers tackling on a limited set of healthcare problems such as daily-living-activities data[1].
- The medical experiments require paperwork, hospital permission, and guardians' consent.

---

[1]Collecting reliable healthcare data is more challenging than daily-living-activities data. For some daily-living-activities such as walking and running, researchers could get a significant amount of the labeled data from the public dataset or the raw sensor data collected and annotated by themselves. However, for some specific human activities related to healthcare, such as the therapeutic activities of patients, researchers could not get enough sensor data since these activities are low-probability available compared with daily-life-activities. While daily-living-activities focus on the differences between various types of activities, other healthcare-problem such as Duchenne Muscular Dystrophy (DMD) focuses on the quality of performing the same type of activities.

- It's difficult to find a clean healthcare-dataset that was not manipulated in any way, such as format, storage, and structure. Healthcare-datasets don't conform to a uniform standard that can be used to automatically extract meaningful information.
- Healthcare-datasets need to be annotated by experts in the domain knowledge, while daily-living-activity annotation can be done with regular people.

There are also other challenges in terms of methods used to collect data from different sources, such as sensory-based/camera-based. On top of these challenges, some challenges need to be addressed regarding AI models selection, models' configuration, models' evaluation, and data environment:

- For the **sensory-based method**, challenges include finding appropriate sensor position, sensor's orientation, number of sensors, types of sensors, and sensor sampling rate. For the **camera-based method**, challenges include finding answers to problems related to fixed/flexible camera location, the number of cameras, available computational resources, and storage limitation.
- From the **AI** and **machine learning** side, challenges include selecting the best model that fits the problem (both classical ML and DL), choosing the appropriate Time-Window (TW), which includes fixed/flexible length-size with/without overlapping.
- The application to **healthcare** creates more challenges from the **data-collection** to the end-to-end **model-evaluation**:
  - **Data-collection** poses challenges such as the limited number of subjects, the difficulty to find enough diversity in the subjects, the unbalanced number of subjects in each category, the noise in the environment interfering with the sensor data, data annotation, data segmentation, storage and structuring of the data.
  - **Model-evaluation**: It's challenging to evaluate the performance of an AI model due to the lack of a benchmark dataset. This is because each disease has its own focus and characteristics. Thus, a generalized model needs to be verified by evaluating with medical data, from other diseases, that share the characteristics with the data that was used to train the current model.

1.1. Overview of Thesis

Our research centers on the application of Human Activity Recognition (HAR) in the context of diseases that impact the patient's mobility. Our main objective is to develop methods for early disease diagnosis, assessing its severity, and tracking its progression. Two primary techniques are available to monitor a patient's movements: sensor-based, which relies on time-series data, and camera-based, which employs pictures and videos.

Chapter 2 offers an overview of HAR using wearable sensors, emphasizing its potential applications in healthcare. In Chapter 3, the focus shifts to feature engineering that is tied to clinical features. Specifically, the chapter explores the use of a single smartphone-based accelerometer to extract temporospatial gait clinical features (CFs) in both typically developing (TD) individuals and those with Duchenne muscular dystrophy (DMD). These CFs, which have clinical significance in gait analysis, are utilized as input features for classical machine learning (CML) models in Chapter 4 to classify gait patterns in children with DMD and TD peers.

Chapter 4 explores the use of a single smartphone-based accelerometer with the extract of CFs from Chapter 3 and the raw acceleration signals without feature engineering to identify the children with DMD and TD peers. Chapters 5 and 6 focus on camera-based data. Chapter 5 describes a camera-based deep learning (DL) method that detects patients' eye blinks and translates them into words. In contrast, Chapter 6 showcases a deep ensemble learning approach that uses kinetics and kinematics data from multiple cameras and markers to identify post-stroke survivors and age-matched controls. These two chapters demonstrate the potential of using camera-based methods for disease detection and monitoring and highlight the power of machine learning (ML) in analyzing complex data from different sources. Chapter 7 presents Walk4me, a telehealth system that uses signal processing and ML to automate the CFs detection in TD individuals and those with muscle disorders. The system collects participant data remotely and in real-time from device sensors, using the approaches presented in Chapters 3 and 4. It then extracts CFs to train CML models, as described in Chapter 4, and DL models based on raw signals, as described in Chapters 4 and 6, to identify patterns that can facilitate the identification of muscle disorders.

While the specific disease detection and monitoring tasks may differ across the chapters, they all likely utilized similar underlying design principles such as data preprocessing, feature engineering,

model selection, hyperparameter tuning, and model evaluation to develop ML-based solutions. By applying these principles, we were able to develop effective ML-based solutions that can help in the diagnosis and monitoring of various diseases affecting mobility. Accordingly, the core idea is divided into the following chapters:

- **Chapter 2** Overview of human activity recognition using wearable sensors [**70**]. This work provides an overview of HAR using wearable sensors, with a focus on healthcare applications. It highlights the essential components of designing HAR systems, including sensor factors, AI model selection, and feature engineering, and proposes several research opportunities for both the medical and computer science communities. The chapter presents two emerging HAR projects in healthcare: early mobility identification for ICU patients and gait analysis for DMD patients and discusses the hardware design and accurate AI models required to recognize patients' activities. The chapter also covers the challenges of building a HAR healthcare system and future opportunities in this field.

- **Chapter 3** Automated Detection of Gait Events and Travel Distance Using Waist-worn Accelerometers Across a Typical Range of Walking and Running Speeds[2]. This work investigates the use of a single smartphone-based accelerometer placed near the body's center of mass to extract CFs in healthy adults during ambulation at varying gait speeds. Machine learning-based algorithms (ML) were developed to estimate temporospatial gait CFs, including the number of steps, distance, duration of the gait cycle, gait speed, and step length. The estimated temporospatial gait CFs were compared with ground-truth CFs determined by expert observers during clinical testing using the 6-minute walk test, 100-meter run/walk, and self-selected free walk. Additionally, the extracted CFs from the accelerometer were compared with CFs extracted by the pedometer feature built into the smartphone. The study findings indicate that a single accelerometer placed near the body's center of mass can accurately measure CFs across different gait speeds in healthy adults, suggesting the potential for measuring CFs in the community without the need for ground reaction force (GRF) measurements.

---

[2]In submitting process to the Gait & Posture journal

- **Chapter 4** Gait Characterization in Duchenne Muscular Dystrophy (DMD) Using a Single-Sensor Accelerometer: Classical Machine Learning and Deep Learning Approaches[3] [**95**]. This chapter discusses the use of a single smartphone-based accelerometer to extract gait clinical features in children with DMD and TD peers during ambulation at varying gait speeds. Machine learning-based algorithms were developed to estimate gait clinical features, which were compared to ground-truth clinical features determined by expert observers during clinical testing. The study found that a single accelerometer placed near the body's center of mass can accurately measure clinical features across different gait speeds in both TD and DMD peers, suggesting the potential for measuring clinical features in the community without the need for ground reaction force measurements.91%.

- **Chapter 5** BWCNN: Blink to Word, a Real-Time Convolutional Neural Network Approach [**91**]. This work discusses one of the brain diseases that cause damage to the central neural system (CNS). This disease impairs the patient's speech ability and motor functions. This study used a camera-based method to collect facial data and built an AI-based system that detects patients' eye blinks and converts them to words. This work focuses on real-time scenarios where we are trading off between latency and accuracy. Our best model achieves an accuracy of 99.20% and 94ms latency in eye blink recognition.

- **Chapter 6** A Deep Ensemble Learning Approach to Identify Stroke-related Features of Gait[4]. This work discusses the use of a deep ensemble learning framework to identify stroke-related features of gait. The study aimed to apply machine learning methods to a gait biomechanical package of post-stroke and age-matched control groups to automatically identify post-stroke population, reveal the biomechanical data signals that best characterize post-stroke gait, and quantify the impact of force platform data on the tasks. The study's current results indicate that up to 100% can identify more vs. less affected sides and 99% in identifying post-stroke vs. age-matched control, showing promise towards gait analysis automation as a tool to better support clinical decisions and extensive studies in subjects' natural environments.

---

[3]In submitting process to the PLOS One journal
[4]In submitting process to the IEEE Transactions on Neural Systems and Rehabilitation Engineering journal

- **Chapter 7** Walk4Me: Telehealth Community Mobility Assessment, An Automated System for Early Diagnosis and Disease Progression [**92**,**93**,**94**][5]. This work presents Walk4me, a telehealth system that uses Artificial Intelligence (AI) to detect gait characteristics in patients and typically developing peers. The system collects data remotely and in real-time from device sensors, extracts gait characteristics and raw data signal characteristics, and uses machine learning techniques to identify patterns that can facilitate early diagnosis, identify early indicators of clinical severity, and track disease progression across the ambulatory phase of the disease. The authors have identified several machine learning techniques that differentiate between patients and typically-developing subjects with 100% accuracy across the age range studied and have identified corresponding temporal/spatial gait characteristics associated with each group. The system has potential to inform early clinical diagnosis, treatment decision-making, and monitor disease progression.

In order to provide a more comprehensive understanding of the content and contributions of this thesis, I briefly describe each chapter.

1.1.1. An Overview of Human Activity Recognition Using Wearable Sensors: Healthcare and Artificial Intelligence.

With the rapid development of the internet of things (IoT) and artificial intelligence (AI) technologies, human activity recognition (HAR) has been applied in a variety of domains such as security and surveillance, human-robot interaction, and entertainment. Even though a number of surveys and review papers have been published, there is a lack of HAR overview papers focusing on healthcare applications that use wearable sensors. Therefore, we fill in the gap by presenting this overview work [**70**]. In particular, we present our projects to illustrate the system design of HAR applications for healthcare. Our projects include early mobility identification of human activities for intensive care unit (ICU) patients and gait analysis of Duchenne muscular dystrophy (DMD) patients. We cover essential components of designing HAR systems including sensor factors (e.g., type, number, and placement location), AI model selection

---

[5]The system has been actively used since 2020 in the Neuromuscular Research Lab at the Department of Physical Medicine & Rehabilitation to collect and analyze data from typically developing (TD) and Duchenne muscular dystrophy (DMD) peers. Its value was especially apparent during the COVID-19 pandemic, enabling remote assessments. Additionally, the system is utilized in both the Neuromuscular Research Lab and the Biomechanics, Rehabilitation, and Integrative Neuroscience (BRaIN) Lab to collect and analyze data from stroke survivors and age-matched controls.

(e.g., classical machine learning models versus deep learning models), and feature engineering. In addition, we highlight the challenges of such healthcare-oriented HAR systems and propose several research opportunities for both the medical and the computer science community.

1.1.2. Automated Detection of Gait Events and Travel Distance Using Waist-worn Accelerometers Across a Typical Range of Walking and Running Speeds. This work[6] investigates the use of a single smartphone-based accelerometer placed near the body's center of mass to extract temporospatial gait clinical features (CFs) in children with Duchenne muscular dystrophy (DMD) and typically developing (TD) peers during ambulation at varying gait speeds. Machine learning (ML)-based method were developed to estimate temporospatial gait CFs, including the number of steps, distance, gait speed, step length, and duration of the gait cycle. The estimated temporospatial gait CFs were compared with ground-truth CFs determined by expert observers during clinical testing using the standard 6-minute walk test, 100-meter run/walk, and self-selected speed free-walk. Additionally, the extracted CFs from the accelerometer were compared with CFs extracted by the pedometer feature built into the smartphone. The study findings indicate that a single accelerometer placed near the body's center of mass can accurately measure CFs across different gait speeds in both TD and DMD peers, suggesting that there is potential for measuring CFs in the community without the need for ground reaction force (GRF) measurements.

1.1.3. Gait Characterization in Duchenne Muscular Dystrophy (DMD) Using a Single-Sensor Accelerometer: Classical Machine Learning and Deep Learning Approaches. Differences in gait patterns of children with Duchenne muscular dystrophy (DMD) and typically-developing (TD) peers are visible to the eye, but quantifications of those differences outside of the gait laboratory have been elusive. In this work[7], we measured vertical, mediolateral, and anteroposterior acceleration using a waist-worn iPhone accelerometer during ambulation across a typical range of velocities. Fifteen TD and fifteen DMD children from 3-16 years of age underwent eight walking/running activities, including five 25 meters walk/run speed-calibration tests at a slow walk to running speeds (SC-L1 to SC-L5), a 6-minute walk test (6MWT), a 100

---

[6]In submitting process to the Gait & Posture journal
[7]In submitting process to the PLOS One journal

meters fast-walk/jog/run (100MRW), and a free walk (FW). For clinical anchoring purposes, participants completed a Northstar Ambulatory Assessment (NSAA). We extracted temporospatial gait clinical features (CFs) and applied multiple machine learning (ML) approaches to differentiate between DMD and TD children using extracted temporospatial gait CFs and raw data. Extracted temporospatial gait CFs showed reduced step length and a greater mediolateral component of total power (TP) consistent with shorter strides and Trendelenberg-like gait commonly observed in DMD. ML approaches using temporospatial gait CFs and raw data varied in effectiveness at differentiating between DMD and TD controls at different speeds, with an accuracy of up to 100%. We demonstrate that by using ML with accelerometer data from a consumer-grade smartphone, we can capture DMD-associated gait characteristics in toddlers to teens.

1.1.3.1. *Highlights on Duchenne Muscular Dystrophy (DMD).* Duchenne Muscular Dystrophy (DMD) is a fatal X-linked congenital genetic disorder affecting 1/5000 males [94]. The disorder leads to progressive, debilitating muscle weakness and loss of ambulation around the age of 12 with early death due to cardiac and respiratory complications in the late teens or early 20s. Weakness is detectable from infancy, but symptoms are typically recognized during toddler or early childhood years. By facilitating early diagnosis, Gene repair interventions and other preventive therapies can be initiated as early as possible to slow the disease's progress and prevent secondary conditions.

Existing works on muscle disorder diagnosis [7] have limitations including high cost, required expertise, and cannot be done remotely. The majority of the studies on muscle disorder diagnoses rely on methods such as multi-camera-based assessment [98], Magnetic Resonance Imaging (MRI)-gene-image [27], muscle-biopsy [52] and genetic-testing [1].

The Multi-camera-based assessment requires high-end cameras, dedicated lab environment and hence can't be administered remotely. Additionally, processing and storing this data requires significant compute capabilities and large storage which can result in large operation costs.

The MRI-based-image method uses MRI images to diagnose children with DMD from TD children. MRI devices are expensive, require operational expertise and as such are not suitable for community use. This method also requires patients to be physically present at the point-of-care since taking MRI images cannot be done remotely.

8

The muscle-biopsy and genetic-testing uses samples to diagnose DMD. Collecting these samples is also time consuming and requires human expertise, which can lead to increased wait times when scheduling tests. Guardian availability can also be a challenge when the point-of-care is located in a remote area. It may require lots of traveling time and further expenses, including travel, transportation, and housing costs.

Researchers in [9] present a promising approach of evaluating disease progression in dystrophin-deficient canines, who exhibit a disease course that closely mirrors human DMD disorder. They extract features from the gait pattern of these canines with an accelerometer and record the readings while the animals walk for a predetermined period of time. While the paper presents promising results, the scope of the study is limited to canines. Additionally, the experimental setup consists of just a single activity (walking) over a short duration of time (10 sec).

There is a real demand for alternative methods that can mitigate some of the limitations of reproducing the methods for community use. We believe that gait analysis can aid in early DMD diagnosis, making it possible to slow the disease's progress and prevent or delay secondary conditions through earlier clinical intervention.

In this work, we present a fast, highly-usable, cost-effective AI-based human DMD detection system. Our system uses a commodity accelerometer found on most modern smartphones as our primary data source, coupled with a state-of-the-art AI system to detecting human DMD. Our system can be easily applied in the community on a large-scale for pre-screen and disease treatment, especially in remote areas. DMD diagnosis is a challenging task since there is no single gait pattern representing the symptoms of a gait disorder. Establishing a diagnosis is mainly based on extracting effective clinical features (CF) of gait disorders [87]. In this study, we identify DMD-related gait abnormalities using two types of features: (1) Eight extracted gait clinical features (CF) similar to an 8 CF where utilized on Canine gait assessments [9]. (2) RAW accelerometer signal.

As part of our prototype, we have developed a system that consists of a smartphone app to collect raw signal data remotely from a built-in accelerometer sensor on patients' smartphones in conjunction with a web application on a cloud server to receive the data from smartphones [94]. Then, we extract the temporal/spatial gait characteristics from the gathered data. Finally, we use

classical machine learning (CML) and deep learning (DL) techniques [**66**] on the data to capture the gait changes associated with DMD.

In our evaluation, we examine the reliability and effectiveness of using raw data and 8 CF with AI to detect DMD gait patterns. The AI used in these experiments can be further classified based on the classical machine learning (CML) and deep learning (DL) approaches. Our results show a great deal of promise with 91% accuracy in predicting DMD gait patterns.

1.1.4. BWCNN: Blink to Word, a Real-Time Convolutional Neural Network Approach. Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disease of the brain and the spinal cord, which leads to paralysis of motor functions. Patients retain their ability to blink, which can be used for communication. Here, We present [**91**] an Artificial Intelligence (AI) system that uses eye-blinks to communicate with the outside world, running on real-time Internet-of-Things (IoT) devices. The system uses a Convolutional Neural Network (CNN) to find the blinking pattern, which is defined as a series of Open and Closed states. Each pattern is mapped to a collection of words that manifest the patient's intent. To investigate the best trade-off between accuracy and latency, we investigated several Convolutional Network architectures, such as ResNet, SqueezeNet, DenseNet, and InceptionV3, and evaluated their performance. We found that the InceptionV3 architecture, after hyper-parameter fine-tuning on the specific task led to the best performance with an accuracy of 99.20% and 94ms latency. This work demonstrates how the latest advances in deep learning architectures can be adapted for clinical systems that ameliorate the patient's quality of life regardless of the point-of-care.

1.1.5. A Deep Ensemble Learning Approach to Identify Stroke-related Features of Gait. Stroke is the leading cause of disability in adults worldwide, where the majority of stroke survivors experience walking dysfunction. To facilitate rehabilitation, the kinematics and kinetics of a person's gait can be measured in a motion capture laboratory instrumented with force platforms. However, this process is costly and requires significant buy-in, not only from the participant but also in terms of laboratory equipment. To address this issue[8], we ask the following question: given the vast biomechanical signals measured in a motion capture experiment, which signals best capture stroke-related features of gait? By identifying such pertinent features, we can move closer

---

[8]In submitting process to the IEEE Transactions on Neural Systems and Rehabilitation Engineering journal

to identifying post-stroke gait outside the lab and into the clinic. Our work aims to provide a more cost-effective and accessible approach to gait analysis in post-stroke patients, ultimately leading to improved rehabilitation outcomes. As a first step to achieve this goal, we develop a deep ensemble learning method to analyze a gait biomechanical dataset consisting of 63 post-stroke subjects and 47 age-matched controls. We have four tasks: 1) to automatically identify post-stroke population when compared to age-matched controls; 2) to automatically identify the more affected side of post-stroke subjects; 3) to reveal the biomechanical data signals best characterize post-stroke gait; and 4) to quantify the impact of ground reaction force (GF) on tasks 1-3. The goal of 1) and 2) is to test the ability of deep learning for automation. Goal 3) and 4) is to sift through the vast quantity of data extracted from a typical motion study to identify which data body location (trunk, arm, leg) and data type (kinematic, kinetic, force) are most pertinent in the above automation tasks. Our study has yielded promising results, demonstrating up to 99% accuracy in identifying post-stroke patients and age-matched controls and 100% accuracy in identifying more versus less affected sides. Furthermore, we have identified a subset of biomechanical signals that can achieve these tasks with high accuracy, both with and without force. These findings have promising implications for automating gait analysis and for conducting more extensive studies in clinical and subjects' natural environments.

1.1.6. Walk4Me: Telehealth Community Mobility Assessment, An Automated System for Early Diagnosis and Disease Progression. We introduce Walk4Me [**93**], a telehealth community mobility assessment system designed to facilitate early diagnosis, severity, and progression identification. Our system achieves this by 1) enabling early diagnosis, 2) identifying early indicators of clinical severity, and 3) quantifying and tracking the progression of the disease across the ambulatory phase of the disease. To accomplish this, we employ an Artificial Intelligence (AI)-based detection of gait characteristics in patients and typically developing peers. Our system remotely and in real-time collects data from device sensors (e.g., acceleration from a mobile device, etc.) using our novel Walk4Me API. Our web application extracts temporal/spatial gait characteristics and raw data signal characteristics and then employs traditional machine learning and deep learning techniques to identify patterns that can 1) identify patients with gait disturbances associated with disease, 2) describe the degree of mobility limitation, and 3) identify characteristics that

change over time with disease progression. We have identified several machine learning techniques that differentiate between patients and typically-developing subjects with 100% accuracy across the age range studied, and we have also identified corresponding temporal/spatial gait characteristics associated with each group. Our work demonstrates the potential of utilizing the latest advances in mobile device and machine learning technology to measure clinical outcomes regardless of the point of care, inform early clinical diagnosis and treatment decision-making, and monitor disease progression.

CHAPTER 2

An Overview of Human Activity Recognition Using Wearable Sensors: Healthcare and Artificial Intelligence

2.1. Introduction

Human activity recognition has been actively researched in the past decade, thanks to the increasing number of deployed smart devices such as smartphones and IoT devices. Based on the type of data being processed, a HAR system can be classified into vision-based and sensor-based. This chapter targets wearable-sensor HAR systems in healthcare, which are the most prevalent type of sensor-based HAR systems [22]. More importantly, wearable-sensor HAR systems do not suffer from severe privacy issues like vision-based HAR systems, making wearable-sensor HAR systems suitable for healthcare applications. In a wearable-sensor HAR system, a user wears portable mobile devices that have built-in sensors. The user's activities can then be classified by measuring and characterizing sensor signals when the user is conducting daily activities.

HAR for healthcare has many potential use cases, including (1) Moving gait diagnosis from expensive motion labs to the community. Gait analysis can be used in many healthcare applications, such as stroke detection, gait modification (to prevent failing), and certain disease early detection. (2) Cognitive behavior monitoring and intervention for children and adults with attention-deficit/hyperactivity disorder (ADHD). We can leverage sensors to investigate whether fidgeting positively or negatively affects attention. (3) Stroke-patient hospital direction. When a patient is in an ambulance, a life-and-death question is whether the patient has extensive brain hemorrhage. If so, the patient should be directed to a hospital that can treat such cases. UCSF has developed a device based on an accelerometer sensor to help make this critical decision. (4) Epilepsy and Parkinson's disease study. Doctors have collected a significant amount of data on electrophysiology and episodic memory in rodents and human patients. The analysis of such sensing data can be used for various disease identification and treatment purpose. (5) An expensive device, called Vision RT,

is used to ensure radiation therapy is delivered safely to cancer patients (due to patient motion). It is worth exploiting sensors to detect the patient's movement while taking radiation therapy for the less affluent communities.

However, building practical wearable-sensor HAR systems for healthcare applications not only has challenges (e.g., sensor setup, data collection, and AI model selection) that are faced by traditional wearable-HAR systems, but also challenges that are unique to the healthcare domain. For example, in addition to the overall AI model accuracy (averaging results of all users), clinicians are concerned about the model stability (i.e., the model has approximately the same accuracy for each user) and model interpretability (e.g., to discover patient movement patterns that are specific to some symptoms).

Therefore, we present this overview chapter in the hope to shed light on designing wearable-sensor HAR systems for healthcare applications. To illustrate the system considerations, we share two of our healthcare systems: one for identifying the early mobility activities of ICU patients [**69**] and the other one for the gait analysis of DMD patients [**96**]. Our projects demonstrate that HAR systems for healthcare not only have commonalities such as data processing pipelines but also differences in terms of sensor setup and system requirements.

We organize this chapter as follows. First, we provide the preliminaries of HAR systems. Next, we introduce our HAR systems for ICU patients and DMD patients. Then, we explain the considerations when designing a HAR system. Last, we highlight the challenges of applying wearable-sensor-based HAR systems to healthcare, and propose several research opportunities. Last, we conclude this chapter.

2.2. Human Activity Recognition: A Primer

Given the short time-length data of wearable sensors, a HAR system needs to recognize the activity from which the data is generated. Thanks to the rapid advancement of AI technology, AI algorithms/models are increasingly adopted for recognizing the activity from the sensor data. Figure 2.1 illustrates the general data flow for an AI-based HAR system, which can be divided into two stages: model training and model deployment.
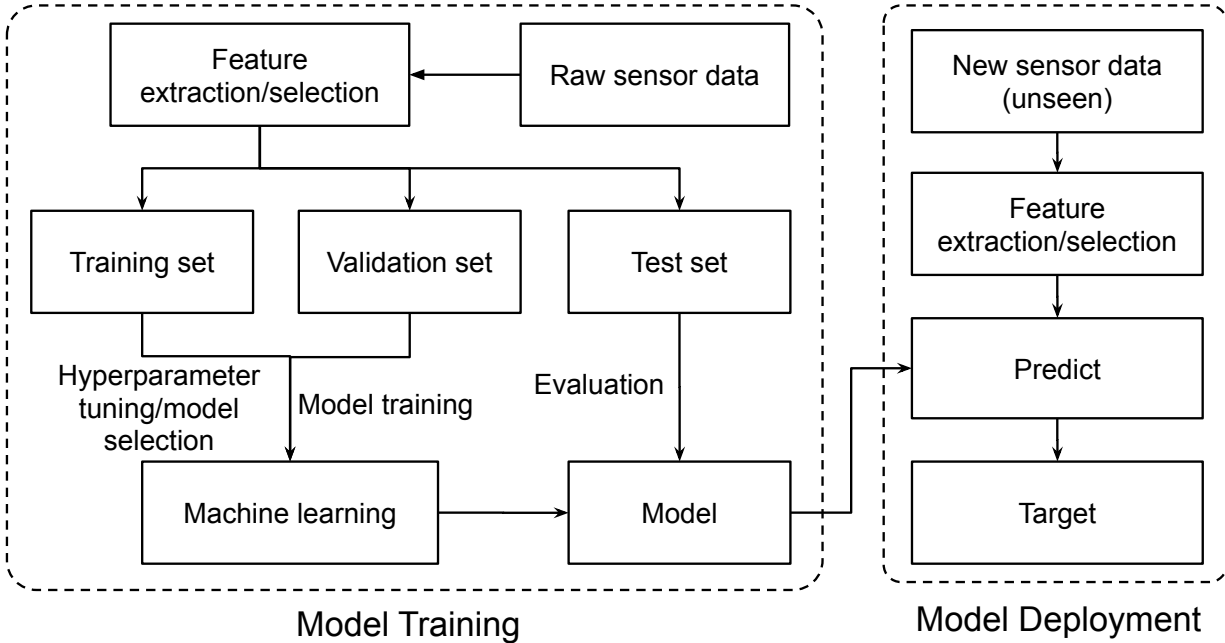
FIGURE 2.1. General data flow for the two-stages of HAR systems: model training and model deployment.

In the model training stage, an AI model is trained and tailored for the specific application. To achieve an accurate AI model, the following steps are often applied. First, raw sensor data from different activities should be collected. The quality of collected data significantly affects the AI model performance. The collected data is required to be diverse, representative, and large in the number of samples. Afterward, the raw data is divided into fixed-length or dynamic-length segments (i.e., time windows) [88]. Then, feature extraction is used to extract potentially useful features from the data segmentation, and feature selection is adopted to remove irrelevant features [56]. To alleviate the overfitting problem of the trained model, the set of processed features are divided into a training set, a validation set, and a test set. During the AI model training, we use the training set to tune the AI model and the validation set to measure the model's accuracy. After we finish the model training, we use the test set to evaluate the trained model. The trained model is deployed to real-world applications if its accuracy is satisfactory. Otherwise, the whole model training stage is performed repetitively by exploring different configurations, such as applying other feature extraction methods and changing AI models.

15

(a) ICU-patient   (b) DMD patient
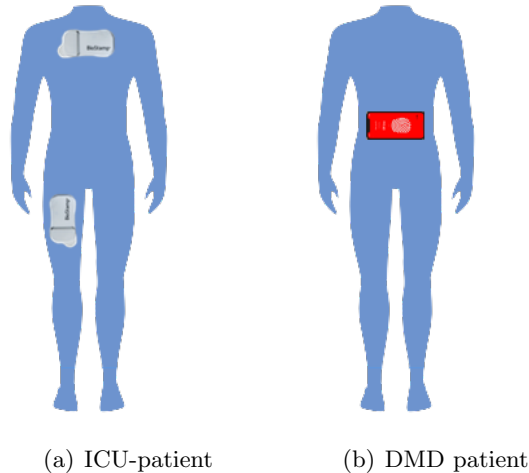
FIGURE 2.2. Device setups in our HAR projects. (a) We use two accelerometer devices to recognize the early mobility activities of ICU patients. One device is on the chest and the other device is on the thigh. (b) We use one smartphone that captures accelerometer data to identify DMD patients. The phone is located at the backside body.

In the model deployment stage, the same data processing (e.g., segmentation, feature extraction, and selection) is applied to the new and unseen sensor data, and the trained model is executed on the processed data. It is possible that the trained model may not work as expected in a real deployment, probably due to the model over-fitting or the lack of generality in the collected dataset [113]. In this situation, the system designer needs to revert to the model training stage.

2.3. HAR Applications in Healthcare

Clinicians have already applied wearable sensor-based HAR systems in healthcare, thanks to the development of more lightweight wearable devices, greater computation capability, and higher accurate AI algorithms. This section presents our two HAR healthcare projects to illustrate the considerations when designing HAR systems for healthcare applications with different goals.

2.3.1. Case 1: Identification of Early Mobility Activity for ICU Patients. Due to long periods of inactivity and immobilization, patients become weak when recovering from major illnesses in ICU [11]. If ICU patients' activities can be accurately recognized, clinicians can provide an optimal personalized dose of mobilities for ICU patients' different illness conditions. Therefore,

16

doctors and researchers are extremely interested in ICU patients' early mobilization, which is an effective and safe intervention to improve functional outcomes [2]. However, early mobility activity (EMA) research is limited by the lack of accurate, effective, and comprehensive methods to recognize patients' activities in ICU.

We propose a wearable sensor-based HAR system for recognizing the EMA of ICU patients [69]. In our system, Each ICU patient wears two accelerometer devices: one on the chest and the other on the thigh, as shown in Figure 2(a). Each device continuously collects 3-axis accelerometer data at a sampling rate of 32 Hz. Figure 3(a) plots the accelerometer data when an ICU patient sits on the cardiac chair to achieve an optimal resting position. This project aims to classify 20 types of ICU-related activities (e.g., reposition, percussion).

This project has two main challenges in designing the HAR system for ICU patients. (1) Label Noise. Because the time lengths for accomplishing an early mobility activity are different for ICU patients with varying health conditions, it is laborious and time-consuming work for clinicians to annotate sensor data for each second in the real world. Therefore, our EMA sensor data are annotated for each minute by a medical expert after data collection. However, one-minute length is exceedingly long for some early mobility activities such as Reposition, which the patient needs less than 20 seconds to accomplish. This annotation process introduces the label noise in our EMA dataset, which decreases the accuracy of the model. (2) Sensor Orientation. In the actual data collection process and possible future applications, we cannot guarantee that the orientations of all accelerometers are the same, and different orientations of the accelerometers lead to different meanings of XYZ coordinate values. Therefore, without careful feature extraction and selection, the AI model generalizes poorly to different patients, affecting the system performance in practice.

To tackle these challenges and improve the accuracy of recognizing ICU patient's activities, we explore the following techniques. (1) We propose a segment voting process to handle the label noise. Specifically, each one-minute sensor data is divided into multiple fixed half-overlapped sliding segments (time windows). We train our AI model using the segments. To predict each one-minute sensor data activity, we apply our trained model to each segment. The final prediction result for the one-minute data is the activity that has the majority vote among the prediction of all segments. Our segmenting method improves the model accuracy by ~4.08% and reduces
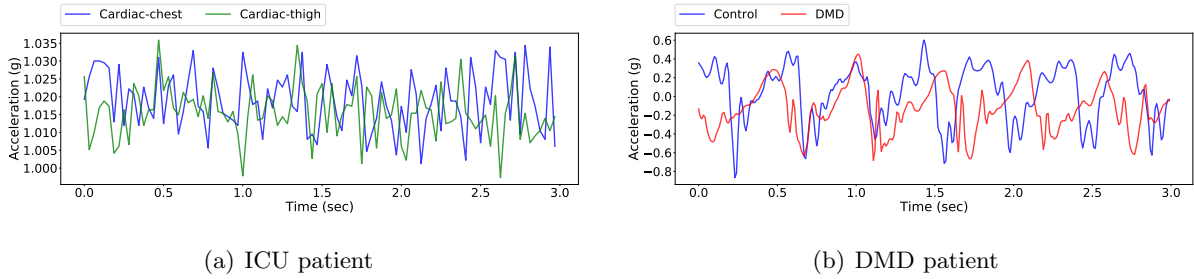
17

(a) ICU patient

(b) DMD patient

FIGURE 2.3. Illustration of accelerometer data in our projects. (a) The z-axis of the accelerometer data from the two on-body devices when an ICU patient is performing the cardiac activity. (b) The z-axis of the accelerometer data, which shows the difference in gait characteristics between a DMD patient and a healthy person.

the model instability by ∼9.77% [**69**]. Our experiments also demonstrate that the number of sensors contributes to eliminating label noise in our dataset. As shown in Figure 3(a), the increase in the number of sensors conveys more information, and thus improves the system's accuracy. (2) We identify and extract features that are not sensitive to sensor orientations to tackle the sensor orientation problem. Our features improve both the accuracy and the stability of AI models compared to the model trained on commonly used features.

2.3.2. Case 2: Identification of Gait Characteristics for DMD Patients. Duchenne muscular dystrophy (DMD) is a genetic disorder disease that affects the dystrophin protein, essential for keeping muscle cells intact. It has an estimated incidence of 1:5000 male births, and untreated boys become wheelchair-bound by the age of 12 years and die in their late teens to early 20s [**117**]. There is presently no cure for DMD disease. Nonetheless, gene repair interventions and other preventive therapies can be initiated as early as possible to slow the disease's progress and prevent secondary conditions. Therefore, it is important to identify children with DMD early in the course of their disease and have tools for quantitative evaluation of their gait in both the clinic and community environments.

We designed a wearable sensor-based HAR system to identify gait characteristics associated with the progression of gait abnormalities in children with DMD and to differentiate those patterns from those of typically developing peers [**96**] [**94**] To leverage this idea, we design a HAR system in which we use a smartphone to capture accelerometer data from the participants. As Figure 2(b) illustrates,

18

participants wear a smartphone at the back of the hips over the spine (lumbosacral junction) at a location that is the closest surface point to the body's center of mass. Each smartphone collects 3-axis accelerometer data at a sampling rate of 30 Hz with the same phone orientation.

We recruited ambulatory participants with DMD between 3 and 13 years of age and typically developing controls of similar ages. We ask participants to perform exercises at various times, speeds, and distances such as free walk and 6-minute walk, as specified by the north star ambulatory assessment (NSAA) standard [90]. Figure 3(b) shows the gait pattern difference between a DMD patient and a healthy person when they are walking.

We found that classical machine learning and deep learning, after hyper-parameter fine-tuning and cross-validation on seven different gait activities, led to the best performance with an accuracy exceeding 91% on the 6-min-walk-test activity [96]. We demonstrate that by using AI techniques and an accelerometer, we can distinguish between the DMD gait and typically developing peers.

There are two main challenges in designing our HAR system for the DMD application: clinical interpretability and data sparsity. (1) Clinical Interpretability. Medical practitioners desire not only a high prediction accuracy but also an interpretation of the prediction result. (2) Data Sparsity. In the healthcare domain, collecting diverse and sufficient data is challenging, especially for fatal diseases such as DMD.

We explore the following techniques to tackle these challenges. (1) To interpret AI model outcomes, we plan to link the clinical measurements with the model's extracted features by leveraging advanced AI models such as interpretable CNN [118]. However, it is an active, challenging task to find which clinical measurements correlated with the AI model features, especially for deep learning models. (2) To overcome the lack of data, we plan to use Generative Adversarial Network (GAN) [39] or synthetic minority over-sampling technique (SMOTE) [13] to generate more data samples.

2.3.3. Summary of Our Projects. Our two projects target different healthcare applications with different goals: recognizing ICU patients' activities and distinguishing DMD gait patterns from those typically developing controls. The ICU project focuses on the system performance to assist the doctor in better understanding patients' recovery. While achieving high system performance, the DMD project interprets the model results further and discovers disease-specific patterns to

determine the patient's condition and progression. Our example projects demonstrate the effectiveness and potential of wearable sensor-based HAR systems in healthcare. However, due to the different goals, different healthcare applications may have additional HAR system considerations. For example, our two projects adopt a different number of devices (2 versus 1) and device position (chest and thigh versus central mass body). In addition, our projects also apply different feature extractions (time and frequency domain versus clinical). In the next section, we present design considerations for building HAR systems.

2.4. System Design

This section covers three design considerations essential for HAR systems, i.e., sensor, feature extraction and selection, and AI model selection.

2.4.1. Sensor. Sensors play an essential role in wearable HAR systems. Different HAR systems adopt various sensor configurations regarding the type of sensors, the sensor position and orientation, and the number of sensors.

2.4.1.1. *Sensor Types.* There are several types of sensors. Each sensor captures a different raw movement signal. The most commonly-used wearable sensors in HAR systems are accelerometer, gyroscope, and electrocardiography (ECG). The accelerometer sensor captures the acceleration signal that is useful for recognizing movements such as walking, running, and jumping. Gyroscopes capture the rotation movements used commonly in recognizing swinging, turning, and repositioning. ECG captures the heart rate and rhythm, which helps distinguish between intensive and light exercises.

However, many activities include both directional and rotational movements. Therefore, using one sensor type is not adequate. As a result, multiple types of sensors (e.g., accelerometer and gyroscope) are used in various application scenarios to maximize accuracy. However, using multiple types of sensors is challenging due to the increased complexity of the system in terms of synchronization issues [62].

2.4.1.2. *Sensor Position and Orientation.* Different positions and orientations of devices affect the data features and thus the model accuracy in predicting different activities [74]. However, there have not yet been systematic comparisons of the number, type, and location of sensors to

determine whether an optimal array design can capture data across a wide range of human activities and disease states. In many cases, the device position and orientation are decided by the empirical experience of clinicians.

2.4.1.3. *Number of Sensors.* Generally, a large number of sensors require demanding storage and computation capability. On the other hand, more sensors can collect more diverse data, which is beneficial for improving model performance [**61**]. Therefore, to decide the optimal number of sensors, researchers need to carefully consider many factors such as cost, power consumption, and accuracy target as well as the feasibility of long-term use in the community to collect real-world information [**49**].

2.4.2. Feature Extraction and Selection. In addition to the hardware setup, feature extraction and selection significantly affect the overall system performance. Before applying feature engineering to the data, the input data needs to be segmented.

2.4.2.1. *Data Segmentation.* HAR systems collect data constantly via wearable sensors to identify possible activities. Data segmentation is applied to divide comparatively long time data into short fragments (time windows) that are suitable for AI models to learn. There are two types of data segmentation: fixed-length and dynamic-length [**88**]. For fixed-length segmentation, if the time window is too short, the extracted features from the fragments are insufficient to capture the activity; on the other hand, if the time window is too long, a fragment is likely to contain multiple activities. The system accuracy deteriorates in both cases. In comparison, a dynamic-length data segmentation adopts an adaptive length of fragments corresponding to the characteristics of input data. Ideally, dynamic data segmentation generates fragments, in which each fragment only contains a single and complete activity. However, dynamic data segmentation is much more complex than fixed data segmentation, and thus are not as widely adopted by existing works as fixed-length segmentation.

2.4.2.2. *Feature Extraction.* Feature extraction is then applied to extract important features from the data fragments [**56**]. It can be broadly classified into time-domain and frequency-domain methods. In time-domain feature extraction, metrics such as median, variance, mean, and skewness are calculated over the amplitude variations of data over time. Time-domain features are lightweight to compute and thus are friendly to low-profile embedded devices and real-time applications. In
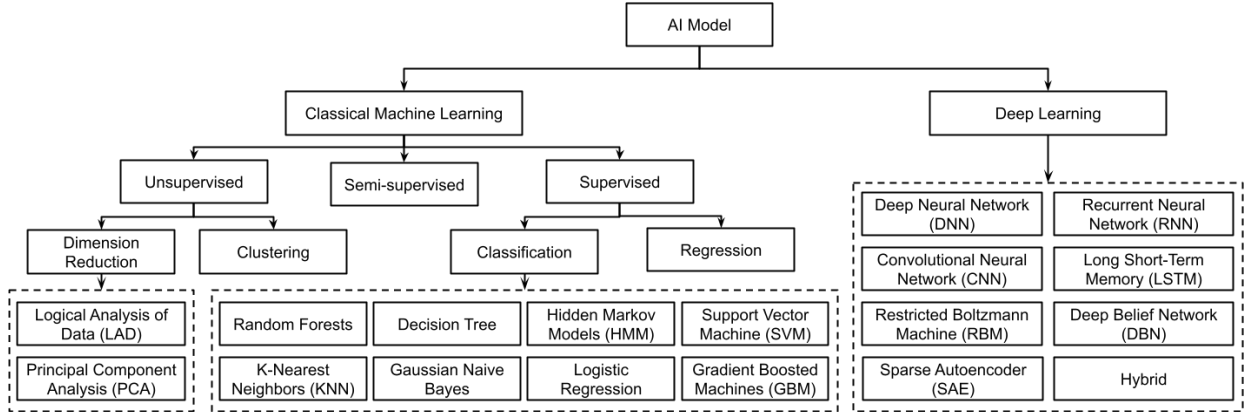
FIGURE 2.4. Classical machine learning and deep learning algorithms used in HAR systems.

comparison, frequency-domain features calculate the frequency variations of data over time. They include metrics such as spectral entropy, spectral power, and peak frequency. The computation overhead of frequency-domain features is generally much greater than time-domain features. In reality, most existing HAR systems adopt both time-domain features and frequency-domain features, in the consideration of the tradeoff among factors such as system accuracy, computation overhead, and power consumption.

2.4.2.3. *Feature Selection.* Feature selection is often adopted in order to reduce system complexity. It measures the importance of features and then removes irrelevant features. Feature selection is roughly divided into three methods: filter methods, wrapper methods, and embedded/hybrid methods [**56**]. Filter methods select a subset of features by exploiting inherent characteristics of features, whereas wrapper methods use classifiers to estimate the useful features. On the other hand, the embedded/hybrid methods combine the results from filter methods and wrapper methods [**22**]. By carefully selecting features, the AI model accuracy can be significantly improved. However, in healthcare HAR systems, pursuing high accuracy is not the sole goal, as the features are often manually decided by medical experts for identifying patients. Therefore, healthcare HAR systems require feature extraction and selection that is meaningful for clinicians and meanwhile achieves high prediction accuracy.

2.4.3. AI Model Selection. In the HAR field, classical machine learning algorithms and deep learning algorithms have been explored and applied, which is summarized in Figure 2.4. Both

classical machine learning algorithms and deep learning algorithms have different advantages and disadvantages.

Dataset requirement and system running overhead. The data collection process in the healthcare scenario is challenging because of the severe privacy issue and rare incidence rate of some medical activities. Therefore, in most healthcare applications, the database size is small. Correspondingly, classical machine learning models are more preferred because they work well with medium-size datasets. In contrast, even though deep learning models achieve better accuracy, they usually require a large amount of data for training. Real-time performance is another critical factor for some healthcare applications [**91**]. For example, [**55**] uses cranial accelerometers to detect stroke in an ambulance to decide whether to send the patient to a specialist stroke hospital for special treatment. Therefore, lightweight models are preferred in this use case. In addition to the running overhead of the AI models, the processing time of feature extraction also affects the model selection, because different model structures adapt differently to the extracted features.

System interpretability. The features extracted from the sensor data are helpful to understand the pattern of some specific diseases to find out the pathological characteristics of the disease. For example, we extract the temporal/spatial gait characteristics from sensor data to evaluate the gait changes associated with DMD. Classical machine learning models are easier to interpret the model's decision, especially in decision tree models. Even though there is a great deal of work in interpreting deep learning models, deep learning models have the reputation of poor interpretability.

2.5. Challenges and Opportunities

Wearable sensor-based HAR systems are promising for a variety of healthcare problems. However, there are several challenges in fully exploiting them to build satisfactory HAR systems for healthcare. In this section, we identify challenges as well as research opportunities of HAR systems for healthcare.

2.5.1. Data Sparsity. The most commonly used algorithms for the HAR system in healthcare are the supervised learning algorithms that need extensive labeled data. For some daily living activities such as walking and running, researchers could get a significant amount of the labeled

23

data from the public dataset or the raw sensor data collected and annotated by themselves. However, for some specific human activities related to healthcare, such as the therapeutic activities of patients, researchers could not get enough sensor data since these activities are low-probability events compared with daily life activities. Furthermore, it also takes time and effort to locate the sensor data of these specific activities from the daily data and label them. For example, when patients recover from surgery, they need some range of motion(ROM) exercises several times a day to make their joints and muscles flexible and strong again. Because of the fixed and limited collection times per day and the limited number of patients are involved, raw sensor data for ROM becomes insufficient, affecting the HAR system's performance. Therefore, building HAR systems with high accuracy on small datasets in healthcare is one of the most significant challenges.

Meta-learning is one of the approaches to solve this challenge. Meta-learning aims to optimize models which can learn efficiently in a small dataset when dealing with new categories. In [115], researchers present a meta-learning methodology based on the Model-Agnostic Meta-Learning algorithm [33] to build personal HAR models. In [30], researchers use few-shot learning to transfer information from existing activity recognition models. However, it is unclear whether these techniques work well for medical applications. So more research is needed to explore the feasibility of transferring knowledge from daily living activities to specific activities related to healthcare.

2.5.2. Model Interpretability. In HAR applications in healthcare, an increasing number of applications focus on the interpretability of the model to extract relevant features, in order to describe the severity of the disease and track the progression of the disease [96]. In addition, notwithstanding the benefit of deep learning in HAR, the underlying mechanics of machine learning are still unclear. So, various studies are trying to explain the deep learning model for the recognition of human activities. The common approach to interpreting the deep learning model is to compute the importance of each part of the input. In [59], researchers propose an interpretable convolutional neural network to select the most important sensor position for some specific activities. Instead of computing the importance of each part of the input, another approach is to make a sequence of selections about which part of the input is essential for the model training [15]. More research is required to adopt these methods to HAR systems for healthcare.

2.5.3. Concurrent Activities. Most of the existing HAR research focuses on single-labeled activity, recognizing only one activity of the given data segment. However, in real-world healthcare scenarios, humans can perform multiple activities concurrently. For example, patients can do ROM exercises and percussion therapy at the same time. The AI model performance deteriorates for concurrent activities. On the other hand, designing models to recognize multiple activities per data segment is a challenging task.

2.5.4. Composite Activities. In healthcare applications, optimizing HAR algorithms to identify composite activities in the community is ultimately more desirable than recognizing a single type of task. For example, when a patient moves from bed to the chair, the patient performs various activities, including sitting from supine in the bed, pivoting to place feet on the floor, standing from sitting, walking a few steps, and then sitting down on a chair. Therefore, it is preferred that an AI model can directly recognize the composite activity.

2.5.5. Privacy. Wearable sensor-based HAR systems do not suffer from severe privacy issues as camera-based vision systems. However, since HAR applications continuously capture user data and recognize user activities, they may leak users' personal information if data are not secured. Therefore, secure data sharing and safe data storage are imperative for healthcare applications. To alleviate sensitive information during model training, adversarial loss functions are leveraged to guard against privacy leakage [48]. In addition, federated learning is a promising solution, which trains a global model without exposing local devices' private data [78].

2.5.6. Opportunities of HAR for Healthcare. Through our experience with HAR systems for healthcare, we identify the following research opportunities.

- **Community-based healthcare**. Community-based healthcare requires that user devices are lightweight and affordable for the public. In addition, instructing the non-expert users/patients should be straightforward to follow. We can use digital sensing capability and the popularity of mobile devices to enable large community-based prescreening for various diseases and early signs of diseases. This can be done in a privacy-preserving manner in the sense that data does not need to leave a local device if necessary. For

example, our DMD project enables community-based diagnosis during the pandemic and in rural areas where specialty labs are hundreds of miles away.

- **Chronic disease prevention and intervention.** For chronic diseases, it is essential to capture the behaviors of patients in the long run. To this end, gait analysis, motion monitoring, ECG, and other vital signals (such as continuous glucose monitoring) can play a key role.

- **Health aging**. With the decreased fertility rates and the increased life expectancy, population aging is becoming common for most countries. Therefore, building HAR systems for healthy aging is beneficial for senior citizens and society as a whole. We anticipate that gait and motion monitoring and diagnosis will play a critical role in healthy aging.

2.6. Conclusion

It is gaining popularity by applying wearable sensors to recognize and analyze human activities for the healthcare domain. For example, we leverage HAR systems to recognizing patients' early mobility activities in ICU and to analyzing the symptoms of DMD patients. This overview chapter covers the system design of HAR systems based on wearable sensors, focusing on healthcare applications. We emphasize the essential components of HAR systems, including sensor factors, data segmentation, feature extraction and selection, and AI model comparison. We also highlight the challenges and opportunities of HAR systems for healthcare.

CHAPTER 3

Automated Detection of Gait Events and Travel Distance Using Waist-worn Accelerometers Across a Typical Range of Walking and Running Speeds

3.1. Introduction

Research indicates that accelerometers are more accurate than pedometers, especially at slower walking speeds and in populations with atypical gait patterns, making pedometers less suitable for evaluating physical activity in said populations [65]. Estimating temporospatial clinical features (CFs) of gait, such as step length, step duration, step frequency, and gait speed, is a fundamental step in gait analysis, and detecting the initial contact (IC) of the heel is crucial for identifying gait events and the beginning of the step cycle. Estimating CFs, such as step length, step duration, step frequency, and gait speed, is a fundamental step in gait analysis. Additionally, detecting the IC of the heel is crucial for identifying gait events and the beginning of the step cycle. In a laboratory environment, detecting events and estimating CFs is relatively straightforward, typically done by measuring ground reaction forces (GRF) and verifying with visual observation. However, using these methods to measure gait events in the community is not feasible or practical.

Research has also investigated the potential of using acceleration signals to estimate CFs. Several studies have demonstrated that step length, gait speed, IC, and incline can be determined from acceleration signals of the lower trunk [119]. Aminian and colleagues explored the feasibility of using a fully connected artificial neural network (ANN) with accelerometers placed on the trunk and heel to predict incline and speed based on ten statistical parameters extracted from the raw signal [6]. The results revealed that a negative peak in the heel accelerometer signal indicates the IC events in each gait cycle (two steps).

Studies comparing accelerometer signals from different body positions during various walking speeds have shown that positions near the body's center of mass, such as the trunk, waist, pelvis, and

27

sacrum, are suitable for capturing gait events [38,54,57]. For instance, in a study by Zijlstra et al., participants walked on a force-transducing treadmill and overground while their trunk acceleration data was recorded to estimate step lengths and walking speed. The IC events were matched with vertical GRF normalized by body weight to anteroposterior acceleration. The study found that the start and end of gait cycles indicated by the GRF corresponded with the time of the peak amplitude value in the anteroposterior acceleration signal [119]. Further research by Lee et al. and Mo et al. demonstrated that IC events can be determined from anteroposterior acceleration measured at the level of the pelvis and sacrum [67,81]. These studies collected simultaneous raw accelerometer signals from the pelvis/sacrum and GRF data, and the IC events on anteroposterior acceleration and vertical GRF were matched. Initial contact events on the force plate corresponded with the instant of the positive peak of pelvis/sacrum anteroposterior acceleration [81]. These studies collected simultaneous raw accelerometer signals from the pelvis/sacrum and GRF data, and showed that the initial contact events on the force plate corresponded with the instant of the positive peak of pelvis/sacrum anteroposterior acceleration.

In this study, we present a machine learning (ML)-based method that automates the detection of IC events and CFs extracted from raw accelerometer signals obtained from consumer mobile devices [69,70,91]. Our findings demonstrate that using a single accelerometer worn close to the body's center of mass is an accurate and reliable approach to estimate CFs and IC events across a wide range of typical walking speeds. Importantly, this method can be applied to both healthy individuals and those with gait disturbances without the need for GRF measurements.

3.2. Materials and Methods

Estimating distance using accelerometer signals is challenging due to the inherent quadratic error of accelerometers, which can result in deteriorating estimates even with short integration times and distances. Many existing methods attempt to estimate the distance from accelerometers by integrating acceleration twice with respect to time, despite incorporating error-limiting mechanisms and setting restrictions, which can result in errors due to noise, drift, and bias [5]. In this work, we propose an ML-based signal processing method that accurately estimates an individual's

distance traveled, step length, and the number of steps across varying walking/running speeds, outperforming the built-in pedometer [67].

To account for the fact that different individuals have different walking/running behaviors that heavily affect acceleration, we built a regression model for each individual to accurately estimate the distance based on their specific walking/running patterns. The regression model uses data from five different speeds per individual, mapping step length to the corresponding acceleration amplitude value. We created a regression model using data from five different speeds per individual, aiming to map step length to the corresponding acceleration amplitude value. For each of the five different speeds (SC-L1 to SC-L5), distance and corresponding acceleration were collected from each speed (Figure 3.1A). The distance for a single speed was calculated by averaging the step distances, while the acceleration was calculated by averaging the maximum values of acceleration in each step (Figure 3.1B).

To ensure a fair comparison, we compared three sources of estimated data: first, ground-truth data based on the estimator observing the individual, measuring the distance the participant took in the scene, and calculating the number of steps visually with the help of a video recorder; second, the pedometer sensor in the iPhone, providing distance and number of step estimates; and third, our Walk4Me system [92, 93, 94], which includes calibration regression models for estimating distance and a signal processing algorithm for measuring number of steps. We estimate the speed, step length, and frequency as derivatives from the regression and signal processing.

3.2.1. Participants. Fifteen children with Duchenne muscular dystrophy (DMD) and fifteen typically developing (TD) peers participated in gait speed experiments. The age of the participants ranged from 3 to 16 years, with a mean age of 8.6 years and a standard deviation of 3.5. Their body weight ranged from 17.2 to 101 kg, with a mean weight of 36 kg and a standard deviation of 18.8. Their height ranged from 101.6 to 165.5 cm, with a mean height of 129 cm and a standard deviation of 15.8. All participants had at least 6 months of walking experience and were able to perform a 10-meter walk/jog/run test in less than 10 seconds. Participants with DMD had a confirmed clinical diagnosis and were either naïve to glucocorticoid therapy or on a stable regimen for at least three months. Northstar Ambulatory Assessment (NSAA) scores for DMD participants ranged from 34 to 8, indicating typical levels of function to clinically-apparent moderate mobility limitation

29
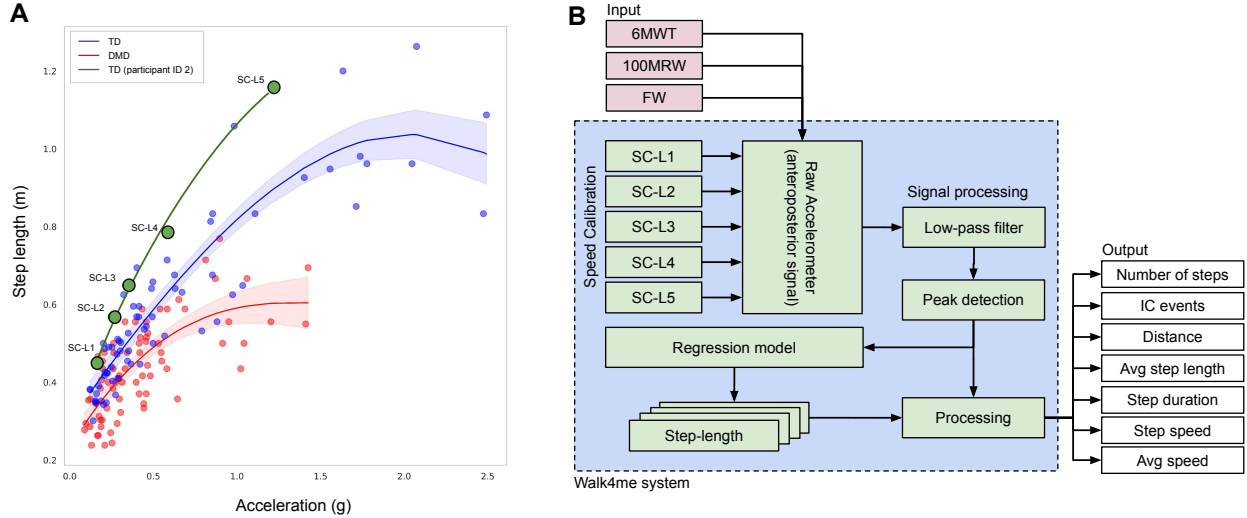
FIGURE 3.1. (A) The relationship between step length and central body mass's acceleration at various speeds for individuals with TD and those with DMD. The plotted curves depict the regression model, with the black line representing all participants, the green line representing TD participants, and the red line representing DMD participants. (B) The diagram depicts the data flow of our model training and prediction process. In the training phase, the model uses five-speed calibration SC-L1 to SC-L5 with ground-true to predict the average step length. The input to our model is the acceleration signal from unseen gait activities (6MWT, 100MRW, and FW). Our model processes these signals to estimate the average step length.

(Table-3.1). The protocol was reviewed and approved by the Institutional Review Board (IRB) at the University of California, Davis and informed consent was obtained from each participant prior to the initiation of study procedures. Measurements were taken at eight different walking/running gait activities, including speed-calibration tests at slow walk to running speeds (SC-L1, SC-L2, SC-L3, SC-L4, and SC-L5), a 6-minute walk test (6MWT), a 100-meter fast-walk/jog/run (100MRW), and a self-selected pace free walk (FW).

3.2.2. Equipment. Acceleration data from each participant were sampled at a rate of 100 Hz using an iPhone 11 and our Walk4Me smartphone application [94, 96]. The phones were securely attached at the waist with an athletic-style elastic belt enclosure, positioned approximately at the level of the lumbosacral junction. The raw accelerometer signal was synchronized with video recordings captured by a GoPro camera at a rate of 30 Hz. An observer marked the events where a participant passed the start and end of the duration or distance assigned to each activity using the web portal of the Walk4Me system.

30

TABLE 3.1. Characteristic of the children included in the study.

ID	Case (status)	Age (years)	Weight (kg)	Height (cm)	NSAA (/34)
1	TD	15	101.0	165.5	34
2	TD	12	57.7	155.6	34
3	TD	11	46.4	146.0	34
4	TD	9	41.8	132.9	34
5	TD	8	29.6	126.5	34
6	TD	8	27.9	136.9	34
7	TD	7	32.8	139.0	34
8	TD	7	25.2	131.1	34
9	TD	7	22.7	122.0	34
10	TD	6	27.2	127.1	34
11	TD	6	20.1	114.8	34
12	TD	5	20.8	119.8	34
13	TD	4	18.7	114.0	34
14	TD	4	18.6	108.5	34
15	TD	6	23.2	122.4	31
16	DMD	3	20.0	101.6	34
17	DMD	9	36.2	128.4	31
18	DMD	6	17.2	106.2	30
19	DMD	14	50.0	147.0	27
20	DMD	10	54.5	142.0	24
21	DMD	10	31.7	131.5	24
22	DMD	5	22.9	111.8	24
23	DMD	12	40.0	129.0	22
24	DMD	11	67.7	145.0	20
25	DMD	15	63.7	153.3	15
26	DMD	11	37.5	125.0	15
27	DMD	8	30.7	133.0	13
28	DMD	7	28.5	120.4	12
29	DMD	16	46.7	130.1	9
30	DMD	5	18.2	102.5	8

3.2.3. Gait and Events Detection and Data Analysis. We collected the raw accelerometer signal from 30 participants, which included the x, y, and z axes (vertical, mediolateral, and anteroposterior), along with the corresponding timestamps. Based on the findings of Zijlstra [**119**], we observed that the IC events were more distinguishable in the anteroposterior axis (z-axis) compared to the other axes. Therefore, we used the anteroposterior signal from the raw accelerometer

data to develop our method for counting the number of steps, estimating step length, and calculating the total distance individuals walked at different speeds.

3.2.3.1. *Method of Step Detection.* Figure 3.2A presents a raw accelerometer signal of the anteroposterior movement (z-axis) from a typically developing (TD) participant during fast walk speed calibration (SC-L4) for 3.9 seconds. The steps in the anteroposterior signal are characterized by long wavelengths (low frequency), while other wavelengths (high frequency) represent noise signals. To extract the steps, we applied a low-pass filter (first-order Butterworth) to the signal to smooth the signal and remove short-term fluctuations while preserving the longer-term trend (Figure 3.2A and Figure 3.2B). We then identified the peak values of the filtered signal, as the peaks occur only once per step in the filtered signal (Figure 3.2C). The number of peaks corresponds to the observed number of steps taken by the participant. Figure 3.3A shows the estimated number of steps using our method as blue dots, compared with the ground truth represented by a black line. The built-in pedometer steps estimation is shown in red.

3.2.3.2. *Method of IC Detection.* To detect the IC events, we find the midpoint boundaries between two peaks in the filtered signal (Figure 3.2D), which corresponds to the toe-off (TO) events during the gait cycle verified by observation. We then identify all the peaks that occur within each step duration in the original acceleration signal (Figure 3.2E). Next, we determine the maximum peak value (anteroposterior G's), which corresponds to the time point of each IC (Figure 3.2F). Using this method, we can identify the IC of every single step and estimate the step duration (Figure 3.2F) without the need to use GRF [119].

3.2.3.3. *Method of Step Length Estimation Using Regression.* We create an individualized regression model for each participant to associate average peak acceleration values with step lengths. Figure 3.1A depicts the data flow of our model training and prediction process. Each regression model with a polynomial of 3rd order degree is fitted using five different participant-selected calibration speeds (SC-L1 to SC-L5). For each speed, we calculate the average acceleration peak values by taking the mean of all the peaks as described in Section 3.2.3.2. To calculate the average step length for training, we divide the observed ground-truth distance by the number of steps obtained as in Section 3.2.3.1. This process is repeated for each of the five calibration speeds (e.g., point SC-L4 in Figure 3.1A)). The resulting individualized equation through all five points allows us to
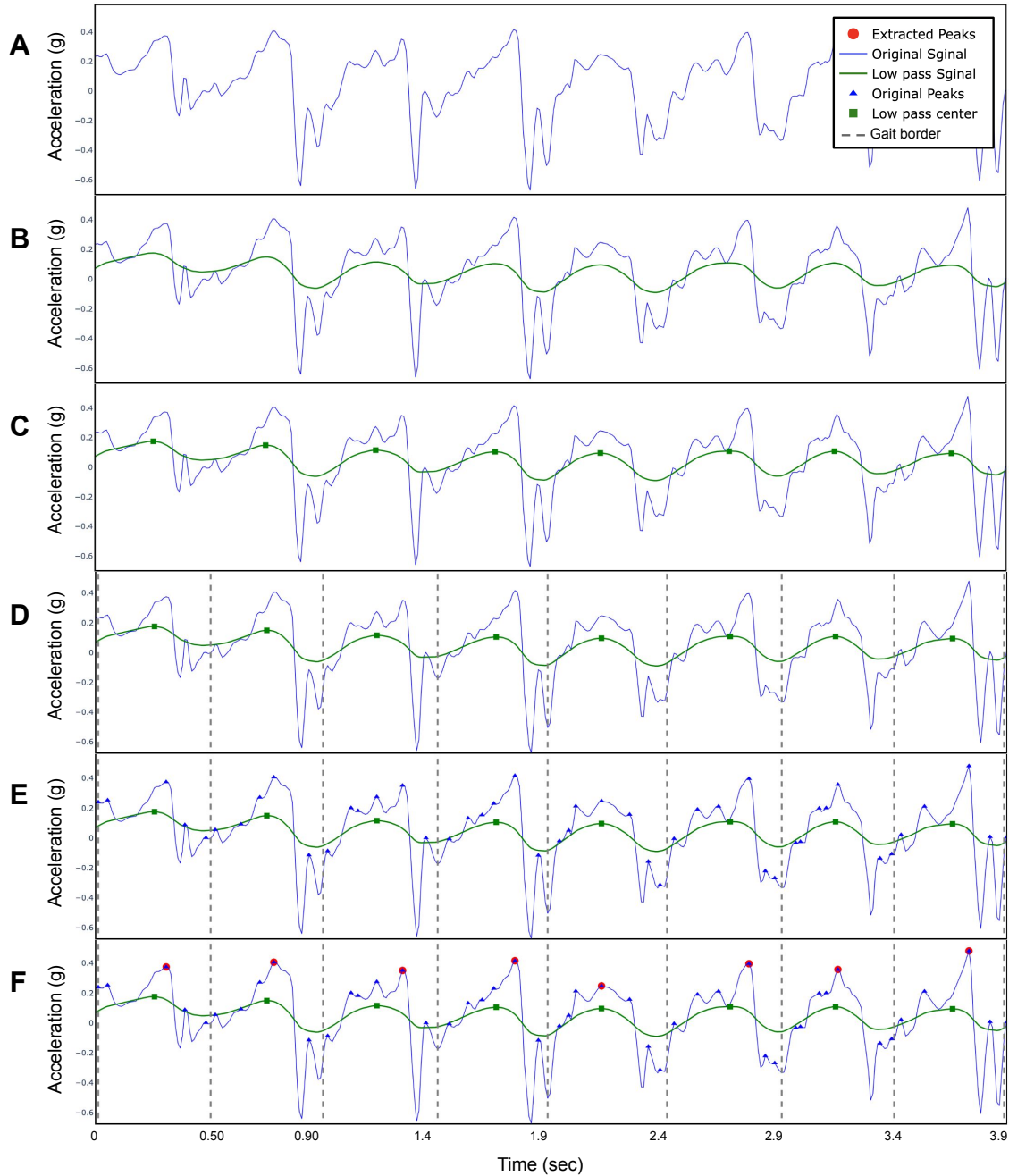
FIGURE 3.2. This figure presents the signal processing of the raw accelerometer signal of the anteroposterior movement (z-axis) of participant ID 2 on fast walk speed calibration (SC-L4) for 3.9 seconds. (A) Original raw accelerometer signal. (B) Filtered signal. (C) Peaks detection of the filtered signal. (D) Locate the being and the end of each step. (E) Peaks detection of the original signal. (F) Locate the highest peak in the original signal.
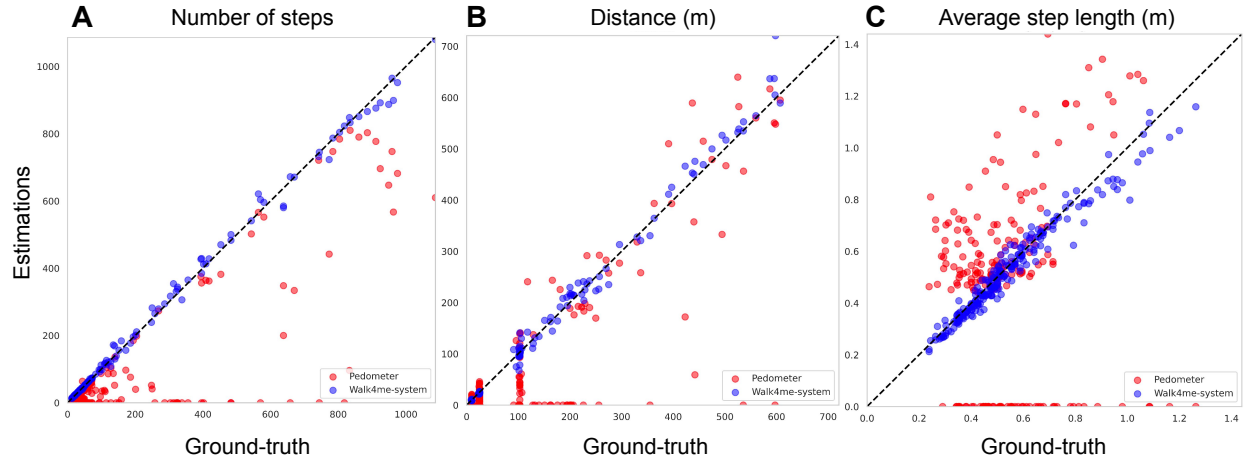
FIGURE 3.3. Comparison between the estimates and ground-truth values for both pedometers and our Walk4Me system to illustrate the accuracy across four key metrics: (A) the number of steps. The adjusted R-squared of our Walk4Me is 0.9973, while the pedometer is 0.6826. (B) the distance in meters. The adjusted R-squared of our Walk4Me is 0.9937, while the pedometer is 0.6987. (C) the average step length in meters. The adjusted R-squared of our Walk4Me is 0.9595, while the pedometer is 0.0094

input the peak acceleration value of any step within the participant's range of ambulatory velocity to estimate that step's length given a peak acceleration value of a step of the individual (shown as the green line in Figure 3.1A)).

3.2.3.4. *Estimating the Distance.* After establishing the individualized model, it can be used to estimate the distance of unseen data of an individual. Using the regression model, the step lengths are estimated for all identified steps of an individual from the forward acceleration of unseen walking activity (Figure 3.1A). These step lengths are then accumulated to calculate the total distance traveled by the individual for that unseen walking activity. In this project, we used 100MRW, 6MWT, and FW as unseen input activities during the inference stage, as shown in Figure 3.3B, and compared the calculated distances with the ground-truth observed distances and the device's internal pedometer.

3.2.3.5. *Calculating the Average Step Length.* During the inference stage, the average step length of an individual is calculated by dividing the distance estimated as in Section 3.2.3.4 by the number of steps obtained as in Section 3.2.3.1. Figure 3.3C shows the estimated average step length using our regression model as blue dots, compared with the ground-truth average step length
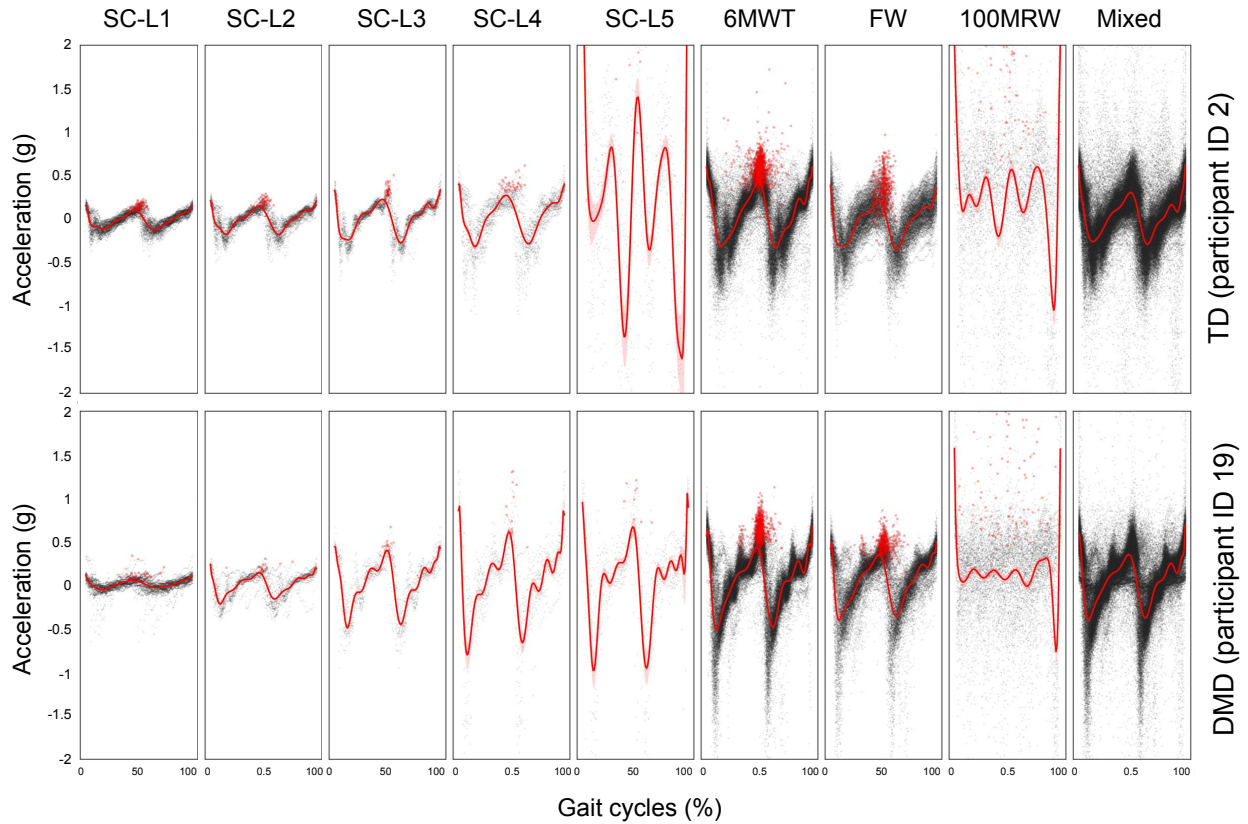
FIGURE 3.4. The gait pattern comparison of an individual with TD and an individual with DMD at varying speeds and a combination of multiple speeds and the effects of different speeds on gait pattern.

represented by a black line. The red dots represent the average step length estimated by the built-in pedometer.

3.2.3.6. *Gait Pattern Representation.* After determining the midpoint boundaries between steps, we generate a composite map of each step normalized to the gait cycle percentage, allowing for visual examination of AI-determined steps for irregularities or comparison of averaged accelerometer patterns between individuals (Figure 3.4). The gait cycle is identified using peak detection at the IC event, marking the beginning and end of each step. The average acceleration patterns are also calculated from all gait cycles across all activities and at various speeds. The forward movement (x-axis) is normalized to a time scale of 0 to 100%. By comparing the gait cycles of two participants (TD and DMD peers) at various speeds, distinct different patterns of acceleration magnitude emerge (Figure 3.4), highlighting differences in gait patterns between the two participants.

35

3.2.3.7. *Error Percentage Rates.* To compare observed ground-truth step counts, distance traveled, and average step lengths with our model's estimates and the pedometer estimates native to the mobile devices, we employed two methods. First, we calculated the aggregated error for all estimates by determining an error percentage rate ($Error_{rate}$) using equation 3.1.

$$(3.1) \qquad Error_{rate} = \left| \frac{\sum\limits_{i}^{n} |V_c - V_o|_i - \sum\limits_{i}^{n} |V_o|_i}{\sum\limits_{i}^{n} |V_o|_i} \right| \times 100$$

The $Error_{rate}$ is calculated by aggregating the residual values of all participants (i) for all activities. The residual is the difference between the proposed methods (V_c) and the total number of ground truth observations (V_o). Then, the total aggregated is subtracted from the total ground truth and divided by the total ground truth. Supplemental Table-3.2 compares the error percentage rate of step count, distance, and average step length between our Walk4Me system and iPhone pedometer measurements.

Second, to evaluate the percentage error for each individual measurement and estimate pair, we subtracted the model estimate from the observed ground truth measure and divided it by the ground truth measure multiplied by 100 for each event. We computed mean (SD) percentage error for step count, distance traveled, and step length parameters for calibration events SC-L1 to SC-L5 combined, and separately for 6MWT, 100MRW, and FW efforts combined, as well as for all efforts combined. We compared the mean percentage error values between control participants and those with DMD using simple t-tests for each contrast.

3.3. Results

In this study, we assessed the accuracy of step counts during walking, jogging, and running using our Walk4Me system compared to the iPhone pedometer. We validated our results by comparing both systems with ground-truth data. Our findings, as shown in Table-3.2, indicate that the Walk4Me system had an average step count error rate of 3.46%, demonstrating reliable performance in accurately tracking steps at different speeds. The combined error rates from participants with

Duchenne muscular dystrophy (DMD) and typically developing (TD) participants ranged from 1.26% during slow walk pace (SC-L2) to 7.26% during the fast 100m run. In contrast, the iPhone's built-in pedometer showed an average error rate of 48.46% during short- to moderate-distance tasks at varying gait velocities. The iPhone pedometer had the lowest error rate of 36.35% during the longer-duration fast walk 6MWT task, and the highest error rate of 85.26% during the short-duration jogging/running task SC-L5.

For distance measurement estimates, our Walk4Me system showed an average error rate of 5.83%, with the lowest error rate of 3.9% during the fast walk SC-L4 pace, and the highest error rate of 7.74% during the fast 100m run. The iPhone's built-in pedometer had an average error rate of 42.23%, with task-specific error ranging from 27.42% during the 6MWT to 82.54% during SC-L5 jogging/running task.

For step length measurement, our Walk4Me system showed an average error rate of 5.80%, with the lowest error rate of 3.68% at a comfortable walking pace (SC-L3), and the highest error rate of 8.64% during the short-term jog/run SC-L5 task. The iPhone's built-in pedometer demonstrated an average error rate of 46.40%, which varied from 30.76% during SC-L5 to 76.10% during SC-L1.

In contrast to overall aggregate accuracy, the mean (SD) accuracy of model predictions for individual events compared to ground truth observations for step counts, distance traveled, and step lengths is presented in Table-3.3 and depicted in Figure 3.3A, Figure 3.3B, and Figure 3.3C. Predicted and observed values for all three parameters showed a strong correlation (Pearson's r = -0.9929 to 0.9986, p¡0.0001). The estimates demonstrated a mean (SD) percentage error of 1.49% (7.04%) for step counts, 1.18% (9.91%) for distance traveled, and 0.37% (7.52%) for step length compared to ground truth observations for the combined 6MWT, 100MRW, and FW tasks. There were no statistically significant differences in mean error percentages between control participants and those with DMD (data not shown).

3.4. Discussion

The use of travel distance and step length as gait metrics is essential for clinical gait assessment in the community setting. However, accurately measuring step length traditionally requires a clinical facility or gait lab with a trained observer present during the assessment session. Clinical

TABLE 3.2. The table compares the error percentage rate of the step count, distance, and average step length between our Walk4me system and iPhone pedometer measurements.

Set	Activities	Case	Number of steps			Total Distance			Avg Step length	
			GT* (steps)	Sys.** (error %)	Ped.*** (error %)	GT* (meters)	Sys.** (error %)	Ped.*** (error %)	Sys.** (error %)	Ped.*** (error %)
Training set	SC-L1	TD	976.0	1.64%	47.44%	359.30	5.37%	54.97%	6.43%	76.93%
		DMD	506.0	1.58%	57.31%	149.90	4.64%	65.7%	6.03%	75.07%
		All	1482.0	1.62%	50.81%	509.20	5.16%	58.13%	6.25%	76.1%
	SC-L2	TD	782.0	1.41%	40.15%	359.80	4.48%	42.2%	5.38%	36.22%
		DMD	410.0	0.98%	49.02%	149.55	6.6%	39.82%	6.28%	59.94%
		All	1192.0	1.26%	43.2%	509.35	5.1%	41.5%	5.79%	46.88%
	SC-L3	TD	670.0	1.64%	49.4%	358.00	3.29%	50.66%	3.03%	34.39%
		DMD	347.0	1.15%	51.01%	149.80	3.79%	45.11%	4.47%	32.27%
		All	1017.0	1.47%	49.95%	507.80	3.44%	49.03%	3.68%	33.44%
	SC-L4	TD	562.0	1.96%	63.88%	358.70	2.93%	62.57%	2.86%	44.15%
		DMD	305.0	1.97%	69.84%	150.10	3.82%	62.64%	4.91%	35.34%
		All	867.0	1.96%	65.97%	508.80	3.19%	62.59%	3.76%	40.28%
	SC-L5	TD	373.0	3.75%	88.2%	352.90	5.63%	87.49%	9.16%	34.32%
		DMD	326.0	1.84%	81.9%	178.90	6.38%	72.78%	7.74%	24.62%
		All	699.0	2.86%	85.26%	531.80	5.88%	82.54%	8.64%	30.76%
Test set	6MWT	TD	10934.0	2.32%	35.44%	7745.50	5.02%	27.92%	4.87%	34.62%
		DMD	7478.0	3.81%	37.68%	4949.00	5.48%	26.63%	5.85%	48.73%
		All	18412.0	2.93%	36.35%	12694.50	5.2%	27.42%	5.24%	39.87%
	100MRW	TD	1729.0	8.68%	62.81%	1543.50	10.95%	57.67%	6.97%	55.9%
		DMD	1906.0	5.98%	56.24%	1134.00	3.38%	50.41%	5.0%	54.62%
		All	3635.0	7.26%	59.37%	2677.50	7.74%	54.59%	6.4%	55.53%
	FW	TD	4460.0	3.27%	60.09%	3232.50	4.96%	58.39%	3.11%	64.92%
		DMD	3767.0	5.07%	72.87%	2408.50	9.58%	67.05%	8.02%	81.02%
		All	8227.0	4.1%	65.94%	5641.00	6.93%	62.09%	5.14%	71.57%
	All	All	35531.0	3.46%	48.46%	23579.95	5.83%	42.23%	5.8%	46.4%

* Ground-truth. ** Walk4me system using our method. *** Built-in pedometer in iphone.

TABLE 3.3. The table shows the percentage error (correlation, p-value, means, and SD) of the predicted and observed values for step counts, distance traveled, and step length vs. ground truth observations.

	Ground-truth		Percentage Error of Calculated vs Observed					
			Calculated vs Obs. Steps		Calculated vs Obs. Distance		Calculated vs. Obs. Step Length	
Activities	# Act.	# Steps	Correlation (p-value)	Mean (SD)	Correlation (p-value)	Mean (SD)	Correlation (p-value)	Mean (SD)
SC-L1 to SC-L5	150	5257	0.9986 (p<0.0001)	1.2% (2.87%)	0.9946 (p<0.0001)	-3.16% (4.81%)	0.9929 (p<0.0001)	-4.23% (5.34%)
6MWT, 100MRW, FW	69	30274	0.9972 (p<0.0001)	1.49% (7.04%)	0.9933 (p<0.0001)	1.18% (9.91%)	0.9652 (p<0.0001)	0.37% (7.52%)
All	219	35531	0.9987 (p<0.0001)	1.29% (4.59%)	0.9969 (p<0.0001)	-1.59% (7.37%)	0.9796 (p<0.0001)	-2.78% (6.46%)

The level of significance was set at 0.05.

assessment methods are considered the most detailed and ideal, but their availability may be limited due to factors such as facility availability, staff availability, difficulties with patient travel to assessment locations, or public health restrictions such as those related to COVID-19. Additionally,

clinical observation methods can be susceptible to human error, such as observer fatigue or distraction, as well as instrument errors, failed video recordings, or obstructed views, which can limit the utility of the collected data. An alternative option to overcome these limitations and facilitate more frequent and convenient collection of gait data in the community setting is to use off-the-shelf technologies such as pedometers, which are commonly built into smartphones and widely used in sports. However, it is crucial to assess the reliability of these devices, particularly when used for clinical purposes. Therefore, we conducted experiments to clinically validate the reliability of using a pedometer and compared the results with those obtained by observers.

We conducted the experiment with 30 participants ambulating at different speeds, including 15 with muscle disorders that restricted their ability to walk or run at faster paces. Our results demonstrate that our system outperforms built-in pedometers and can be used to extract steps and stride lengths as clinical features of gait. Our study shows that during short-term measurements and at walking speeds commonly seen in people with muscle diseases, mobile device-based pedometers can have higher error rates when counting steps and estimating travel distances, making them unsuitable for clinical or research use even in the community setting. To address this challenge, we propose an ML-based signal processing method using our Walk4Me system that can estimate step counts, distance traveled, and step lengths with increased levels of accuracy. The advantage of our method is that it requires less observed interaction, only necessitating a short duration of time for five observed speed-calibration tests. After that, our system can automatically estimate distance and step length without the need for human interaction.

3.5. Conclusion

This study introduces a novel method for accurately estimating contact forces during various gait speeds, including walking, running, and jogging. By leveraging signal processing and ML techniques, this method provides accurate estimates of the number of steps, distance, and step length. Notably, the model is calibrated based on the individual's gait style, resulting in more personalized and precise results. Our findings demonstrate that using a single accelerometer worn near the body's center of mass is a more accurate and reliable method for CFs estimation compared to a standard pedometer. Importantly, this method can be applied to both healthy individuals and

people with DMD without the need for GRF measurements. To our knowledge, this is the first study to propose a novel method that can extract reliable CFs from raw accelerometer data across different gait speeds and in healthy participants and those with muscle disease. On average, our method of counting steps and estimating stride length and distance traveled performs well when applied to longer structured sub-maximal clinical testing efforts and free-roaming self-selected pace travel. Our methods surpass the pedometer functions native to the mobile devices we use, and with further development, they will allow us to extend basic elements of gait analysis to the community setting using commonly available consumer-level devices.

CHAPTER 4

Gait Characterization in Duchenne Muscular Dystrophy (DMD) Using a Single-Sensor Accelerometer: Classical Machine Learning and Deep Learning Approaches

4.1. Introduction

Patterns of gait disturbance in children with Duchenne muscular dystrophy (DMD) demonstrate biomechanical compensatory substitution to overcome strength loss and progressive joint contractures. Disease progression yields temporal and spatial changes in gait analysis metrics as described by Sutherland [107], D'Angelo [28], Heberer [44], and Gaudreault [37]. Perturbations and compensatory adaptations in gait are present from the onset of walking, are progressive, and follow a predictable pattern of increasing anterior pelvic tilt, increasing foot internal rotation and decreasing hip extension in stance phase, lateral trunk lean toward the supporting limb, increased hip flexion and hip abduction and decreased ankle dorsiflexion in the swing phase [107] [44] [37] [19]. The center of pressure at foot contact shifts laterally and anteriorly until an equinus posture at foot strike predominates [107]. These progressive adaptations lead to decreased step length and cadence during ambulation, decreased relative power of anteroposterior movement, and increased relative power of mediolateral movement with concomitant impairment of gait velocity.

Recent advancements in novel muscle-sparing therapeutics highlight the desirability of initiating early disease modifying treatment in the toddler years, but relatively few reliable tools exist for quantitative measurement of strength, function, and mobility in this age group, underscoring an urgent need to develop new tools that include that age group and extend upward to the limits of ambulation [19]. Wearable accelerometers can accurately measure variation in step rates in children with DMD in both the laboratory and community settings, produce natural history data that is suitable for analysis in clinical trials [34], and hold promise to provide a complete picture of the effect of strength limitation on community mobility and daily activities. To maximize their

41

effectiveness, wearable devices will need to detect and record well-understood quantitative temporal and spatial features of gait patterns while being unobtrusive and affordable.

The increasing availability of high-quality, low-cost triaxial accelerometers and inertial measurement units as stand-alone devices or integrated into commonly-available smartphones yields new opportunities to gather community level data across a wide range of typically-developing (TD) individuals and those affected by movement-related disorders. Because of this, researchers are developing a better understanding of how to extract and interpret temporal and spatial features of single accelerometer data that include not only step counts and frequencies but also a wide variety of other features [16] [49] including step lengths, step velocities [37] [109] and triaxial power spectra [45] [105] in order to use those features in principal components analysis (PCA) to evaluate between-group differences and changes over time [79]. The utility of these measures in describing disease severity and tracking disease progression has been demonstrated in the golden retriever muscular dystrophy (GRMD) of DMD [9], as well as in children with DMD [36].

Because of the increasing availability of such sensing data, there is a strong demand for automated systems that can thoroughly analyze and utilize such data. In this project, we take the first step towards this goal. We evaluate the utility of various classical machine learning (CML) and deep learning (DL)-based approaches to differentiate between children with and without DMD using data from consumer-level mobile phone accelerometers during walking and running activities. We developed a system (Walk4Me) [92,93,94] consisting of a smartphone-based application to collect raw data remotely using the phone's built-in accelerometer sensor, combined with a web-based tool to aggregate, store and analyze data. We extracted the temporal/spatial gait characteristics and used CML and DL techniques [66] to evaluate the gait changes associated with DMD, using both extracted features and raw data.

We note that our project is only a first step towards an automated disease monitoring system that can be used for disease prescreening, diagnosis assistance, progression monitoring, and possibly in a subject's natural environment (which we call community-based) instead of motion labs. It should be noted DMD needs to be diagnosed in specialty clinics, which are sparsely located across the country, and a potential patient could be hundreds of miles away from the closest clinic. The tool described in this work can be used to facilitate prescreening for such patients. Additionally,

this tool, further developed, could be used to continue monitoring the progression of the disease and quantify the effectiveness of medical treatment. Furthermore, while our current system and work focus on DMD, it has the potential to expand to other mobility-related diseases, such as post-stroke recovery and healthy aging.

4.2. Materials and Methods

4.2.1. Participants. The University of California, Davis institutional review board (IRB) reviewed and approved the protocol. Informed consent was obtained for each participant prior to initiation of study procedures. We studied thirty male participants (15 with DMD, 15 TD) who were between the ages of 3 and 16 years old, had at least 6 months of walking experience, and could perform a 10 meters walk/jog/run test in less than 10 seconds. Participants with DMD had a confirmed clinical diagnosis and were glucocorticoid therapy-naïve or on a stable regimen for at least three months. Northstar Ambulatory Assessment (NSAA) scores for DMD participants ranged from 34 to 8, indicating typical levels of function to clinically-apparent moderate mobility limitation (Table 4.1).

TABLE 4.1. **Characteristic of the children included in the study.**

Case (status)	Value	Age (years)	Weight (kg)	Height (cm)	NSAA (/34)
DMD (N=15)	Mean	9.5	37.7	127.1	20.5
	(SD)	(3.9)	(16.0)	(16.2)	(8.2)
	Min	3	17.2	101.6	8.0
	Max	16	67.7	153.3	34.0
TD (N=15)	Mean	7.7	34.2	130.8	33.8
	(SD)	(3.0)	(21.6)	(15.8)	(0.8)
	Min	4	18.6	108.5	31.0
	Max	15	101.0	165.5	34.0
	p-value	0.1664	0.6229	0.5331	**<0.0001**

Scores for TD participants ranged from 34 to 31, indicating a maximal range of task performance, with a low score of 31 in a six-year-old participant indicating age-appropriate achievement of developmental milestones [**23**].

4.2.2. Materials. Participants wore an NGN Sport fitness phone belt (Engine Design Group, LLC, Los Angeles, CA) carrying an Apple iPhone 11 (Apple, Inc. Cupertino, CA), which includes

a single, built-in triaxial accelerometer. The belt was placed close to the body's center of mass at the lumbosacral junction at the approximate level of the iliac crest, with the phone oriented so that the right lower side was consistently oriented in the upper right position to measure acceleration in the vertical, mediolateral and anteroposterior axes (Figure 4.1A and Figure 4.1B). We developed an iPhone app (Walk4Me) to continuously stream raw sensor data to a cloud server at a sampling rate of 100 Hz. At the conclusion of each assessment session, data was processed via Walk4Me's web application to extract clinical gait features and train the CML and DL models.
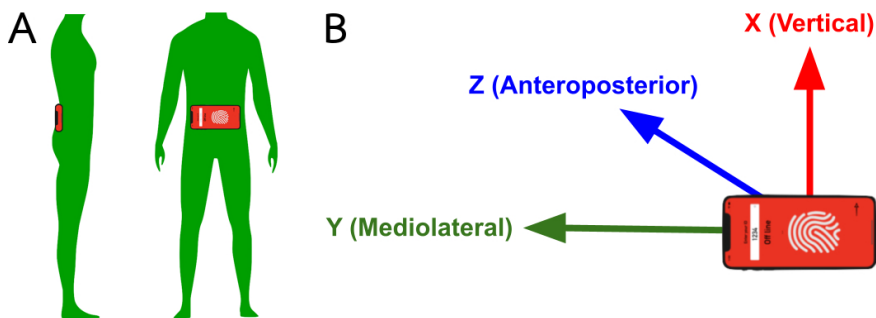

FIGURE 4.1. **The position and the orientation of the accelerometer.** (A) The smartphone with its built-in accelerometer is positioned near the body's center of mass, fastened by a belt around the waist. (B) The accelerometer records acceleration in three axes. The x-axis is vertical, the y-axis is mediolateral, and the z-axis is anteroposterior.

4.2.3. Gait and Functional Testing. Children performed eight walking and running activities along a 25 meters straight-line course according to established protocols [76]. The first five activities were speed-calibration-tests (SC) where the child walked along a 25 meters long corridor at incrementally increasing gait velocities every 25 meters from slowest speed, speed-calibration-L1 (SC-L1), to a slow-walk (SC-L2), to a comfortable self-selected-walk pace (SC-L3), to a fast-walk (SC-L4), to a run or fastest possible gait (SC-L5) [37]. The evaluator recorded the observed number of steps taken during each effort. Participants then performed a 6-minute walk test (6MWT), followed by a 100 meters fast-walk/jog/run (100MRW) using previously described methods [76] [3]. The evaluator recorded the 6MWT distance and the number of steps in the first 50 meters, and the time to complete the 100MRW.

Participants also completed the NSAA according to established protocols [75]. The NSAA consists of a set of 17 graded tasks representing common mobility and positional transfer activities.

Tasks are graded individually with a total score of 34 points indicating full functional mobility. All evaluations were recorded on video for later verification of step counts, distances, task times, and task quality.

4.2.4. Data Analysis. Data analysis was performed using both CML and DL approaches (Figure 4.2). In our first approach, we processed raw data from each activity to extract temporospatial gait CFs including speed, step length, step frequency, total power, percent of power in each axis, and force index in a manner similar to those described by Barthelemy [9] and Fraysse [35]. We normalized speed and step length to standing height. For information purposes, we also compared the means of features between DMD and TD control groups for each activity using simple two-tailed T-tests. For convenience, we bold the p-values when it is smaller than 0.05, without a Bonferroni adjustment. We note p-values are reported here primarily to describe relationships between temporospatial gait CFs and model dimensions across DMD and control groups - they are not considered in ML algorithms. A feature with a non-statistically-significant p-value could still be useful in ML tasks. This is especially true when the sample size is relatively small, as in this study. We applied CML approaches to the temporospatial gait CFs both with and without dimensional
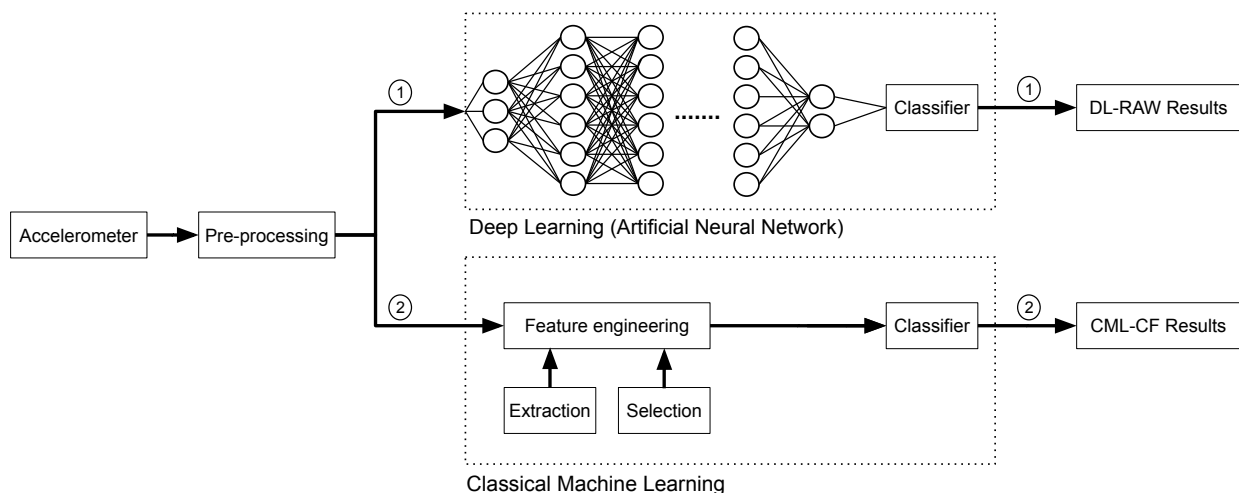

FIGURE 4.2. **A typical process diagram of classical machine learning (CML-CF) and deep learning (DL-RAW).** The two methods used are (1) DL-RAW and (2) CML-CF.

reduction to identify data characteristics and models that were most accurate at predicting group membership. We used Pearson's Correlation Coefficients with a Bonferroni adjustment to describe

relationships between temporospatial gait CFs and latent domain scores after dimensional reduction, except during comparisons with ordinal measures where we used Spearman's Rank Ordered Correlation.

4.2.4.1. *Extraction and Evaluation of Temporospatial Gait Clinical Features.* We extracted the following eight temporospatial gait CFs from the raw accelerometer data:

Speed (SP) is measured in meters per second and normalized by the height of the child in meters. This feature is calculated by dividing the distance by the time spent performing each activity. **Step Length (SL)** is measured in meters and normalized by the height of the child in meters. This feature is calculated by dividing the total distance by the number of steps. The number of steps is calculated using a low pass filter on the anteroposterior (z-axis) acceleration signal (Figure 4.3C and Figure 4.3F). Then, we calculate the number of peaks, where each peak represents one step, and a complete gait cycle is composed of two steps. **Step Frequency (SF)** is measured in steps per second. This feature is calculated by dividing the number of steps by the time for each activity. **Total Power (TP)** is measured in W/kg. This feature is calculated by first transferring the time domain to the frequency domain of the three axes using Fast Fourier Transform (FFT): vertical (x-axis), mediolateral (y-axis), and anteroposterior (z-axis) (Figure 4.3D, Figure 4.3E, and Figure 4.3F, respectively). Then, we sum the integral of the power (normalized by weight) in each of the three axes (Figure 4.3G, Figure 4.3H, and Figure 4.3I). **Mediolateral Power (MP)** is measured in % of TP. This feature is calculated by first transferring the time domain of the accelerometer's y-axis (Figure 4.3E) to the frequency domain using FFT. Then we calculate the integral of the power spectrum density (PSD) (Figure 4.3H). Finally, we normalize the value by weight and TP. **Anteroposterior Power (AP)** is measured in % of TP. This feature is calculated by first transferring the time domain of the accelerometer's z-axis (Figure 4.3F) to the frequency domain using FFT. Then we calculate the integral of the PSD (Figure 4.3I). Finally, we normalize the value by weight and TP. **Vertical Power (VP)** is measured in % of TP. This feature is calculated by first transferring the time domain of the accelerometer's x-axis (Figure 4.3D) to the frequency domain using FFT. Then we calculate the integral of the PSD (Figure 4.3G). Finally, we normalize the value by weight and TP. **Force Index (FI)** is measured in N/kg. This feature is calculated by first transferring the time domain of the accelerometer's z-axis (Figure 4.3F) to the

46

frequency domain using FFT. Then, we divide the integral of the PSD (Figure 4.3I) by the average speed in order to average the force index.
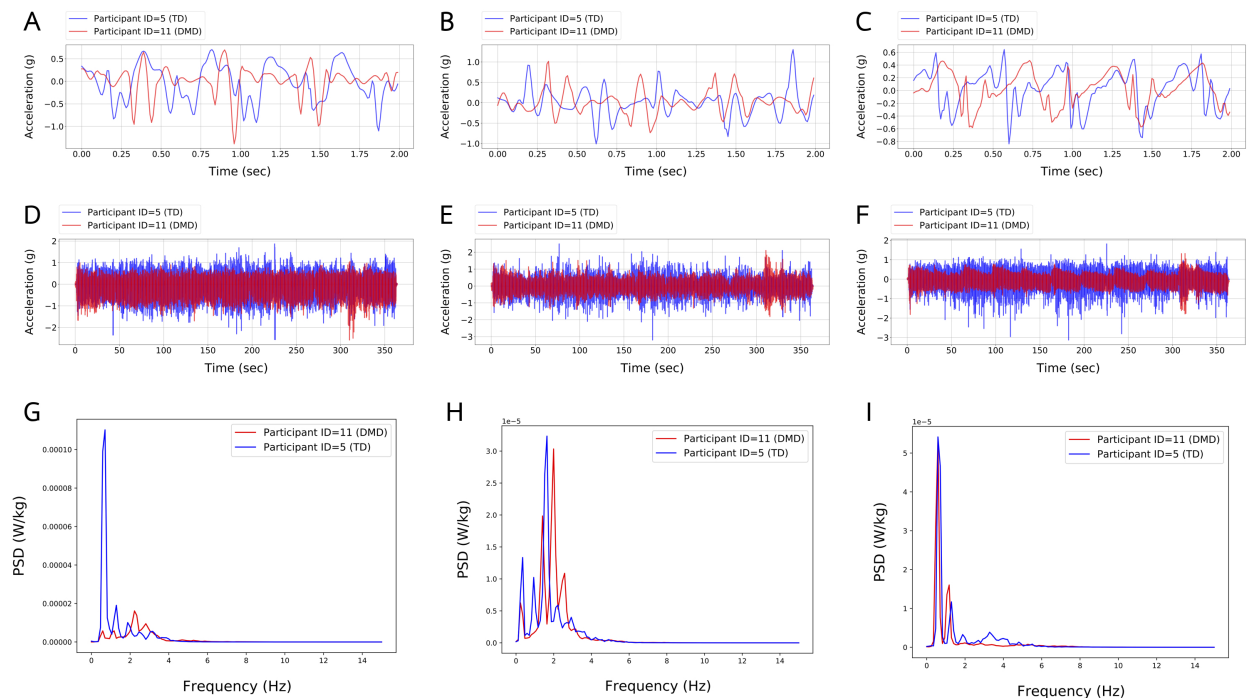
FIGURE 4.3. **Figures Illustrate a comparison of two children, DMD (ID=11) and TD (ID=5), in the three-axial accelerations Vertical (x-axis), Mediolateral (y-axis), and Anteroposterior (z-axis) of 6MWT.** Figures (A, B, and C) represent a 2 second TW acceleration signal of a TD child performing 5 steps, and a DMD child performing 4 steps. Figures (D, E, and F) represent the acceleration signal of a TD child and a DMD child performing the whole 6MWT. Figures (G, H, and I) represent the Power-Spectrum-Density (PSD) of a TD child and a DMD child performing the whole 6MWT.

4.2.4.2. *CML and DL Analytical Methods.* We used two different ML approaches to classify accelerometer data collected at a range of gait speeds as belonging to a child with DMD or a TD. For the **CML-CF** approach, 6 different classifiers were implemented: Random Forest (RF), Decision-Tree (DT), Support-Vector Machine (SVM), K-Nearest Neighbors (KNN), Gaussian Naive Bayes (GNB), Logistic Regression (LR). In this method, we used the eight extracted temporospatial gait CFs as input into CML classifiers with and without dimensionality reduction. The eight extracted temporospatial gait CFs for each child were used to train the CML model for all eight activities.

For the **DL-RAW** method, we used a Convolutional Neural Network (CNN) model with the time-windowed raw accelerometer signal as an input in the DL classifier. Figure 4.4 shows the process diagram of the DL-RAW and CML-CF approaches.
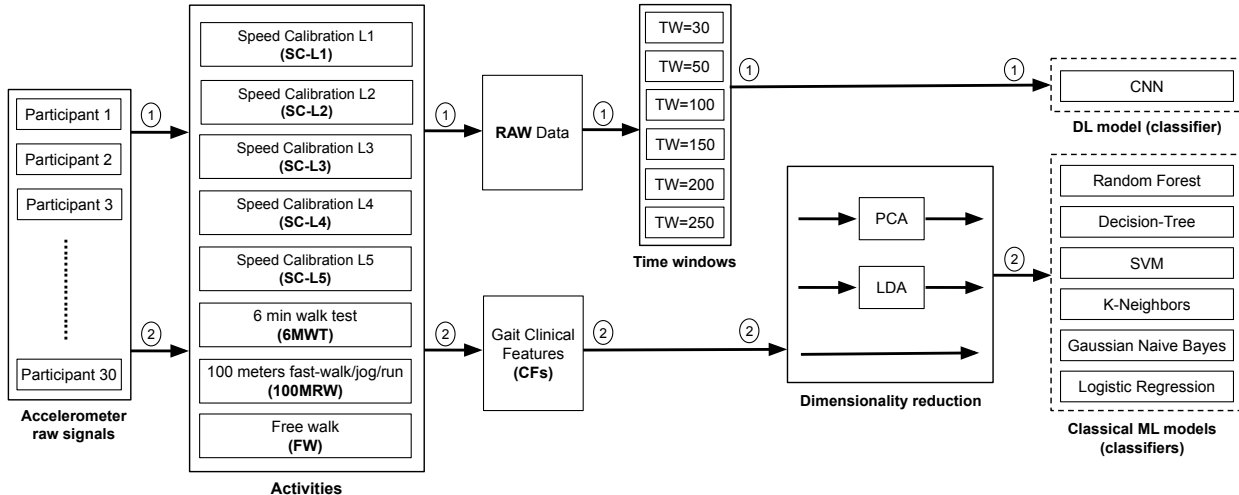
FIGURE 4.4. **The process diagram of (1) the DL-RAW approach and (2) the CML-CF approach.**

4.2.4.3. *Dimensionality Reduction.* To analyze how the participants are distributed on the projection plane, we used dimensionality reduction techniques in a manner similar to those described by Fraysse and Barthelemy [8] [35]. We investigated whether using principal component analysis (PCA) and linear discriminant analysis (LDA) techniques would affect the discrimination accuracy between DMD and TD groups. We fed the eight extracted temporospatial gait CFs to CML models using PCA and LDA. We used PCA and LDA to reduce the dimensionality of the input features of all participants and projected their models' results into a two-dimensional (2D) and one-dimensional (1D) representation, respectively. We compared PCA and LDA accuracy with the original models without using any dimensionality reduction techniques.

4.2.4.4. *Preprocessing of Raw Accelerometer Signals Using Time-windowing.* In the CML-CF method, data from each activity for each participant represents an individual input to the model. The temporospatial gait CFs of each activity must be extracted entirely before using it as input to the model. In the DL-RAW method, we used raw acceleration values in each of the three axes as input to the DL model (Figure 4.4). We used the window-slicing method [21] [70] to segment

48

the raw data from each activity into multiple fixed time-windows (TWs) (Figure 4.5) in order to augment the data input [64] and to facilitate DL model convergence [91] [103]. We examined six distinct TWs (i.e., 0.3, 0.5, 1, 1.5, 2, 2.5 seconds) to determine the signal duration required for the highest model accuracy. For each activity, the model predicted whether an individual TW was labeled as DMD or TD. We then used those predictions to calculate the overall percentage of the TWs predicted as DMD or TD. At inference time, we examined the percentage of typical/atypical decisions for each TW, and used majority voting to determine whether the participants were correctly labeled as having DMD.
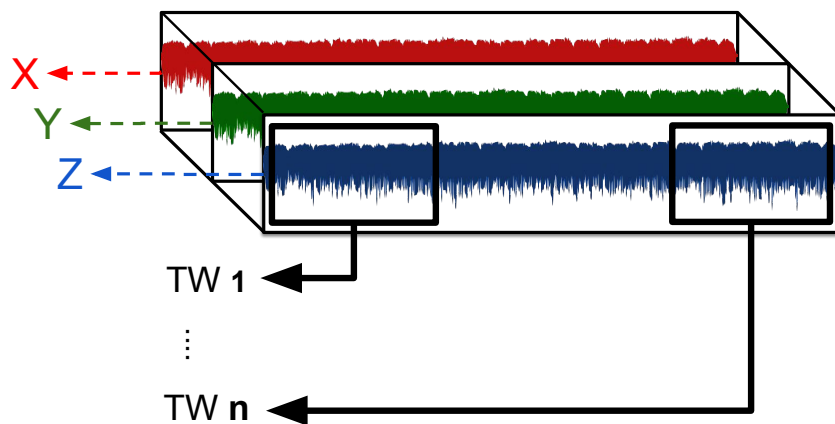
FIGURE 4.5. **Time-windows (TWs) of the accelerometer (x, y, and z axes)**

4.2.4.5. *Cross-Validation of ML Models.* We used Leave-One-Subject-Out (LOSO) cross-validation to evaluate classification accuracy in the presence of slight variations in dataset contents. We use LOSO to ensure that the model does not have information leakage. Having data from a particular participant in both training and testing datasets makes the model more familiar with the data of that participant. Thus, using a completely unseen participant in the testing-set will ensure that the model is not biased toward that participant because it has never been trained on any part of that participant's data before. The present dataset included 30 participants. To ensure the accuracy of our predictions, we utilized data from 29 of the participants as the training-set and predicted the label of the remaining participant. We repeated this process 30 times, each time leaving out a different participant for testing, and calculated the accuracy of each model by averaging the 30-fold accuracy. By using this method, we were able to ensure a robust and reliable analysis of our data.

49

4.3. Results

4.3.1. Extracted Temporospatial Gait Clinical Features. We compared temporospatial gait CFs of DMD and TD participants extracted from our sensor data, shown in Table 4.2. For the 25 meters course at a slow walking pace (SC-L1 and SC-L2), walking speed and step length were significantly lower, and the mediolateral percent of accelerations was significantly higher in people with DMD. Those observations are consistent with clinical descriptions of a slower and more lateral DMD gait. On our 25 meters and more extended efforts, self-selected-walk and fast-walk paces (SC-L3, FW, SC-L4, and 6MWT), we saw a similar pattern of differences that added a lower percent of vertical accelerations in people with DMD. At jogging to running paces (SC-L5 and 100MRW), significantly reduced step frequency of people with DMD became apparent. When combining all efforts, differences in speed, step frequency, step length, total power, and percent power in vertical and mediolateral axes differed significantly between the two groups. Interestingly, there was little between-group-difference in anteroposterior force at any but the slowest of velocities.

4.3.2. Comparison between CML-CF and DL-RAW Approaches. Two different ML approaches were performed: CML-CF and DL-RAW. Figure 4.6 and Table 4.3 summarize the best results obtained from each approach. Figure 4.6 represents the optimal accuracy obtained by performing the eight different gait activities: SC-L1, SC-L2, SC-L3, SC-L4, SC-L5, 6MWT, 100MRW, and FW (described in Section 4.2.3). LOSO was used to verify the accuracy results.

CML-CF achieved the best accuracy of 100% at SC-L3, while DL-RAW achieved 86.67% accuracy at both SC-L2 and FW. In comparison, the CML-CF approaches utilize CFs while DL-RAW only uses raw accelerometer data. CFs in the CML-CF approach include features related to the activity as a whole, such as speed and step length, and participants in particular, such as age, height, and weight. We note that the DL-RAW approach does not use participant features. It also cannot extract other activity features such as speed and step length because it does not have information on the distance in each experiment. These suggest the effectiveness of CFs.

Furthermore, CML-CF considers the average metrics of each participant, while DL looks at TWs of raw signals of the activity. DL requires substantial data to train a model, so its accuracy drops with larger TW sizes in short-duration activities. The drops continue as activities get shorter

50

TABLE 4.2. **The summary of temporospatial gait CFs for the eight different gait activities.**

Activities	Case	Value	Temporospatial Gait Clinical Features (CFs)							
			SP (%)	SF	SL (%)	TP (10^{-6})	VP (%)	MP (%)	AP (%)	FI (10^{-3})
SC-L1	TD	Mean	0.35	1.29	0.27	64.84	28.45	36.24	35.31	11.54
		(SD)	(0.06)	(0.19)	(0.02)	(48.52)	(4.07)	(7.22)	(5.45)	(8.94)
	DMD	Mean	0.26	1.12	0.23	62.0	27.82	42.1	30.08	5.79
		(SD)	(0.1)	(0.32)	(0.04)	(92.77)	(5.48)	(6.17)	(4.36)	(6.95)
		p-value	**0.0077**	0.093	**0.0004**	0.9172	0.7246	**0.0239**	**0.0072**	0.0594
SC-L2	TD	Mean	0.55	1.62	0.34	155.3	31.48	34.29	34.23	16.02
		(SD)	(0.07)	(0.16)	(0.02)	(104.07)	(7.78)	(8.4)	(5.63)	(13.4)
	DMD	Mean	0.46	1.58	0.29	170.27	28.25	40.16	31.59	9.85
		(SD)	(0.12)	(0.31)	(0.03)	(179.9)	(6.2)	(6.8)	(6.63)	(11.03)
		p-value	**0.0139**	0.6133	**<0.0001**	0.7823	0.22	**0.0445**	0.2487	0.1801
SC-L3	TD	Mean	0.76	1.89	0.4	355.84	36.93	31.36	31.7	19.54
		(SD)	(0.12)	(0.23)	(0.02)	(255.55)	(10.22)	(7.9)	(5.72)	(16.06)
	DMD	Mean	0.63	1.89	0.33	334.35	30.16	39.32	30.52	14.05
		(SD)	(0.1)	(0.24)	(0.03)	(317.31)	(5.98)	(7.48)	(5.79)	(13.76)
		p-value	**0.0024**	0.9457	**<0.0001**	0.8396	**0.0348**	**0.0085**	0.5786	0.323
SC-L4	TD	Mean	1.28	2.55	0.5	1907.94	38.92	30.94	30.14	58.39
		(SD)	(0.24)	(0.32)	(0.04)	(1690.83)	(8.75)	(8.73)	(5.84)	(58.04)
	DMD	Mean	0.94	2.4	0.39	1807.78	29.07	40.11	30.83	49.58
		(SD)	(0.28)	(0.51)	(0.06)	(3060.77)	(10.71)	(9.29)	(3.86)	(65.31)
		p-value	**0.0016**	0.3496	**<0.0001**	0.9125	**0.0101**	**0.0095**	0.7091	0.6992
SC-L5	TD	Mean	2.44	3.61	0.67	9219.81	50.13	21.53	28.34	93.53
		(SD)	(0.48)	(0.52)	(0.09)	(6775.47)	(8.93)	(10.08)	(10.09)	(71.93)
	DMD	Mean	1.22	2.82	0.42	4235.65	35.67	35.01	29.32	68.55
		(SD)	(0.48)	(0.73)	(0.07)	(5913.64)	(13.68)	(10.86)	(9.18)	(88.33)
		p-value	**<0.0001**	**0.002**	**<0.0001**	**0.0406**	**0.0019**	**0.0015**	0.7822	0.4029
6MWT	TD	Mean	1.18	2.36	0.5	1521.86	41.68	29.51	28.81	915.74
		(SD)	(0.11)	(0.17)	(0.03)	(1054.76)	(11.07)	(9.01)	(6.71)	(677.16)
	DMD	Mean	0.79	2.05	0.38	992.04	31.61	38.79	29.6	840.77
		(SD)	(0.25)	(0.39)	(0.06)	(1441.95)	(6.02)	(6.93)	(5.46)	(1210.56)
		p-value	**<0.0001**	**0.0088**	**<0.0001**	0.2664	**0.0056**	**0.0046**	0.7327	0.837
100MRW	TD	Mean	2.27	3.39	0.67	8402.12	52.89	18.05	29.06	480.5
		(SD)	(0.4)	(0.47)	(0.08)	(5348.93)	(8.86)	(9.36)	(10.53)	(410.11)
	DMD	Mean	1.1	2.57	0.42	2944.45	39.68	34.82	25.51	599.57
		(SD)	(0.42)	(0.67)	(0.07)	(5073.6)	(17.19)	(12.28)	(7.58)	(955.96)
		p-value	**<0.0001**	**0.0013**	**<0.0001**	**0.0148**	**0.0172**	**0.0006**	0.3512	0.6685
FW	TD	Mean	0.83	1.96	0.42	617.5	41.38	28.46	30.17	321.34
		(SD)	(0.16)	(0.29)	(0.04)	(478.61)	(8.92)	(7.49)	(4.87)	(263.17)
	DMD	Mean	0.61	1.83	0.33	529.12	32.87	36.47	30.66	638.46
		(SD)	(0.18)	(0.42)	(0.06)	(983.01)	(8.51)	(6.99)	(7.19)	(1252.37)
		p-value	**0.0015**	0.3626	**<0.0001**	0.7565	**0.0124**	**0.0052**	0.826	0.3454
All	TD	Mean	1.21	2.33	0.47	2780.65	40.23	28.8	30.97	239.57
		(SD)	(0.76)	(0.83)	(0.14)	(4680.29)	(11.6)	(10.13)	(7.35)	(417.76)
	DMD	Mean	0.74	2.01	0.35	1333.61	31.62	38.47	29.91	262.26
		(SD)	(0.4)	(0.69)	(0.08)	(3175.88)	(10.1)	(8.54)	(6.43)	(740.37)
		p-value	**<0.0001**	**0.0016**	**<0.0001**	**0.0062**	**<0.0001**	**<0.0001**	0.2423	0.7714

The level of significance was set at 0.05. SP: speed in meters per second normalized by height in meters. TP: total Power with a unit of 10^{-6} W/kg. SF: Step Frequency with a unit of step/second. SL: Step length as a percentage of standing height. MP: the percentage of power in the x-axis normalized by total power TP (%). AP: the percentage of power in the y-axis normalized by total power TP (%). VP: the percentage of power in the z-axis normalized by total power TP (%). FI: force index with a unit of 10^{-3} N/kg.
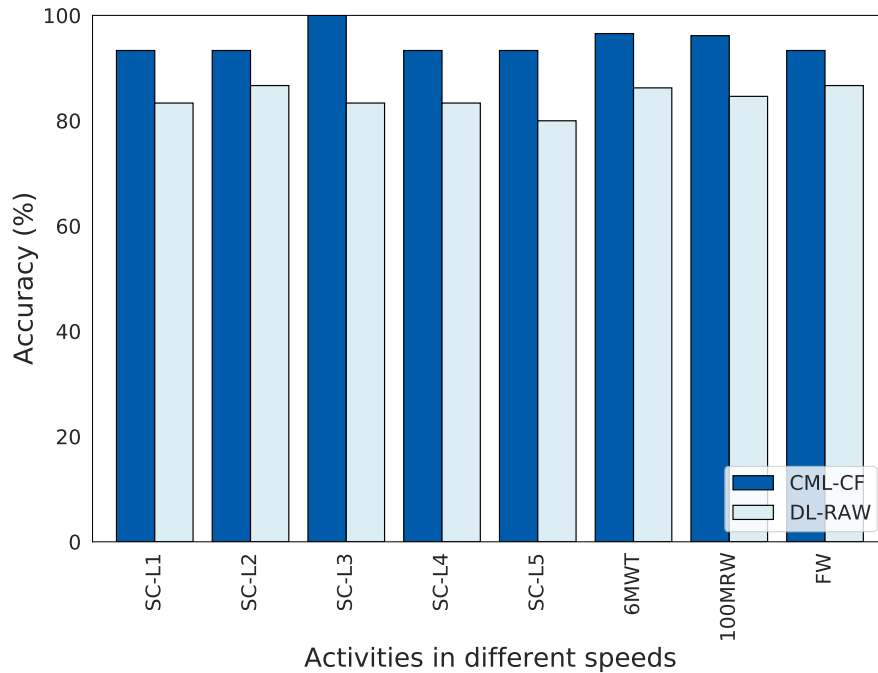
FIGURE 4.6. **The optimal classifier's accuracy of CML-CF and DL-RAW.** The y-axis represents the classifier accuracies, while the x-axis represents different activities.

with fewer TWs available for analysis (e.g., from SC-L1 to SC-L4). In more extended activities like 6MWT and FW, the differences in accuracy between TW sizes tend to decrease.

4.3.2.1. *CML-CF Approach.* We evaluated the CML approach using temporospatial gait CFs. We report the classifier's accuracy among 6 different CML techniques: RF, DT, SVM, KNN, GNB, and LR. For CML-CF, three different methods were used: no projection (CML), PCA projection (CML-PCA), and LDA projection (CML-LDA). Each group of results represents one of the eight different activities mentioned previously. Our models achieved the best accuracy of 100% in self-selected-walk speed (SC-L3) and exceeded 96% in both fast-walk/jog/run speeds (100MRW) and fast-walk speed (6MWT).

4.3.2.2. *DL-RAW Approach.* We evaluated the DL approach using raw data. Since DL algorithms require a large amount of data to train, each activity was divided into fixed TWs using

Table 4.3. **Classification performance of CML-CF and DL-RAW approaches.**

| Activities | Alg. | CML-CF | | | DL-RAW | |
		CML (%)	PCA (%)	LDA (%)	TW* (Samples)	CNN (%)
SC-L1	RF	76.67	**86.67**	**93.33**	30	79.98
	DT	66.67	80.0	**93.33**	50	**83.35**
	SVM	73.33	73.33	**93.33**	100	**83.35**
	KNN	**83.33**	73.33	**93.33**	150	79.98
	GNB	70.0	66.67	**93.33**	200	79.98
	LR	80.0	73.33	**93.33**	250	79.98
SC-L2	RF	70.0	53.33	**93.33**	30	**86.67**
	DT	70.0	56.67	**93.33**	50	83.35
	SVM	**83.33**	**70.0**	90.0	100	79.98
	KNN	66.67	**70.0**	86.67	150	76.66
	GNB	76.67	66.67	90.0	200	60.01
	LR	76.67	60.0	90.0	250	63.33
SC-L3	RF	**90.0**	53.33	**100.0**	30	**83.35**
	DT	83.33	53.33	**100.0**	50	76.66
	SVM	70.0	70.0	96.67	100	63.33
	KNN	70.0	50.0	96.67	150	60.01
	GNB	73.33	70.0	96.67	200	63.33
	LR	80.0	**73.33**	**100.0**	250	56.69
SC-L4	RF	63.33	63.33	83.33	30	76.66
	DT	**80.0**	70.0	83.33	50	**83.35**
	SVM	**80.0**	**73.33**	90.0	100	79.98
	KNN	76.67	70.0	83.33	150	56.69
	GNB	76.67	70.0	**93.33**	200	66.65
	LR	76.67	66.67	**93.33**	250	60.01
SC-L5	RF	**90.0**	66.67	**93.33**	30	**79.98**
	DT	86.67	60.0	**93.33**	50	66.65
	SVM	86.67	**83.33**	**93.33**	100	56.69
	KNN	83.33	**83.33**	**93.33**	150	73.34
	GNB	86.67	76.67	**93.33**	200	50.0
	LR	**90.0**	**83.33**	**93.33**	250	73.34
6MWT[a]	RF	86.21	72.41	93.1	30	**86.23**
	DT	86.21	**89.66**	93.1	50	**86.23**
	SVM	**89.66**	79.31	**96.55**	100	82.76
	KNN	82.76	82.76	82.76	150	82.76
	GNB	**89.66**	75.86	93.1	200	79.3
	LR	82.76	86.21	93.1	250	79.3
100MRW[b]	RF	88.46	88.46	92.31	30	80.76
	DT	88.46	**92.31**	92.31	50	80.76
	SVM	88.46	80.77	**96.15**	100	69.24
	KNN	**92.31**	80.77	92.31	150	**84.62**
	GNB	**92.31**	88.46	**96.15**	200	80.76
	LR	**92.31**	88.46	92.31	250	**84.62**
FW	RF	80.0	83.33	**93.33**	30	83.35
	DT	80.0	76.67	**93.33**	50	**86.67**
	SVM	**93.33**	**86.67**	90.0	100	79.98
	KNN	83.33	83.33	86.67	150	73.34
	GNB	80.0	83.33	86.67	200	76.66
	LR	90.0	80.0	86.67	250	76.66

[a] One DMD child was unable to perform the activity.
[b] Four DMD children were unable to perform the activity.
* Each second contains 100 samples.

the window-slicing method to provide more data for the DL model. Table 4.3 reports the detailed results of DL-RAW analyses and optimization of TW segment length.

Our model demonstrated the highest accuracy of 86.67% in slow-walk speed (SC-L2) and self-selected-walk (FW) activities. We hypothesize this was achieved using TWs long enough to capture a portion of the gait cycle with a strong correlation between the accelerometer signal's x, y, and z-axes. However, it is important to note that different portions of the gait cycle may have varying levels of distinguishability, resulting in some parts having a higher likelihood of being classified correctly than others. When the portions of gait cycles are located in a region where the correlation between the x, y, and z-axis is weak, the characteristics of those portions may not be captured effectively by the DL model. As a result, the TW may not have enough distinctive features to be classified correctly. One possibility is to use a TW that includes at least one complete gait cycle can ensure that all the portions that contain sufficient correlation between the x, y, and z axes are captured. By setting a threshold for the percentage of the contribution of each portion to the identification task, we can further enhance the accuracy and reliability of our model. This approach can potentially improve clinical explainability and enhance performance. The only limitation of this approach is that it requires more data to train the model. We plan to explore it further in future studies.

4.3.3. Relationship Between Extracted Step Length, Gait Speed, and Functional Ability. To help illustrate how step length measures relate to the overall clinical function, we compared step lengths for DMD participants with NSAA clinical function scores ≥ 30 (near-TD based on a 34-point maximum score) with TD participants at each gait speed from slow walking to jogging or running (SC-L1 to SC-L5) as shown in Figure 4.7. At all but the lowest slow walking speed, this highest-functioning group of DMD children demonstrated significant reductions in step length compared to their TD peers. We then subdivided all DMD participants by NSAA score into groups of near-TD (≥ 30), mildly affected ($20 - 29$), moderately affected ($10 - 19$), and severely affected (< 10) individuals, as shown in Figure 4.8. In those DMD children, average step lengths differed significantly between each velocity activity from a slow walk (SC-L1) to a fast walk (SC-L4). Furthermore, there were significant differences in step length between the mildly and moderately

affected NSAA groups within the medium-slow (SC-L2), comfortable (SC-L3), and fast walk (SC-L4) activities, suggesting that the largest drop in step length appears to occur near the middle of the NSAA scoring range (scores $10 - 29$). We should note, however, that the largest difference in step lengths seen in the moderate NSAA group could be due to the non-linear nature of NSAA scoring. The NSAA score is a composite of multiple ordinal item scores rather than a continuous variable.
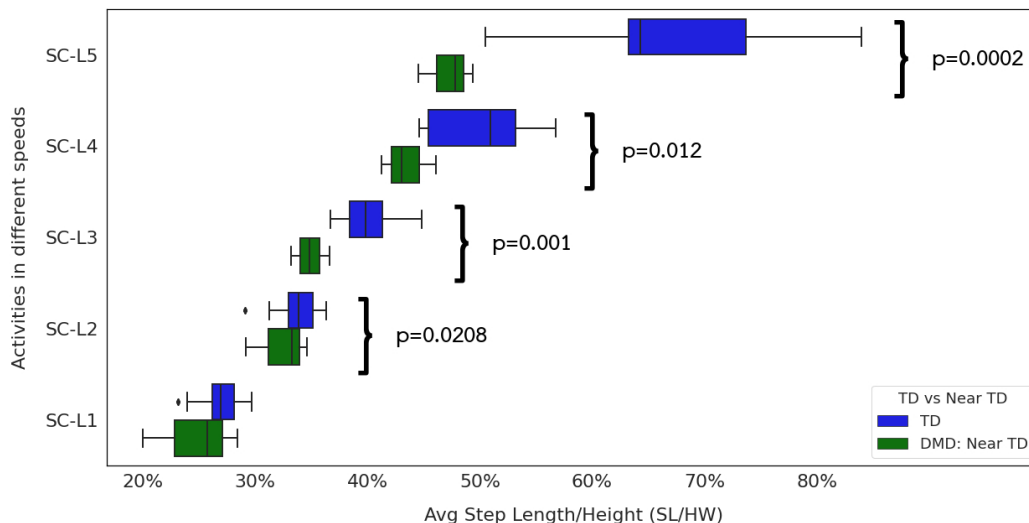

FIGURE 4.7. **Step Length by Different Gait Speeds.** TD and ambulatory DMD children with NSAA$\geq$ 30.

We calculated correlations between temporospatial gait CFs and PCA and LDA dimension scores to determine which temporospatial gait CFs helped describe CML tool outputs (Table 4.4). For PCA-based analyses, the PC1 component score was significantly and highly correlated with speed, step frequency, stride length, and total power across all velocity groups. The PC1 component was also significantly and moderately correlated with NSAA clinical and functional scores. In contrast, the PC2 component score was predominantly associated with the axial percentages of total power. The unidimensional LDA score was significantly and highly correlated with step length and significantly and highly to moderately correlated with the NSAA score. The relationship between step length and the LDA coordinates at different velocities is illustrated in Figure 4.9. Step length increased with velocity across all velocity activities (SC-L1 to SC-L5), with positive LDA scores on average
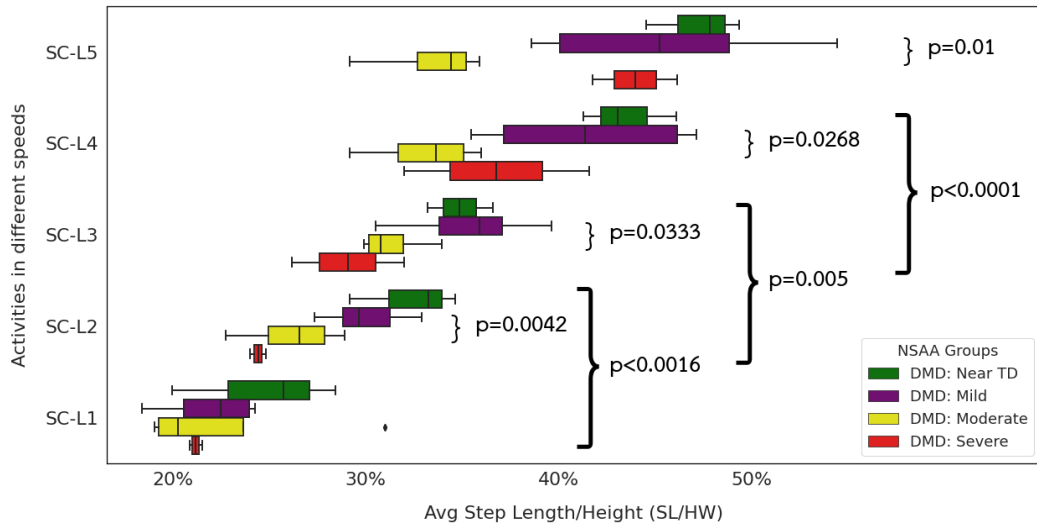
FIGURE 4.8. **Step Length by Different Gait Speeds.** NSAA severity group in ambulatory DMD children.

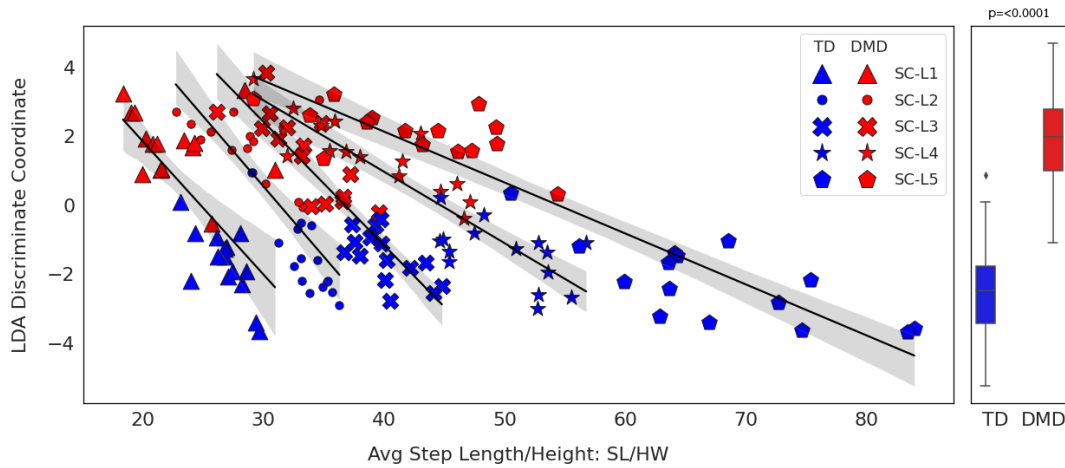indicating DMD participants and negative LDA scores indicating TD with little overlap between groups ($p < 0.0001$).

FIGURE 4.9. **LDA coordinates vs. step length at different gait speeds.** DMD vs. TD.

Table 4.4.

TABLE 4.4. **Correlation of Dimensional Reduction Scores with Extracted Clinical Features and NSAA Score**

Components	Activities	Value	Temporospatial Gait Clinical Features (CFs)								NSAA*
			SP	SF	SL	TP	VP	MP	AP	FI	
PC1[a]	SC-L1	Correlation	0.93	0.85	0.72	0.82	0.03	-0.44	0.55	0.81	0.56
		(p-value)	(**<0.0001**)	(**<0.0001**)	(**0.0002**)	(**<0.0001**)	(1.0)	(0.5332)	(0.0589)	(**<0.0001**)	(**0.0014**)
	SC-L2	Correlation	0.92	0.85	0.61	0.83	-0.28	-0.32	0.74	0.74	0.4
		(p-value)	(**<0.0001**)	(**<0.0001**)	(**0.0116**)	(**<0.0001**)	(1.0)	(1.0)	(**0.0001**)	(**0.0001**)	(**0.0273**)
	SC-L3	Correlation	0.93	0.85	0.54	0.83	0.29	-0.41	0.17	0.72	0.51
		(p-value)	(**<0.0001**)	(**<0.0001**)	(0.0679)	(**<0.0001**)	(1.0)	(0.8368)	(1.0)	(**0.0003**)	(**0.0041**)
	SC-L4	Correlation	0.91	0.85	0.66	0.83	-0.78	0.82	-0.23	0.7	0.71
		(p-value)	(**<0.0001**)	(**<0.0001**)	(**0.0027**)	(**<0.0001**)	(**<0.0001**)	(**<0.0001**)	(1.0)	(**0.0007**)	(**<0.0001**)
	SC-L5	Correlation	-0.93	-0.85	-0.83	-0.89	-0.66	0.88	-0.2	-0.76	-0.62
		(p-value)	(**<0.0001**)	(**<0.0001**)	(**<0.0001**)	(**<0.0001**)	(**0.0023**)	(**<0.0001**)	(1.0)	(**<0.0001**)	(**0.0002**)
PC2[b]	SC-L1	Correlation	0.12	0.15	0.01	0.35	-0.76	0.89	-0.52	0.14	-0.28
		(p-value)	(1.0)	(1.0)	(1.0)	(1.0)	(**0.0001**)	(**<0.0001**)	(0.1481)	(1.0)	(0.1278)
	SC-L2	Correlation	0.19	0.11	0.2	-0.43	0.85	-0.9	0.19	-0.4	0.19
		(p-value)	(1.0)	(1.0)	(1.0)	(0.7552)	(**<0.0001**)	(**<0.0001**)	(1.0)	(1.0)	(0.3069)
	SC-L3	Correlation	0.19	-0.32	0.57	-0.53	0.85	-0.72	-0.25	-0.57	0.3
		(p-value)	(1.0)	(1.0)	(**0.0485**)	(0.1187)	(**<0.0001**)	(**0.0003**)	(1.0)	(**0.0453**)	(0.1049)
	SC-L4	Correlation	-0.19	0.3	-0.61	0.27	-0.12	0.45	-0.65	0.29	-0.41
		(p-value)	(1.0)	(1.0)	(**0.0144**)	(1.0)	(1.0)	(0.603)	(**0.004**)	(1.0)	(**0.0253**)
	SC-L5	Correlation	-0.09	-0.17	0.05	-0.42	0.56	0.11	-0.93	-0.54	0.29
		(p-value)	(1.0)	(1.0)	(1.0)	(0.9396)	(0.0645)	(1.0)	(**<0.0001**)	(0.0945)	(0.1268)
LDA	SC-L1	Correlation	-0.54	-0.35	-0.7	-0.04	-0.07	0.47	-0.55	-0.42	-0.72
		(p-value)	(0.0695)	(1.0)	(**0.0007**)	(1.0)	(1.0)	(0.3336)	(0.0601)	(0.7212)	(**<0.0001**)
	SC-L2	Correlation	-0.51	-0.11	-0.78	0.05	-0.26	0.42	-0.25	-0.3	-0.63
		(p-value)	(0.1305)	(1.0)	(**<0.0001**)	(1.0)	(1.0)	(0.7424)	(1.0)	(1.0)	(**0.0002**)
	SC-L3	Correlation	-0.64	-0.02	-0.91	-0.05	-0.46	0.56	-0.13	-0.23	-0.68
		(p-value)	(**0.0054**)	(1.0)	(**<0.0001**)	(1.0)	(0.3801)	(**0.044**)	(1.0)	(1.0)	(**<0.0001**)
	SC-L4	Correlation	-0.65	-0.19	-0.89	-0.01	-0.54	0.54	0.1	-0.07	-0.79
		(p-value)	(**0.004**)	(1.0)	(**<0.0001**)	(1.0)	(0.0732)	(0.0795)	(1.0)	(1.0)	(**<0.0001**)
	SC-L5	Correlation	-0.87	-0.59	-0.92	-0.42	-0.59	0.61	0.05	-0.18	-0.67
		(p-value)	(**<0.0001**)	(**0.0195**)	(**<0.0001**)	(0.8081)	(**0.0204**)	(**0.0134**)	(1.0)	(1.0)	(**0.0001**)

* The Spearman rank-order correlation coefficient. The level of significance was set at 0.05.
[a] Captures the most variation.
[b] Captures the second most variation.

4.4. Discussions

The purpose of our study was to explore the utility and feasibility of collecting clinically meaningful gait data using consumer-level accelerometers outside of the formal gait laboratory setting and to explore a range of CML and DL methods to differentiate between children with DMD and TD of different ages. Extracting and describing a combination of well-known and understood temporospatial gait CFs allowed us to identify some of the characteristics that CML and DL tools used to differentiate between DMD and TD groups.

4.4.1. Utility and Feasibility of CML and DL Approaches to Extracted Temporospatial Gait CFs.
We investigated two different ML approaches, CML with extracted temporospatial gait CFs and DL with RAW data. We reported the outcome for each gait velocity in Table 4.3. We also showed that using CML with extracted temporospatial gait CFs to predict membership in the

DMD group yielded satisfactory results, with correct predictions for up to 100% of participants. The CML-CF approach typically shows improved accuracy with gait CFs, and we thus expect that identifying and extracting more correlated temporospatial gait CFs will improve the current model's outcome in future studies. This improvement in the outcome depended on how these new temporospatial gait CFs related to the gait cycle quantitatively.

In comparison, DL has been shown to have a high accuracy in several medical fields [80]. It also does not require feature engineering as in CML. At the same time, DL requires a large amount of data to train and DL models lack explainability, which might be concerning in medical fields. However, our promising results comparing DL-RAW with CML-CF approaches should encourage researchers to conduct further research so that we can transfer the knowledge yielded from DL-RAW to improve the existing temporospatial gait CFs and, and conversely, use temporospatial gait CFs to aid in interpreting DL-RAW results.

4.4.2. CML-CF Approach.

4.4.2.1. *Extracted Gait Features are Consistent with Clinical Observations.* It is commonly known that in people with DMD, the temporospatial gait characteristics of speed, step frequency, and step length are on average lower than those in TD peers. We extracted temporospatial gait CFs from signals derived from a single mobile phone-based triaxial accelerometer using methods similar to those described by Barthelemy [9], and our data demonstrated that across a range of commonly-attained speeds, our extracted gait features differ between DMD and TD controls with many differences reaching statistical significance.

Analysis of power spectra of time-series gait data using single sensors during walking has been explored in comparison to some mobility-limited human populations but has not, to our knowledge, been previously applied in a DMD population. The series of papers by Barthelemy [9] [8] using single sensors to measure gait characteristics of dystrophic GRMD dogs and to evaluate extracted temporospatial gait CFs using LDA methods demonstrated the utility of such methods and provided inspiration for our approach. The overall decreased vertical percent power and increased lateral power that we observed appear consistent with clinical observations of the development of a lateral, trendelenburg gait pattern in children with DMD as described by D'Angelo [28]. In TD individuals, as walking speed increases and progresses to running, ground clearance for the foot and leg in

the swing phase is achieved through a gradual but proportional increase in vertical movement. In people with DMD with progressing weakness, however, swing phase leg clearance is achieved through substitution using a more pronounced lateral shift of the center of mass to the stance phase leg with the elevation of the contralateral hip. This more lateral gait style is effective but also less efficient, and results in greater work for reduced forward motion as demonstrated by prior studies showing increased heart rate-based energy expenditure of ambulation with increasing step frequency in DMD children measured by COSMED portable metabolic testing combined with StepWatch activity monitoring [77].

4.4.2.2. *Interpreting extracted clinical features.* A major challenge with the use of CML approach in evaluating health status is to interpret model outputs relative to well-known temporospatial gait CFs of a disease. Here we have built on our colleague's work in GRMD dogs to extract similar clinically-salient temporospatial CFs of gait from our accelerometer data in humans to help explain differences between our DMD and TD cohorts. In addition, we evaluated our participants using common quantitative timed motor performance tests (25 meters, 100 meters fast-walk/jog/run tests, and the 6MWT) and the NSAA, which all have demonstrated utility as outcome measures used in clinical care and clinical trial contexts.

We do not show correlations between timed motor performance measures and extracted clinical features because the features themselves (velocities, step lengths, and step frequencies) are derived directly from the test's times and distances (Table 4.4). NSAA, however, is not a feature in our models and serves as an external "anchor" to help with the overall interpretation of model performance and to provide additional cross-validation of step length.

At near-TD levels (NSAA ≥ 30) near the ceiling (Figure 4.7), children with DMD show reduced height-normalized step length compared to TD peers at all but the slowest of walks. The most dramatic differences occur in the fast walking and jogging or running paces. Looking in more detail at the DMD children (Figure 4.8), we can see significant differences in all gait paces except in SC-L1 pace. At the same time, we fail to see any significant differences between step lengths for fast walks and jogging or running paces. This supports the clinical observation that some children with DMD fail to develop a "double-off" running pattern that allows TD individuals to lengthen their step lengths well past what they can achieve at a fast walk (Figure 4.10).
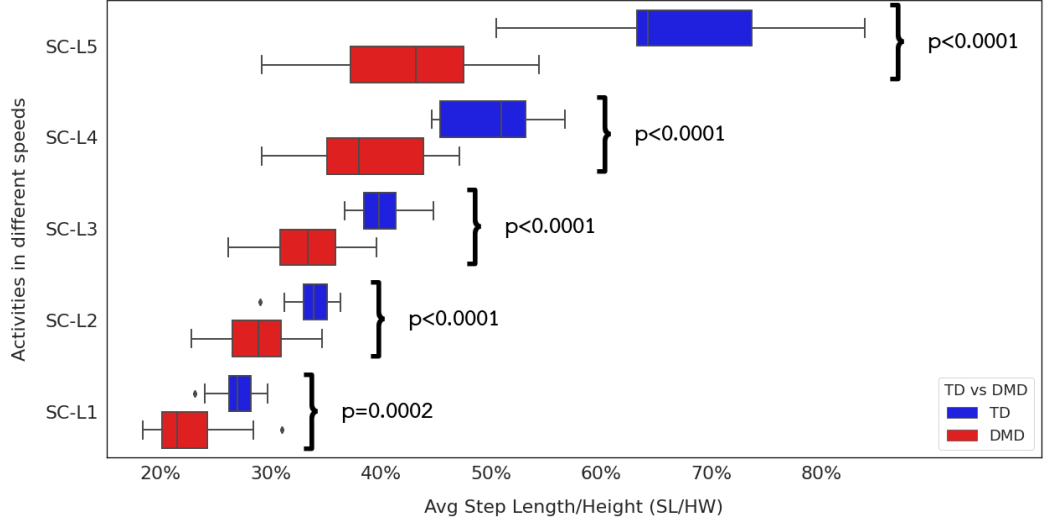
59

FIGURE 4.10. **Step Length by Different Gait Speeds.** TD vs. DMD children. This boxplot displays the data distribution based on: minimum, first quartile (25th percentile), median, third quartile (75th percentile), and maximum, as well as the outliers.

4.4.2.3. *Utility of CML Models as Classifiers.* CML with LDA dimensionality reduction (CML-LDA) surpasses both the original CML (without dimensionality reduction) and CML-PCA in the self-selected-walk to run speeds. For example, CML-LDA achieves a maximum accuracy of 100% in SC-L3 and exceeds 96% in 6MWT and 100MRW.

Our results indicate that gait speed is an essential component in single accelerometer gait analysis [**37**] and that changes in gait speed can affect classification accuracy. We believe the observation that speed and accuracy are correlated would encourage researchers to consider gait speed's effect on gait characteristics. In this respect, CML approaches seem to rely on elements that are highly affected by gait speed, e.g., the observed differences in lateral movement or shortened relative step length that are apparent to the eye and thus picked up by domain experts. It would be interesting to further investigate if additional CFs could be developed and used to improve classification accuracy.

4.4.2.4. *CML with Dimensional Reduction.* Using PCA and LDA provides a 1D and 2D representation of the DMD and TD group's distribution. We reduce the dimensionality of the temporospatial gait CFs to 2D using PCA and 1D using LDA. Both TD and DMD participants' temporospatial gait CFs after PCA and LDA reduction tend to form different groups. LDA (supervised) maximizes the separability between classes, and PCA (unsupervised) maximizes the variance within the classes. The resulting visual representation of these group separation methods provides valuable feedback about model performance as well as a degree of difference between groups, even in the presence of complex, multi-dimensional data. When the distance between the groups becomes more obvious, the two groups tend to be separable. This gives an indicator that the model would yield a high classification accuracy. On the other hand, when the two groups intersect, which indicates high similarity between the two groups, DMD and TD, the classifier yields low accuracy. In CML-LDA (Table 4.3), we notice that the model's accuracy is best in the self-selected-walk to fast-walk speeds (SC-L3 and 6MWT), and fast-walk/jog/run speed (100MRW). We can see that the accuracy of the CML-LDA overall performs better than both the original CML and CML-PCA models. In 6MWT and 100MRW activities, we observe that both the original CML and CML-PCA reach their highest accuracy of 89.66% and 92.31%, respectively. Therefore, CML-LDA is more accurate and provides a helpful visual representation in the self-selected-walk to fast-walk speeds (SC-L3 and 6MWT), and fast-walk/jog/run speed (100MRW) activities which aids to better understand the differences in gait patterns of the participants.

4.4.2.5. *Correlation of PCA and LDA models with clinical features.* PCA-based classification tools perform slightly worse than LDA-based approaches. However, the 2-major components of PCA results (PC1 and PC2) are interesting in that the first factor correlates well with temporospatial gait CFs of height-adjusted step velocity, step frequency, and total power. The other one correlates more strongly with proportional axial portions of power in the vertical, mediolateral, and anteroposterior directions. Viewed from the perspective of progressive DMD symptoms with increased weakness, shortening overall step lengths, and more lateral gait, the selection of these two sets of features as drivers of differences between DMD and control groups makes clinical sense. The LDA-based approaches, however, appear more accurate in differentiating between DMD and TD children. From the perspective of the correlation of LDA location scores and clinical features,

61

it is not surprising that the model outputs correlate most with height-normalized step length as an indicator of typical walking patterns versus those affected by the DMD diagnosis (Table 4.4) and that the models are able to differentiate between groups across a broad range of self-selected velocities (Figure 4.9).

4.4.2.6. *Impact of reduced stride length on community mobility.* Besides differentiating patients with DMD from similarly-aged TD children, reductions in stride length are interesting for another reason. Children with DMD with mild to moderate gait disturbance can take similar numbers of steps during most days relative to TD peers – a key reason that step-counting devices have been less informative as outcome measures for clinical trials [71]. However, the incorporation of distance traveled (whether adjusted to standing height or not) may provide us with valuable additional information about the effect of step length reduction on overall community travel. For instance, a recent meta-analysis by Conger et al. [18] suggests an expected average daily step count of approximately 11,000 steps in TD children (mixed male/female). If we assume that a 50th percentile height American 9-year-old is 128 centimeters tall [60], and that their activity is primarily a comfortable self-selected-walk, we can expect based on our control data for step length to be approximately 40% of their standing height, or 51.2 centimeters, and that their total daily distance traveled would be 5,632 meters. A similarly-sized mildly-affected ambulatory child with DMD would have a step length of approximately 35% of their standing height, or 44.8 centimeters and their total distance traveled would be 4,928 meters, an overall reduction of 704 meters (12%) traveled per day compared to their TD peer. If that same DMD child at a future date was more severely affected because of disease progression, holding their step count and height equal, their step length might be 30% of their standing height, or 38.4 centimeters and their total distance traveled would be 4,224 meters, an overall reduction of 1,408 meters (25%) per day compared to their typical peer and 704 meters (14.2%) less compared to their own prior performance. Granted, this brief exercise makes some unlikely assumptions, including that step counts would match, that all travel would be at a comfortable self-selected pace, that our DMD child did not grow between testing events. Clinical observations would tell us that there is also a progressive impact of fatigue that reduces overall step counts and active time [77], and that linear growth is often reduced in children with DMD both due to disease and due to the effects of steroid treatments. These factors might

further reduce daily travel, or they might cancel each other out. Further research on community travel patterns and behaviors in people with DMD will be required if these questions are to be answered to most readers' satisfaction, but it is not difficult to imagine that reduced stride lengths could noticeably limit a person's community-level travel and participation in daily activities.

4.4.3. DL-RAW Approach.

4.4.3.1. *Evaluation of Raw Data Using DL Models.* DL models (supervised in our study) depend heavily on input data. When we apply the DL approach to raw accelerometer data (DL-RAW), the model achieves a maximum accuracy of 86.67% in slow-walk speed (SC-L2) and self-selected-walk speed (FW). The advantage of using DL-RAW is this method does not need feature engineering. We believe a further investigation into improving the ML algorithms that target raw data could be a promising direction. Still, one challenge is that they are still more difficult to explain from a clinical perspective than methods anchored to clinical features and evaluations. Therefore, when the clinical explanation is a primary focus, it is favorable to extract temporospatial gait CFs that could help understand how the decisions are made in ML models. Our results indicate that gait speed is an essential component in single accelerometer gait analysis [37] and that changes in gait speed can affect classification accuracy.

4.4.3.2. *Time-windowing of Raw Data.* By examining different TW sizes, we found that the optimal TW size should be long enough to include a portion of the gait cycle where contrasts are expected between DMD and TD children. The model determines per each single TW whether it belongs to DMD or TD children by identifying the difference in gait patterns among the DMD and TD groups. This allows us to simplify our methods and use the DL method on minimally processed raw data in a manner that requires less expertise to extract temporospatial gait CFs. Having enough samples in each TW ensures a solid correlation between the x,y, and z-axis and helps the model correctly classify. Using a small TW that does not have enough correlation between the three axes could lead to insufficient data that cannot capture gait patterns. On the other hand, using a large TW results in a smaller number of TWs per participant. At different velocities, people with DMD and TD have similarities in some portions of their gait cycles, and even at faster velocities that are more difficult for people with DMD to achieve, some portions of gait cycles may appear more typical than others. By examining the typical/atypical decisions for each TW as

mentioned in section 4.2.4.4, it may be possible to use that percentage of typical to atypical portion of each gait cycle across a range of speeds to indicate the severity of disease at a given point in time and to track and quantify disease progress over time.

4.4.4. Effectiveness of CML and DL Models Differs Depending on Gait Velocity and Type of Gait. Using the CML-CF approach performs better than the DL-RAW. Overall, DL-RAW accuracy increases with the length of the activity. Since SC-L2 is a slow-walk and long-period activity compared with SC-L3 to SC-L5, DL-RAW achieves an accuracy of 86.67%. In FW, where participants walk as fast as possible for a long duration, DL-RAW in this activity achieves an accuracy of 86.67% while CML-CF exceeds an accuracy of 93%. Additionally, the running efforts on the 25 meters course (SC-L5) are of short duration, which further reduces available training data. In 100MRW, where the duration of the activity is longer and provides more data to the model, the model achieves an accuracy of 84.62% despite the fact that only some participants run while others walk.

Study Limitations. The main limitation of this study is the small number of participants. Our ongoing work aims to expand the scope of this experiment to a larger group of participants spanning multiple diagnoses and age groups. An additional limitation is the short-term nature of our data collection. Multi-day data will be required to demonstrate an actual reduction in travel distance or the ability to differentiate between DMD and typical-developing groups in free-living conditions.

4.5. Conclusion

Use of ubiquitous and widely available mobile devices with single accelerometers to remotely measure differences in common clinical gait parameters represents an opportunity to expand the study of temporospatial gait characteristics into the community setting. Our initial laboratory-based studies demonstrate ability to measure selected gait parameters across a range of typical ambulatory velocities in DMD and TD children to detect significant differences in temporospatial gait CFs that are consistent with previous studies, as well as to detect differences in proportional power of accelerations in vertical, mediolateral and anteroposterior axes. By using these clinical gait parameters and raw data and employing both CML and DL models, we are able to correctly

predict whether sensor data is derived from children with DMD up to 100% of the time at self-selected-walk pace (SC-L3). By combining these approaches, we anticipate that through ongoing studies, we will be able to improve predictive accuracy and identify additional clinically-useful parameters indicating typical growth and development, gait impairment and disease progression across a wide range of individuals with neuromuscular disease.

CHAPTER 5

BWCNN: Blink to Word, a Real-Time Convolutional Neural Network Approach

5.1. Introduction

A plethora of clinical conditions, such as brain trauma and amyotrophic lateral sclerosis (ALS), cause damage to the central neural system (CNS) or brain, in such as way that the ability of speech and motor functions cannot be sustained. In those cases, the ability to communicate is limited, to non-verbal forms of communication, such as eye blinking.

Certain approaches have been used in the past to solve this problem. Researchers in [85] use Infrared (IR) sensors to estimate the state of the eyes ('Open' or 'Closed') in order to detect blinking, which is then converted to Morse code. Challenges in this approach include the IR sensor being irradiated by other sources resulting in false eye-blinks, and the risk of cataract formation in the case of prolonged use. [85]. In another study, a traditional computer vision-based system that detects users' eye-blinks, measures their duration, and interprets the blinks in real-time was proposed [40]. Although the system has achieved an accuracy of 95.6% in ideal conditions, traditional image processing techniques are prone to failure under limited lighting conditions, arbitrary changes in image texture, changes in user's pose, among others, which limits their real-time accuracy, and make these systems not robust in real applications.

A wearable device to detect eye-blinks for alleviating dry eyes was proposed in [24]. This method captured 85.2% of all the blinks that occurred during testing. But, the IR sensors tend to show false readings when the orientations are altered and hence, are unreliable in realistic scenarios. Any facial movements such as laughing, or yawning can induce errors.

To avoid these shortcomings, we here present a deep-learning vision system that detects eye-blinks and maps them to words in real-time. Our system uses the InceptionV3 [108] architecture and achieves an accuracy of 99.20% with a latency of 94.1ms on IoT devices. Our method is

safe to use, has better performance than previous methods, it is robust to changes in lighting conditions and facial orientation, and its architecture is modular so its output can be mapped to other tasks, such as controlling software applications or devices. The main contribution of this work resides in designing, implementing, and evaluating the first deep-learning solution for eye-blink communication with performance, latency, and safety specifications that can be used in a real-world environment.
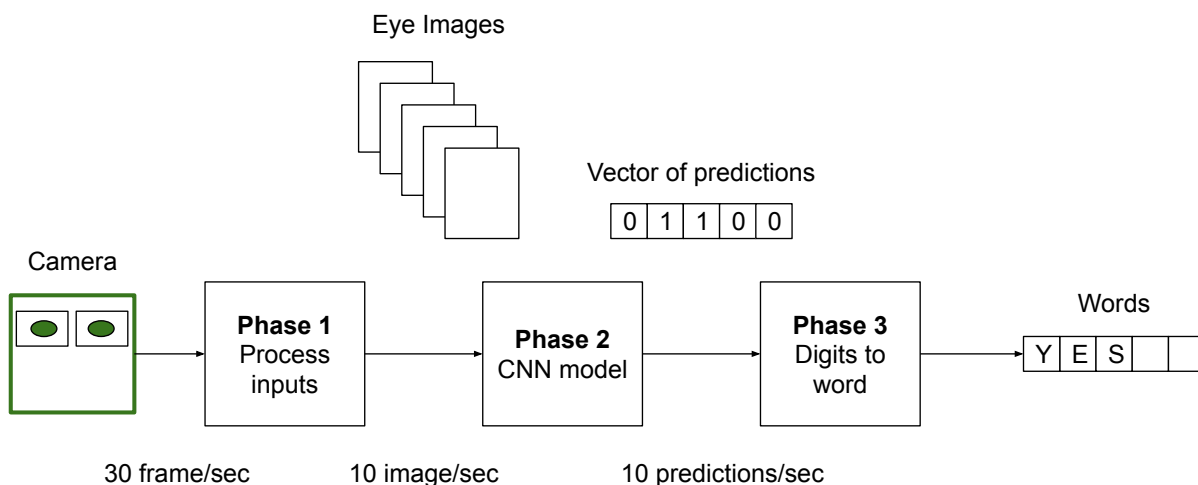

FIGURE 5.1. The 3 phases of the BWCNN system.

5.1.1. Previous work. An efficient system for eye-blink detection is presented in [**82**]. This method uses Haar-cascade classifiers for face detection and eye positions. The performance of Haar-cascade classifiers is not invariant to the change in lighting conditions. Hence there is a decay in performance.

A low-cost implementation of an eye-blink-based communication aid for ALS patients is presented in [**106**]. Template matching is used to track the eye and detect eye-blinks using hierarchical optical flow. The implementation has an accuracy of 94.75% during the typing test. However, the algorithm takes approximately 2 seconds to generate a single scan of the eye. This is excessive for a single character.

67

This paper [104] presents a real-time detection and classification between eye-blink (with both eyes), left wink, or right wink. with 96, 92, and 88% accuracy. The latency for the detection of a single blink was 250ms.

A wearable device to detect eye-blinks for alleviating dry eyes was proposed in [24]. This method captured 85.2% of all the blinks that occurred during testing. But, the IR sensors tend to show false readings when the orientations are altered and hence, are unreliable in realistic scenarios. Any facial movements such as laughing, or yawning can induce errors.

5.2. Design and Implementation

Our system detects the state of eyes, 'Open' and 'Closed', even under poor lighting conditions. We have a pre-defined set of patient inputs corresponding to the blink pattern, which we map to actions in real-time. These inputs could correspond to movements (up, down, left, right), clicks, etc., which would enable patients to use different applications (browser, email, etc.) or devices (mouse, keyboard, etc). As a proof-of-concept, we mapped these inputs to specific words. Since we use predefined words instead of Morse code or other encoding patterns, it is simple for patients to spell out a sentence. We want to create a system that works almost flawlessly in real-time and is safe to use. This is represented by equation (1). $\mathcal{P}$ refers to the performance of the system, which in our case is the accuracy. $\mathcal{S}$ refers to the System parameters. $\mathcal{W}$ refers to the weights of the neural network. $\mathcal{A}$ refers to the architecture. $\mathcal{H}$ is the set of all architectures that can be used for this purpose. $\mathcal{T}$ refers to time constraint which is the prediction time (response time, or latency) for the model on the validation set. It is important to reduce the response time of the system since it will be running over a real-time IoT device such as Raspberry Pi.

(5.1)
$$\max_{\mathcal{A} \in \mathcal{H}} \mathcal{P}(\mathcal{S}, \mathcal{W}, \mathcal{A})$$

$$s.t \; prediction \; time \leq \mathcal{T}$$

The goal of this approach is to detect if the patients blink their eyes and to map the sequence of blinks to a particular entry in a dictionary of words. In order to achieve this, one has to detect the state of eyes ('Open' or 'Closed'). If an 'Open' state is followed by a 'Closed state, the system

detects an eye-blink, as shown in Figure 5.2. The system is divided into three phases as shown in Figure 5.1.

5.2.1. Phase 1: Capturing and saving a stream of frames. In this section, we present the method of obtaining the data and preprocessing it to be given as input to the ConvNet. The system uses a camera device attached to IoT for capturing the frames. Regular webcams are capable of capturing about 30 frames per second (fps). The fps will directly affect the user experience. A higher fps will increase the latency.

Since the system runs in real-time, it is more effective to reduce the latency. At the same time, using lower fps can lead to missing a blink. Our experiment shows that 10 frames per second are a reasonable frame rate for real-time application. As there is a frame being captured every 100ms, our model must predict each frame in less than 100ms to avoid delay. We further impose a constraint on the user that the 'Closed' state should be maintained for at least 200ms. The system saves each frame as a gray-scale image of dimensions 80×70 pixels. Each image is 2KB in size.

5.2.2. Phase 2: Predict the content of the image. In this section, we present the experiments to choose the best fitting neural network architecture to predict the state of the eye. We compare four state-of-the-art architectures (SqueezeNet [**47**], ResNet [**43**], InceptionV3 and DenseNet architecture). Both the architecture and the hyperparameters play a large role in model performance. We start by training the networks from scratch for different batch sizes and find the batch size that gives the best results. We further explore the chances of improving performance by using transfer learning.

5.2.2.1. *Training Dataset:* We used the eye dataset from Media Research Lab (MRL). The dataset contains 84,898 images of eyes taken from 37 individuals consisting of 33 men and 4 women. Each image in the dataset was collected from one of the following sensors: Intel RealSense RS 300 sensor with a resolution of 640×480, IDS Imaging sensor with a resolution of 1280×1024, and Aptina sensor with a resolution of 752×480. The original dataset contains 6 classes: 'gender', 'glasses', 'eye state', 'reflections', 'lighting conditions', and 'sensor resolution'. We split the dataset into an 80% training set and a 20% test set.

69

TABLE 5.1. ResNet with and without transfer learning

Batch S.	Epoch	Training a model (from scratch)		Transfer learning (from our model)		Transfer learning (from official)	
		Acc. (%)	Ep.imp.	Acc. (%)	Ep. imp.	Acc. (%)	Ep. imp.
8	100	99.21	32	99.22	31	99.22	29
16	100	99.26	55	99.23	51	99.17	16
32	100	99.22	48	99.22	49	99.17	25

TABLE 5.2. DenseNet,SqueezeNet,InceptionV3

Batch S.	Eep.	DENSENET		SQUEEZENET		INCEPTIONV3	
		Acc. (%)	Ep. imp.	Acc. (%)	Ep. imp.	Acc. (%)	Ep. imp.
8	100	99.24	55	49.40	1	99.14	35
16	100	99.18	70	49.40	1	99.20	22
32	100	99.21	52	49.40	1	99.17	38

5.2.2.2. *Training experiments:* Two important considerations when training the model were accuracy and latency. Latency for detection is the time taken to make an accurate classification. To find the model which can provide the best accuracy with the least latency, we implemented SqueezeNet, ResNet, InceptionV3, and DenseNet.

5.2.2.3. *Train ResNet architecture from scratch:* We trained the ResNet architecture for 100 epochs using 6 batch sizes. The performance metric used is the overall accuracy. Comparing the performance, we selected the three best batch sizes from 1, 2, 4, 8, 16, and 32. Table 5.1 shows that batch sizes 8, 16, and 32 provide the best accuracy. Since our experiment stops at 100 epochs, training the network for more epochs might improve the performance. To test this, we ran ResNet, for all batch sizes for 500 epochs. The results show that there is no further improvement in accuracy.

5.2.2.4. *Transfer learning:* We took the weights from the most accurate architecture, ResNet with batch size 16, and transferred these weights to all other ResNet architecture with the 3 best performing batch sizes. We also did transfer learning by using weights from the official pre-trained ResNet on our ResNet models with the 3 best performing batch sizes. As before, after 100 epochs, there was no significant improvement in the accuracy of the network.

# blinks	1	2	3	4	5	6	7
Pattern	1	101	10101	1010101	101010101	10101010101	1010101010101
Words	Yes	No	Hi	I am	Good	Thanks	How are you?
Mouse	Right	Left	Click R.	Click L.	Up	Down	Hold click
Keyboard	Tab	Enter	Back	Esc	Scroll up	Scroll down	Space

5.2.2.5. *Train InceptionV3, SqueezeNet and DeneseNet architectures from scratch:* We assume that the 3 best-performing batch sizes from ResNet would be the best performing in the other architectures as well. To investigate this assumption we run the same experiment again but with DenseNet, Inception, and SqueezeNet as shown in Table 5.2.

5.2.3. Phase 3: Mapping. Our system stored the output of the neural network as a vector of 0s and 1s, where Zero represents the 'Open' state for the eye and a One represents the 'Closed' state as shown in Figure 5.2. We normalize the vector by truncating repeated instances of a state with a single instance. For example, vector 00000110000 becomes 010. The system provides a special word to start and end a session of blinking, which is keeping the eyes closed for 4 seconds. The system recognizes the end of a word when the patient's eyes remain open for 1 second. Based on the output vector, the number of blinks is calculated and is mapped with the words in the dictionary (Table 5.3). The dictionary consists of basic words that we use as a proof-of-concept and it can be modified.

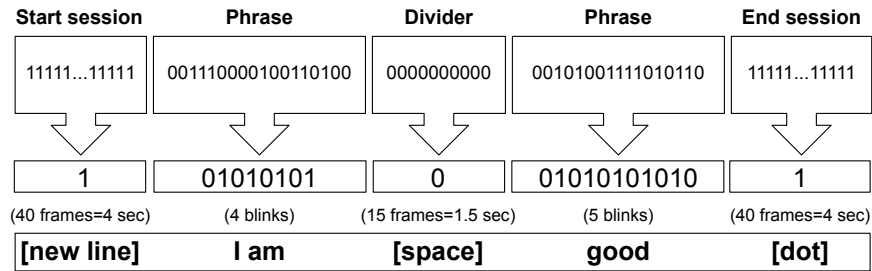| Start session	Phrase	Divider	Phrase	End session
11111...11111	001110000100110100	0000000000	00101001111010110	11111...11111
1	01010101	0	01010101010	1
(40 frames=4 sec)	(4 blinks)	(15 frames=1.5 sec)	(5 blinks)	(40 frames=4 sec)
[new line]	**I am**	**[space]**	**good**	**[dot]**

FIGURE 5.2. BWCNN system predicting a series of frames in the real-time

5.3. Results

This section provides the performance results of the different architectures. Table 5.1 lists the results obtained from training ResNet for 100 epochs on batch sizes 1, 2, 4, 8, 16, and 32. Training ResNet from scratch for 100 epochs on a batch size of 16 yields the best accuracy, 99.26%. There is no improvement in accuracy after 55 epochs.

TABLE 5.4. Final results: 3.1 GHz Quad-Core Intel Core i7, 16 GB 2,133 MHz

Model Archit-ecture	Batch size	Ep. imp.	Total params	Trainable params	Non-trainable params	Model size	Model Accu-racy	Avg latency
ResNet	16	55	23,591,810	23,538,690	53,120	283MB	99.26%	117.28ms
DenseNet	8	55	7,039,554	6,955,906	83,648	85MB	99.24%	146.09ms
SqueezeNet	16	1	723,522	723,522	0	8MB	49.40%	13.64ms
InceptionV3	16	16	21,806,882	1,772,450	34,432	262MB	99.20%	94.1ms

We trained a new ResNet model using transfer learning. We took the weights from the best performing network that we had trained form scratch and used them to train a ResNet network with batch sizes of 8, 16, and 32. We also did transfer learning with weights obtained from the official pre-trained ResNet architecture and used it to train ResNet architecture with batch sizes of 8, 16, and 32. The results are shown in Table 5.1.

Table 5.2 presents the results for DenseNet, SqueezeNet, and InceptionV3 architectures for the batch sizes of 8, 16, and 32. Figure 5.4 shows the training and loss curve for InceptionV3. DenseNet has the best accuracy of 99.24%, which is achieved for a batch size of 8, as training for 55 epochs, after which it starts over-fitting.

SqueezeNet seems to be the worst among all the architectures with an accuracy of 49.40%. There is no change in accuracy after the first epoch, which is in Table 5.2. The final comparison of the best results across the different architectures in Table 5.4. Apart from the accuracy of the networks, Table 5.4 also contains the latency (in milliseconds). We can see that ResNet has the best accuracy of 99.26%, but has a high latency of 117.28ms. InceptionV3 has an accuracy of 99.20% which is close to ResNet, but with a lower latency of 94.1ms. We can see that InceptionV3,
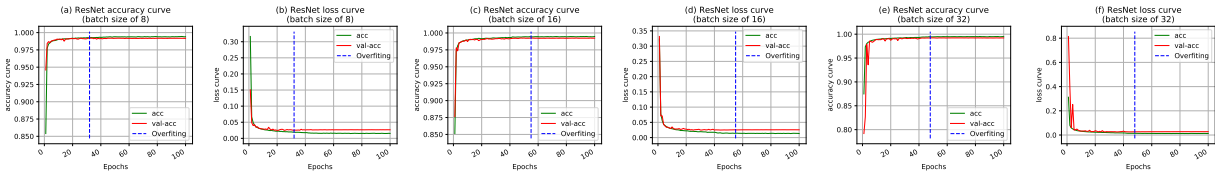
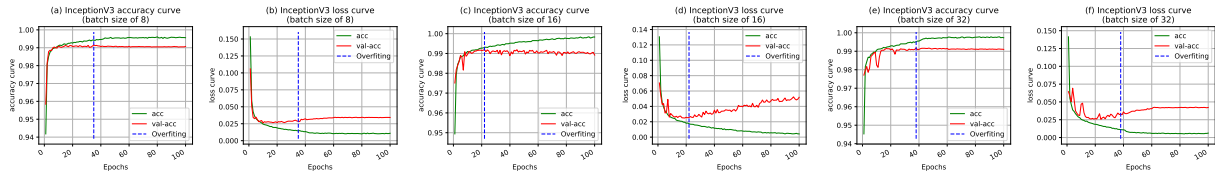FIGURE 5.3. ResNet: training and validation (accuracy and loss) curves.

FIGURE 5.4. InceptionV3: training and validation (accuracy and loss) curves.

DenseNet, and ResNet have similar accuracies but the InceptionV3 model has the lowest latency. Thus, we chose InceptionV3 architecture.

5.4. Conclusion

In this work, we designed an AI system for non-verbal communication that converts eye-blinks to words using a deep learning CNN architecture. The system predicts the state of the eyes of the patient and finds the blinking pattern. We compared several CNN architectures and hyper-parameter selections in model training. For the evaluation, we tested our system using 16,979 facial images and found that our proposed prediction model was efficient and effective. Results demonstrate that overall prediction accuracy is 99.20% and the average prediction time is 94ms. We trained different architectures with different hyperparameters to identify parameter combinations that lead to high accuracy and low latency. For the sake of conducting a clear comparative analysis, we compare the results of each architecture with batch sizes of 8, 16, and 32. SqueezeNet received the lowest accuracy with the fewest parameters. The DenseNet, ResNet, and InceptionV3 acquired accuracies in the range of 99.20% and above. As InceptionV3 had the lowest latency, we chose this architecture. We introduced transfer learning, which improved the convergence when compared to random initialization, to otherwise similar accuracy and latency in the response. Future work includes the training on a more generalized training dataset, the application of hybrid

technologies that fuse computer vision techniques such as the one presented together with other natural language processing methods, including recurrent neural networks, to introduce memory in the actions and system. The dataset, source code, demo, and results of this system are available at `https://albara.ramli.net/research/bwcnn/`

CHAPTER 6

A Deep Ensemble Learning Approach to Identify Stroke-related Features of Gait

Stroke is the leading cause of disability in adults worldwide, affecting movement, speech, vision, and cognitive function [17, 29, 42, 51]. Most stroke survivors experience walking dysfunction since it affects motion control of the entire left or right side of the body [10, 89]. Thus, rehabilitation therapy after a stroke is important for patients to regain walking function. It is informative to the rehabilitation process to accurately identify and assess gait abnormality to better understand the impact of stroke on mobility for post-stroke survivors. Gait studies are often conducted in motion capture laboratories where a subject's movement is precisely measured and abundant biomechanical data quantified (i.e., kinematics and kinetics) [83]. This abundant data inspires the development of automated gait analysis systems that can explore and observe the relationship between these biomechanical signals and identify which signals are mostly revealing in terms of gait abnormality.

In this work, we hope to take a data-driven approach to study the vast volume of data generated at motion labs toward the goal of systematic and automated gait analysis and potentially in broader settings outside motion labs. We start by addressing the following questions:

- **Q1**: Can we accurately identify the gait of post-stroke survivors when compared to age-matched controls using machine learning (ML) algorithms?
- **Q2**: Can we accurately identify gaits of more vs. less affected sides of post-stroke survivors using ML algorithms?
- **Q3**: In the above tasks, what are the most discriminative biomechanical signals using ML algorithms?
- **Q4**: How important is it to include ground reaction force (GF) data, generated from force plate (FP), in the above tasks?

Answering Q1 and Q2 is the first step towards **automated** gait monitoring using machine learning algorithms. Such automation will allow more extensive and less expensive gait monitoring and analysis, e.g., in clinical setting, or even in community settings or homes in a 24/7 manner. Answering Q3 can help researchers and clinicians focus on the most relevant biomechanical signals to the gait abnormality. Answering Q4 allows us to evaluate the possibility of larger-scale community-based data collection and patient care outside the motion lab environment without the force plate, which is much more expensive and difficult to use. To answer the above questions, we extensively study the biomechanical signal sets generated in motion labs, including:

- based on a **single biomechanical signal**, i.e., signal determined by the data source, data positions, and data types, e.g., AngAcc (type) of the AR (position) from KK (source).
- based on **data position** (bases, joints, and segments). Integration of all signals at the position from different data sources, e.g., hip (position) from LMB (source).
- based on **data type**. Integration of a signal type from all signal positions of a data source, e.g., AngAcc (type) from KK (source).
- based on **data source**. Integration of all signal positions and types from a data source, such as (KK, LMB, and FP) or the integration of (KK+LMB+FP).

To conduct the above analysis, this work proposes a machine learning framework based on deep ensemble learning that uses a fusion architecture of convolutional neural networks (CNN) [**91**] [**69**]. We choose deep learning because of its ability to fully utilize the potential of big data to improve performance and its ability to automatically extract patterns from raw signals without feature engineering. We choose ensemble learning to obtain more complete and accurate identification results.

The importance of different biomechanical signals is revealed during the process, as well as the importance of the force plate, as shown in the above examples with details in Sec. 6.4. In summary, our proposed deep ensemble learning framework allows us to systematically analyze gait-related biomechanical signals generated from marker systems and force plates. We demonstrate the feasibility of ML tools in such tasks (Q1 and Q2) as well as answering clinically relevant questions quantitatively (Q3 and Q4). To the best of our knowledge, we are the first to thoroughly and

systematically studying the above questions and data combinations in the field of gait analysis. Our work is a positive step toward the goal of automated gait analysis.

6.1. Related Work

There exists a significant amount of literature on gait analysis and re-habitation research under various conditions, such as stroke recovery [**53**, **63**], Parkinson's disease [**32**, **110**], aging population [**58**, **72**], and cerebral palsy [**116**], using diverse analytical tools. Here, we focus on machine learning (ML) based approaches, including both classic machine learning (CML) approaches and deep learning (DL) approaches. CML algorithms have been more widely used until very recently due to their model and computational simplicity, better interpretability, and faster training. Such algorithms require feature engineering, i.e., selecting a set of preprocessed clinical features (e.g., stride length, variance in angle velocity). In other words, it assumes that researchers are aware of the biomechanical signals that affect the gait pattern of interests in advance. In comparison, DL algorithms do not require such preprocessed features and can apply directly to raw signals. DL can learn complex patterns from raw signals, improve the overall performance, and achieve better robustness against signal noise, but it is difficult to interpret and, in general, requires more data to train. We discuss related work in both categories below. We note that while our work focuses on biomechanical signals from motion capture systems, the ML techniques used for IMU (inertial measurement unit) sensors are often similar and thus also included in the discussion.

6.1.1. Classical Machine Learning Approaches for Gait Analysis. Khera et al. [**58**] present an overview on ML techniques applied to various aspects of gait research, including gait activity detection, gait event detection, identification of gait disorders due to pathological conditions and aging, gait asymmetry detection, and the influence of neurological factors on gait, from 1980 to 2019. A majority of the studies in the review employ CML techniques along with engineered features. Among these techniques, support vector machines (SVM) are frequently employed and consistently achieve high accuracy. Similarly, Figueiredo et al. [**31**] focus primarily on CML algorithms and also emphasize the effectiveness of SVM in addressing gait-related problems. Chau et al. [**12**] survey earlier research involving shallow neural networks and wavelet methods. Diaz et

al. [26] survey recent studies that utilize wearable sensors, such as IMUs, smart devices, electromyography sensors, insole pressure sensors, and force sensors, in gait, balance, and range of motion research. Their survey demonstrates that ML plays a crucial role in these studies, often achieving the highest accuracy for parameter quantification and population or condition classification. We discuss a few representative examples below and refer readers to the above-mentioned survey papers for more references.

Lau et al. [63] apply SVMs to classify five different walking conditions of hemiparetic patients with unilateral dropped foot (post-stroke) using data from two IMUs, each comprising of an accelerometer and a gyroscope. The study involves seven participants under varying walking conditions. Engineered features, such as anteroposterior acceleration and angular velocity of the shank and foot segments, are used. The results highlight the potential of SVMs in differentiating between distinct walking conditions based on IMU data. In [20], an SVM fusion algorithm is utilized to classify post-stroke hemiparetic gait from normal gait. The study employs thirty-eight reflective markers with six cameras (MT), ground reaction force (GRF), and electromyogram from 42 post-stroke and healthy subjects while they walk. The data input is processed using principal component analysis to reduce the dimensionality.

Williams et al. [116] employ a multiclass SVM for the classification of six distinct traumatic brain injury gait patterns, ranging from spastic hemiparesis to ataxia/dyspraxia-bilateral. The authors extract 48 statistical features from four joint-angle curves (pelvis, hip, knee, and ankle) in three planes of motion using data obtained from ViCon motion analysis systems for individuals with cerebral palsy.

Jimenez et al. [50] utilize KNN, SVM, and LDA algorithms to classify subjects with and without chronic neck pain. The study employs IMU-based gait kinematic biomarkers from 39 subjects performing two gait trajectories: linear walking with their head straight (single-task) and linear walking with continuous head rotation (dual-task). The temporal (time) and spectral (frequency) kinematic features are extracted and used for analysis.

In [110], Trabassi et al. compare the performance of five CML algorithms, including SVM, artificial neural network, decision trees (DT), random forest (RF), and K-nearest neighbors, in classifying people with Parkinson's disease (pwPD) from speed-matched healthy subjects (HS),

based on a selected set of IMU-derived gait features, extrapolated from the trunk acceleration patterns of 81 pwPD and 80 HS.

Lugade et al. [72] use an Artificial Neural Network (ANN) to study the gait balance control of elderly adults through analysis of the center of mass, using an eight-camera motion capture system and 29 retro-reflective markers attached to each participant. The study involves 56 elderly adults who walk at a self-selected speed along a 10-meter walkway.

6.1.2. Deep Learning Approaches for Gait Analysis. Alharthi et al. [4] review recent work on DL for human gait monitoring using various data sources, including video sequences, wearable sensors, floor sensors, and multi-sensor, multi-modal fusion. The review concludes that DL generally outperforms traditional shallow learning models, emphasizing the potential of DL in gait analysis.

Using IMU Data. Wang et al. [112] employ DL models to classify four common gait abnormalities: drop-foot gait, circumduction gait, hip hiking gait, and back knee gait. The study utilizes angular velocity data from the shank on the sagittal plane, collected using IMU sensors from eight stroke patients and seven control subjects.

The paper [100] proposes three long short-term memory (LSTM) based models to identify gait phases, including foot-off, mid-swing, and foot-contact, using gait data from two tri-axial inertial sensors located above each ankle. Five existing datasets are used to evaluate the performance of the proposed algorithms. In [99], a Convolutional Neural network (CNN) is used to detect gait events, i.e., initial and final foot contact, using data collected from IMUs on the left and right ankle and shank while walking for 5 m at slow, preferred, and fast walking speed. The cohort consists of 157 participants in both healthy and neurologically diseased groups.

In [86], deep CNN models are used to classify whether the subjects are semi-professional athletes, normal participants, or participants with foot structural abnormalities. The study uses 7 IMUs from 69 subjects with either foot abnormalities or athletic performance who are walking on a 20-meter straight path, with raw signals transformed into spectrogram images. In Sharifi et al. [102], multiple deep neural networks are employed to predict 12 specific spatial-temporal gait

79

parameters using IMU data collected from 29 subjects with osteoarthritis and total knee arthroplasty. The study analyzes walking patterns at slow, normal, and fast paces using raw signal data.

In [**101**], a recurrent neural network (RNN) is employed to predict 3D hip and knee joint rotations using synthetically generated IMU data from 13 osteoarthritides and 17 with total knee arthroplasty. The study focuses on multiple gait trials at various speeds, with the synthetic IMU data generated from measured marker positions with 13 cameras and four Bertec force plates.

Using Motion Capture Data. Much less research has been done using DL methods on raw motion capture data sources, discussed below. Horst et al. [**46**] propose the use of layer-wise relevance propagation (LRP) techniques for explaining classifier decisions by decomposing the contributions of each input variable to the overall prediction. In their study, lower-body joint angles and ground reaction forces are used as input data, which are derived from raw motion capture signals. Applying LRP or other explanation techniques in the context of deep ensemble learning represents a challenging and important direction for future research.

Kaczmarczyk et al. [**53**] use several ML techniques, including Cluster Analysis (CA), Single-linkage (nearest neighbor), k-mean, Discriminant Function (DF), and ANN, to classify three types of pathological gait patterns in post-stroke patients. The study employs motion capture with two analogue cameras and 18 markers attached to each patient as they walk unassisted at a self-selected speed along a 10-meter walkway. Kinematic gait parameters are engineered for CA and DF, while the ANN utilizes raw signals of changes in joint angles, specifically knee and hip angles. The study includes 74 stroke patients. While the work mostly focuses on CML algorithms, because it uses neural networks and raw signals, we include it here.

Filtjens et al. [**32**] apply Graph Neural Networks to identify freeze of gait events in Parkinson's disease research. The study uses Vicon motion capture systems that record the 3D movements of nine optical markers. The analysis involves two motion capture datasets with 42 and 17 subjects, respectively.

Moro et al. [**84**] leverage computer vision and deep learning techniques to compare markerless and marker-based systems for 3D gait analysis. The markerless system uses RGB video data from three cameras, while the marker-based system employs eight cameras and 22 infrared passive

markers. The study involves 16 unimpaired participants who walk naturally in a straight line for 6 meters, repeated ten times for a total of 20 trials in an indoor setting. The researchers extract spatio-temporal parameters such as stride length, stride time, stance phase, swing phase, stride width, and speed, as well as joint angles for the central gait cycle. The results provide insights into the differences between the two systems for 3D gait analysis. Similarly, Vafadar et al. [111] develop a deep learning algorithm for a markerless system that estimates the 3D positions of 17 key points or joint centers. The study demonstrates high accuracy on a dataset of 24 asymptomatic adults.

6.1.3. Comparison. Compared with prior work, our research distinguishes itself in several ways:

- We employ a deep ensemble method, a powerful ML tool that often outperforms CML algorithms and single deep learning models (without ensemble). To the best of our knowledge, this is the first work using deep ensemble models on motion capture signals.
- We conduct an extensive evaluation of 215 raw biomechanical data sources from motion capture data, while previous work uses a few such data sources. This extensive evaluation allows us to better understand, compare, and combine different types or sources of gait data.
- We utilize a relatively large dataset consisting of 110 subjects, which is uncommon in previous studies.

6.2. Data Generation

6.2.1. Subjects. This work involves the analysis of gait biomechanics measured in a motion capture laboratory from 110 subjects categorized into two groups: 63 post-stroke survivors and 47 age-matched controls (Table 6.1). All study participants met the following inclusion criteria: 18 years or older, ability to provide consent, follow three-step motor commands, walk 15 meters independently with or without the use of an assistive device or brace, and stable of significant medical complications which could interfere with walking (i.e., osteoporosis, pain, cardiopulmonary disease). Participants provided written informed consent prior to enrollment and participation. Stroke-survivors must be greater than 1-month post-stroke, clinical presentation of at least one

but no more than three strokes on the same brain hemisphere (confirmed by CT or MRI). All procedures were approved by University of Florida Institutional Review Board (UF-IRB-01).

TABLE 6.1. **Subject Demographics.**

Variable	Unit	Age-Matched Control	Post-Stroke
Age	Years	60.49 (±9.6)	62.63 (±10.71)
Gait Speed	m/s	0.933 (±0.221)	0.478 (±0.219)
Biological	Male	23	48
Sex	Female	24	15

Table 6.1 lists the demographic data of the two groups of subjects consisting of age, speed, and gender.

6.2.2. Data Collection. Subjects were instrumented with 14mm reflective markers in accordance with a modified Cleveland Clinic marker set [**14**]. Three-dimensional (3D) marker position data were measured using a 12-camera motion capture system (200Hz, Vicon, Oxford, UK) while participants walked on a split-belt treadmill (2000Hz, Bertec, Columbus, OH, USA), instrumented with two 6-degree-of-freedom force platforms which measured ground reaction force (GRF) interaction with both legs. Subjects were first acclimated to the treadmill, and then the speed was incrementally increased until a self-selected walking speed (SSWS) was identified (Figure 6.1). A minimum of 30 seconds of continuous walking was recorded as the subject walked at their SSWS.

Tracked marker position data was filtered (lowpass 4th order Butterworth, fc=6Hz) and used to solve a full body model, scaled to the subject-specific parameters mass and height (Visual3D v6, C-Motion, Germantown, MD, USA). The model contained 12 linked rigid segments which correspond to bones: pelvis (PV), thorax/abdomen (TA), thigh (TH), shank (SK), foot (FT), upper arm (UA), and forearm (FA). The segments are modeled as 3D shapes (cone) with mass and inertial properties [**25**,**41**]. Two additional foot segments were defined (MF) to measure ankle angle relative to the lab floor.

The 3D position and orientation of the segments were used to define joint angles as the distal segment relative to the proximal segment. The pelvis and trunk segments were also calculated relative to the global reference frame. To analyze the interaction of forces on the segments, inverse
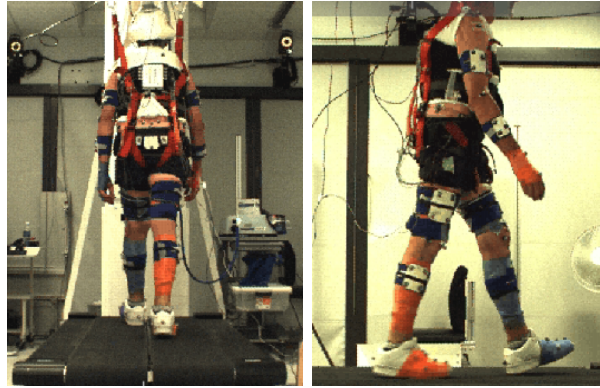
(a) Posterior view.   (b) Lateral view.

FIGURE 6.1. Motion capture cameras track the position of markers attached to a subject who walks on a split-belt treadmill instrumented with two 6-degree-of-freedom force plates.

dynamics methods were applied using filtered force data (lowpass 4th order Butterworth, fc=12Hz) across the ankle, knee, and hip to calculate joint moment and power in the global reference frame.

6.2.3. Data Preprocessing. Consider a subject in a motion laboratory shown in Figure 6.1. Marker positions (MP) and ground reaction force (GF) are collected synchronously for each subject and then processed to generate three main gait data sources: kinematics and kinetics (KK), link-model-based (LMB), and force plate (FP) shown in Figure 6.3. Then, based on positions and types, individual biomechanical signals are generated, i.e., the angular accelerator (AngAcc) from the upper arm (UA). In our experiment, using the raw signals from MP and GF, 205 biomechanical signals are generated for each subject, summarized in Figure 6.3. Examples are shown in Figure 6.2. Each biomechanical signal contains three axes, producing a total of 615 time series signals per subject. This is a huge amount of data, yet representative of gait measurement from a motion lab.

The biomechanical analysis of marker position (MP) and force plate (FP) data generated three data sources considered for these analyses: 1) Force Plate (FP), 2) Kinematics and Kinetics (KK), and 3) Link-Model-Based (LMB) (Figure 6.3). These data are further organized into two subcategories: 1) Data Position (bases, joints, and segments), and 2) Data Types. Thus, in this work, a single biomechanical signal is referenced using three identifiers: 1) Data Source, 2) Data Position, and 3) Data Type. In total, 205 different biomechanical signals are collected per subject. Each signal is measured about three axes (x, y, and z). Therefore, the biomechanical dataset for each

subject in this study consists of 615 time series signals. Figure 6.2 shows a sample of a 10 second time series data set of a 3-axis biomechanical signal from both an age-matched control and a post-stroke study participant. The biomechanical data were uploaded to the automatic gait assess-
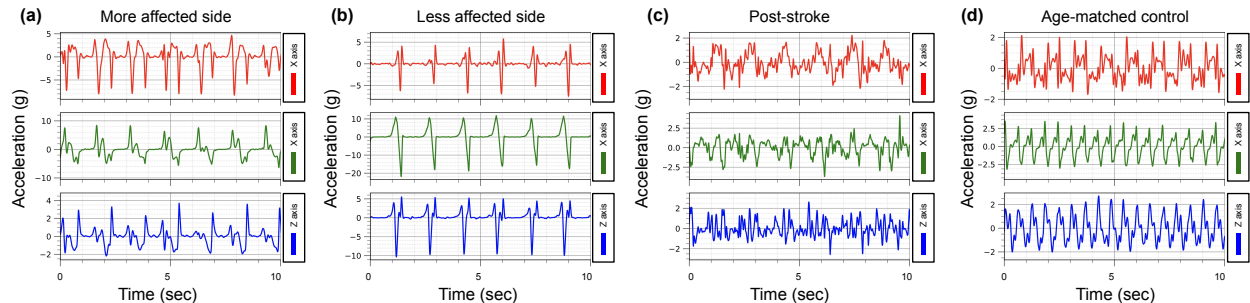

FIGURE 6.2. Among various signals, this signal is 10 seconds sample. (a) and (b) is the foot, kinetics of Linear acceleration in the center of gravity (FT-CGAcc) for the less affected side vs. the more affected side of a Post-stroke survivor. (c) and (d) is the Pelvis of Linear acceleration in the center of gravity (PV-CGAcc) for a Post-stroke survivor vs. Age-matched control.

ment system (Walk4me) [92, 93, 94] for machine learning analysis. The Walk4Me system enables both open-ended and focused exploration by providing both classical machine learning algorithms (with custom clinical feature extraction) and deep learning algorithms (with automated feature extraction).

6.3. Deep Ensemble Learning

In this work, we develop a deep ensemble learning framework to analyze the biomechanical signals collected. There are two main reasons for choosing deep learning. First, deep learning is well known for its capability to scale up the performance with the increase in data input [73]. In comparison, the performance of classic machine learning methods often improves much slower or hits a plateau as the data volume increases. Second, deep learning uses raw signals as input, thus alleviating the need for human-engineered features and, more importantly, has the potential to identify and utilize hidden patterns unknown to human experts for better performance. Furthermore, we choose ensemble learning to obtain more robust and accurate identification results. In addition, when the number of samples (of a single subject) used to train the model is insufficient, the high dimensional data can deteriorate the generalization performance of the model. Here,
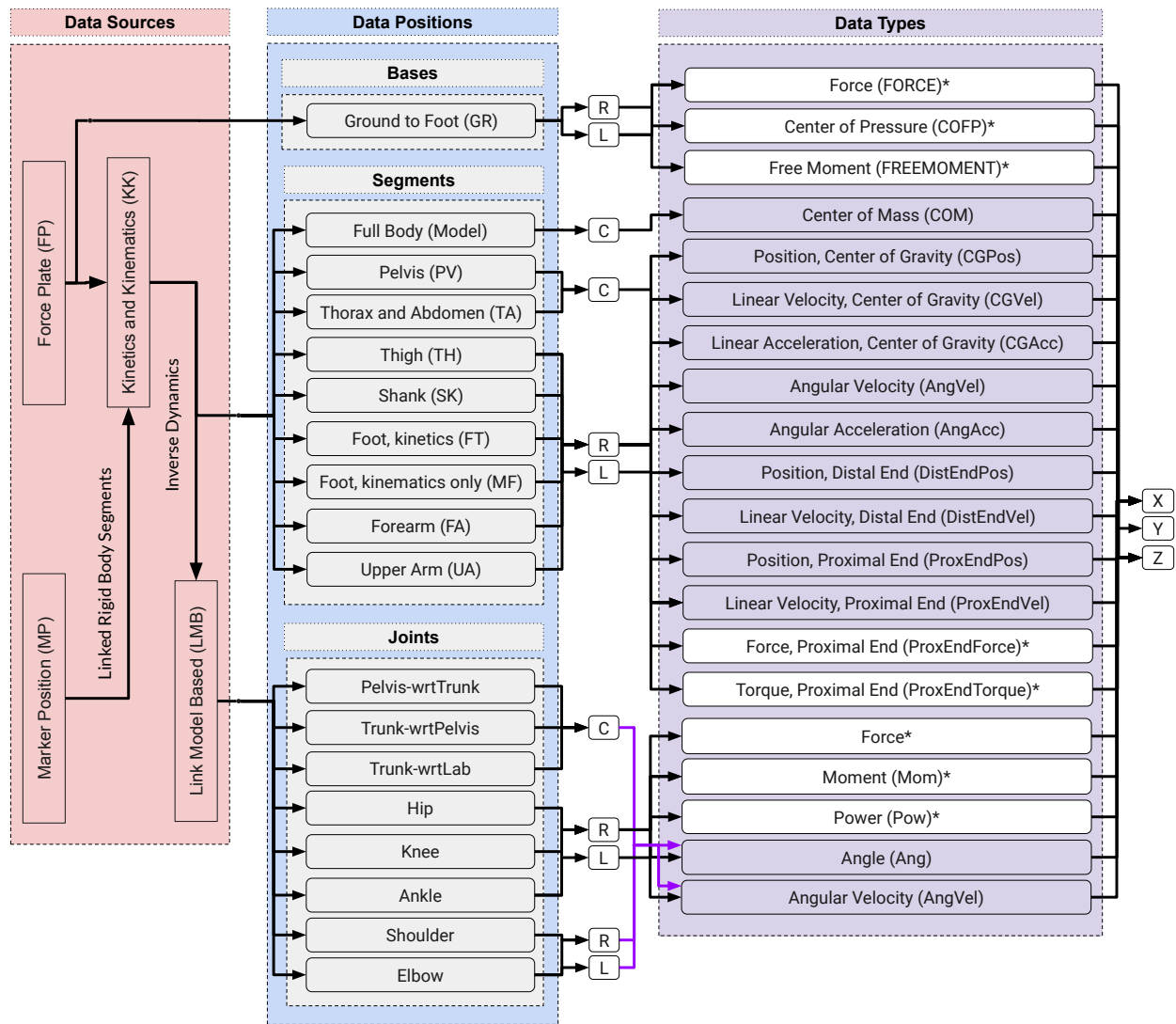
FIGURE 6.3. A flow chart demonstrating the path of extracted biomechanical raw signals based on source, positions (bases, joints, and segments), and types.* Refers to the signals extracted using ground reaction force (GF) data.

ensemble learning is used to aggregate the different biomechanical signals to reduce variance and produce a more generalizable model with better accuracy.

6.3.1. System Overview. After the data preprocessing step, the biomechanical signals of a subject are fed into the deep ensemble learning model. The output of the model is an identification or label of the subject. In Q1, the label is either post-stroke or age-matched control, and in Q2, more or less affected sides. In this section, we first explain how deep ensemble learning is used

to identify the label of the subject based on a single biomechanical signal and then integration of biomechanical signals.

- **Single Biomechanical Signal**: A single biomechanical signal is referred to by three parts: data source, position, and type. (i.e., KK/UA/AngAcc). It can be used as an independent input to identify the subject's label. Specifically, the raw signal is fed into the deep ensemble learning model, which consists of a set of base learners. A base learner is a neural network model trained using a random initialization method. Majority voting is then applied to the output of this set of base learners to decide the label of the subject.

- **Integration of Biomechanical Signals**: Building on the identification result of a single signal, we can further aggregate different signals based on type, position, and source to generate more robust identification results for a particular subject. Again, we use majority voting to aggregate the results. In this work, we consider three ways of integration: based on the different data sources (KK, LMB, FP, KK+LMB, KK+LMB+FP), position (bases, joints, and segments, where each position includes subcategories), and type (in which each signal type integrated from different data sources and positions).

6.3.2. A Single Biomechanical Signal. We first focus on the identification of a single biomechanical signal. One such signal consists of ~30 seconds of time series data in three dimensions. It is first segmented into multiple (five in our case) time windows. These signals are used to train multiple (seven in our case) base learners. Then the results of base learners are aggregated to generate the predicted label. We explain the steps in detail next.

6.3.2.1. *Window slicing.* First, a single biomechanical signal is segmented into five time windows using the window slicing method [**64**,**68**,**97**], each consisting of five seconds and containing multiple complete gait cycles. We divide each subject's data into time windows for the following reasons. First and foremost, this step serves the goal of data augmentation [**64**]. Deep learning is a method that requires a vast amount of training data. In this work, even though there are a massive amount of data regarding the number of biomechanical signals, the number of subjects is only 110. Segmenting data would augment the data input of each subject (each subject's data is multiplied by the number of time windows), which allows the model to train on the same subject in multiple iterations (improving the accuracy). Furthermore, the multiple time windows of a single subject

allow us to use majority voting in identification and thus improve the algorithm's accuracy and robustness. This approach is also more robust to overfitting by increasing the number of training samples and reducing memorization.

We have experimented with different time window lengths. We note that a window length needs to be long enough to contain several complete gait cycles. The longer the time window, the more information each window contains, but the less the number of windows per subject. Numerical results show that a 5 seconds window demonstrates the best performance and is thus chosen here. Given each signal is a minimum of 30 seconds long, the first 30 seconds are segmented into five time windows of 5 seconds in length, and the remaining signal is discarded.

6.3.2.2. *A Base Learner.* Figure 6.4 shows the architecture of a base learner. In the figure, s_i represents subject i, and b represents a biomechanical signal of the subject. It is divided into $N = 5$ time windows $(t_1, \cdot, t_n)$. Each time window is fed into a base CNN model, as shown in Figure 6.5. The CNN has five hidden layers (Convolution, Max pooling, Dropout, Short Term Long Memory, and Dense) and uses (Rectified Linear Unit and Softmax) activation functions. We chose the CNN
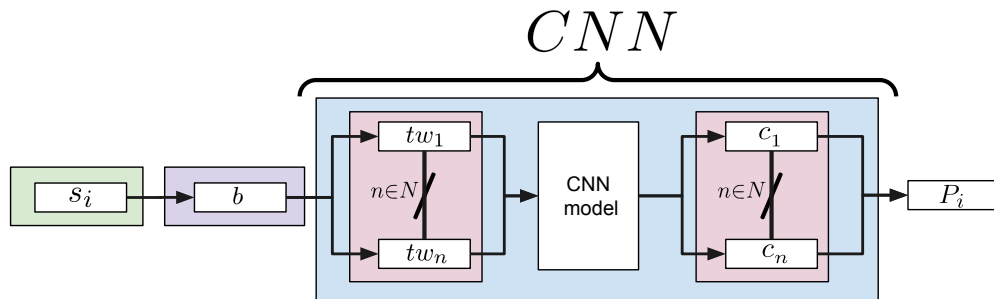

FIGURE 6.4. CNN is as a single base learner where P_i is the predicted output using a majority voting of the time windows.

approach because the biomechanical signals are time series data, and recent studies show that CNN works effectively with time series data [114] [70]. This work proposes a novel input format that handles biomechanical signals efficiently. As shown in Figure 6.5, the input given to the model is a single time window tw while c is an output prediction of that single time window tw. The input is given to the model in a 2D design where the three vectors (x, y, and z-axis) are three rows in the input matrix, as shown in Figure 6.5. In this format, the convolution layers deal with spatial information in two aspects; (1) vertically, across different axes, to capture the relationship between
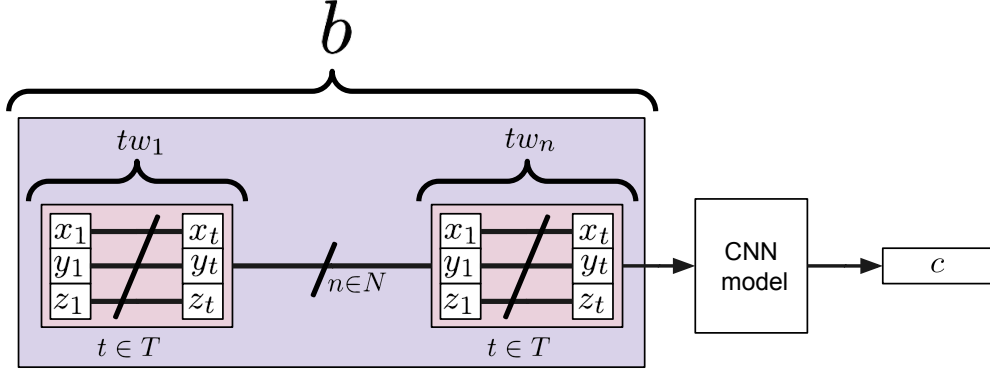
FIGURE 6.5. Training a CNN model based on a single time window tw input, which consists of (x,y, and z axes).

the three direction movements in a single biomechanical signal, and (2) horizontally across a single axis, to capture the movements in a single direction of the biomechanical signal across the time.

The output of the model (for subject i) is c_i, which is the prediction of the subject's label. For example, in the post-stroke vs. age-matched control identification, $c_i = 0$ means the CNN predicts that the subject i is in the control group and $c_i = 1$ otherwise. Because each biomechanical signal has $N = 5$ time windows, the predicted output P_i is the majority voting of the time windows; i.e.,

$$(6.1) \qquad\qquad P_i = \underset{n \in N}{mode} \left(c_i\right)_n,$$

where the $mode(.)$ is the majority voting, taken across N time windows of a single biomechanical signal of subject i.

6.3.2.3. *Ensemble Model.* Figure 6.6 illustrates the architecture of the ensemble model. Specifically, there are seven base learner models ($k \in Z$, and $Z = 7$ base learners in this work), each as described above but with a different random initialization during training. Using the same input (i.e., a biomechanical signal of subject i), majority voting is used to generate the output prediction of a single subject i over all base learners of a single biomechanical signal:

$$(6.2) \qquad\qquad L_i = \underset{k \in Z}{mode} \left(P_{ik}\right)$$

6.3.3. Integration of Biomechanical Signals. Different biomechanical data can be effectively integrated to fully exploit the complementarity between various gait data signals. After we
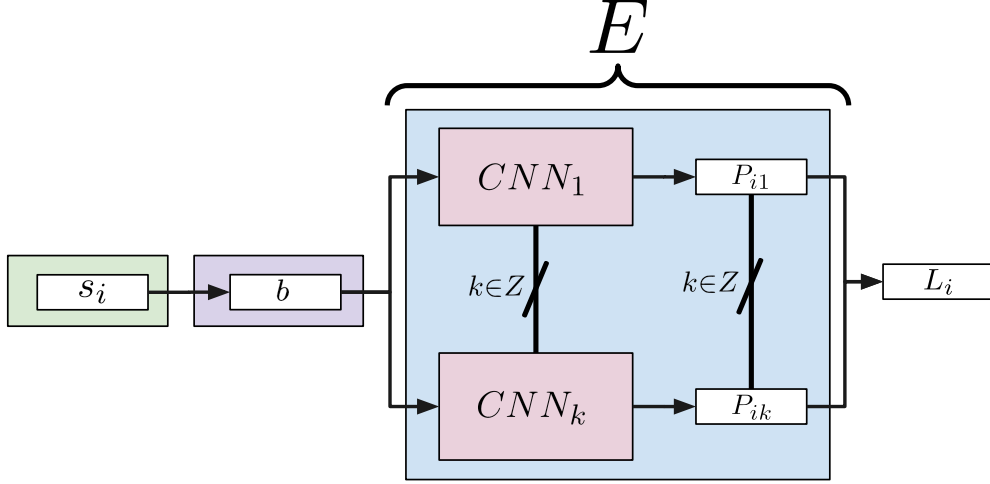
FIGURE 6.6. E is a fusion architecture that consists of several base learners $CNN_{k \in Z}$ targeting a single biomechanical b of a subject i.

identify the respective label of subjects based on a single biomechanical signal, we aggregate the label of each subject using groups of biomechanical data, i.e., data source, types, and position (bases, segments, and joints).

Figure 6.7 illustrates the architecture of the ensemble architecture E^B for a group of independently trained ensemble models $E_{j \in B}$, also called sub-fusion models. Specifically, there are B such sub-fusion models as described in the previous subsection where each sub-fusion is trained on a separate biomechanical signal b_j. Note here we abuse the notation B to represent both the signal set and the number of signals in that set.

For each signal group, majority voting is employed to aggregate the prediction output from each biomechanical signal in the group as follows:

$$(6.3) \qquad L_i^B = \underset{j \in B}{mode} \left(P_{ikj} \right),$$

where B refers to the set of biomechanical signals in each of the four categories, P_{ikj} is a base learner prediction of a single biomechanical signal j of the subject i using base learner k, and L_i^B in Eq. 6.3 as the prediction output of subject i using biomechanical signal group B.

6.3.3.1. *Integration Categories.* Since a single biomechanical signal b may not perform optimally independently, we examine the following integration of signals $b_{j \in B}$ in order to select the best fusions:
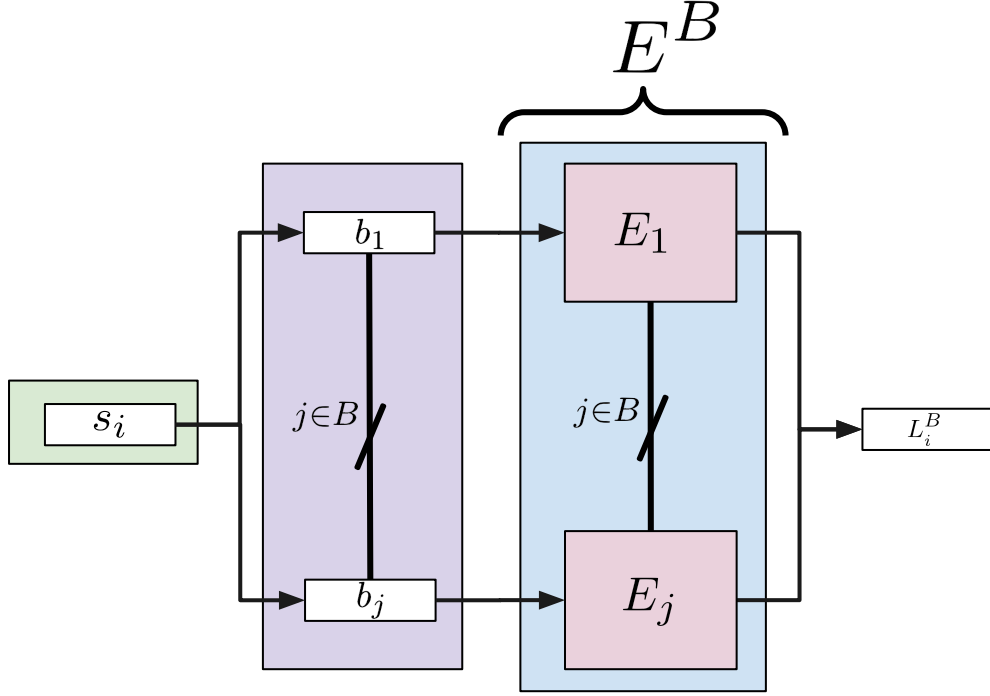
FIGURE 6.7. Architecture of a single subject over integration of biomechanical signals $b_{j\in B}$. E^B is a fusion architecture that consists of a set of biomechanical signals where each signal b_j is fed to a designated architecture E_j.

- Based on the **data positions** as follows: bases (i.e., GR), segments (i.e., PRV, RTA, TH, etc.), joints (i.e., Ankle, Hip, Knee, etc.)

- Based on the **data types** (e.g Pow, Ang, and Mom).

- Based on the **data sources** such as follows: KK, LMB, FP, LMB+KK, and LMB+KK+FP.

6.3.4. Performance Metric. In this work, we use **Identification Accuracy (IACC)** as the performance metric to evaluate the identification performance of the deep ensemble learning given a biomechanical signal or an aggregation of signals. IACC is mathematically expressed as:

(6.4)
$$IACC = \frac{1}{S} \sum_{i}^{S} I_s(L_i^B = g_i),$$

90

where S is the total number of subjects, g_i represents the ground-truth label, and $I_s(\cdot)$ is an indicator function that can be expressed as:

$$(6.5) \qquad I_s(\cdot) = \begin{cases} 1, & \text{if } (L_i^B = g_i) \text{ true} \\ 0, & \text{if } (L_i^B = g_i) \text{ false} \end{cases}$$

The pseudocode for calculating IACC for a single signal ($B = 1$) and for integration of signals ($B > 1$) are shown in Algorithm 1. Furthermore, we use leave-one-subject-out for training and testing. This approach allows us to use as much training data as possible while ensuring signals of a single subject in the test were not seen at all during training and thus allows for better generalization.

Algorithm 1 IACC for a single signal and integrations of singlas

---
Initialize $L_i^B \leftarrow []$
for $i \in \mathcal{S}$ **do**
   $P_i \leftarrow []$
   for $j \in \mathcal{B}$ **do**
     for $k \in \mathcal{Z}$ **do**
       $P_{ikj} \leftarrow \underset{n \in N}{mode}[(c_i)_n]_{jk}$
     end for
   end for
   $L_i^B \leftarrow mode(P_{ikj})$
end for
$IACC \leftarrow avg(L_i^B)$

6.4. Results

In this section, we present the identification accuracy results to answer the four questions Q1-Q4 described in Sec. 6. We divide the results into four tables; Table A1 on a single biomechanical signal; Table 6.2, Table 6.3, and Table 6.4 on the three integrations. We report the identification accuracy (IACC) and the standard deviation (SD) among random initialization models per a single signal (b) or integration of signals ($b_{j \in B}$) to measure the confidence among base learners. The accuracy of the proposed architecture is calculated as in equation (6.4).

- Table A1: shows the results based on a **single biomechanical signal**: such as (FORCE, MOMENT, COFP, etc.) where each signal is referred to by 3 columns (data source, position, and type).

- Table 6.2: shows the results based on the **data positions**: bases (GR), segments (PRV, RTA, TH, etc.), and joints such (Ankle, Hip, Knee, etc.) where each integration of signals are referred to by two columns (data source, position).

- Table 6.3: shows the results based on the **data types**: such as (Pow, Ang, Mom, etc.) where each integration of signals is referred to by two columns (data source, type).

- Table 6.4: shows the results based on the **data types**: such as (KK, LMB, and FP, and KK+LMB+FP, etc.).

In answering each question, we first report the identification accuracy (IACC) using different single biomechanical signals based on sources, positions, and types. By comparing these results, we first identify which data type/position/source is the most revealing. We next explore the integration of signals based on position, type, and source. Not all data fusions demonstrate the same identification performance for each combination form. Still, in most cases, the accuracy of combined fusion decisions is higher than that of a single biomechanical signal.

Using the proposed framework, we achieved up to 100% accuracy in Q1 and Q2 while identifying the most revealing biomechanical signals (Q3) and quantifying the GF importance (Q4) in the process. For example, to identify **more vs. less affected sides**, in terms of a **single signal**, we achieve an accuracy of 99.21% using Ankle from LMB. In terms of **positions**, we achieve an accuracy of 100% using UA from KK and 99.21% (w.o. GF). In terms of **types**, we achieve an accuracy of 100% using Ang from LMB (w.o. GF). In terms of data sources, we achieve an accuracy of 100% using KK+LMB+FP and 99.21% (w.o. GF). To identify **post-stroke vs. age-matched control**, in terms of a **single signal**, we achieve an accuracy of 97.73% using Knee/Ang from LMB. In terms of **positions**, we achieve an accuracy of 100% using PV from KK and 98.18% (w.o. GF). In terms of **types**, we achieve an accuracy of 97.73% using CGPos from KK (w.o. GF). In terms of **data sources**, we achieve an accuracy of 100% using KK+LMB+FP and 100% (w.o. GF).

6.4.1. Post-stroke vs. Age-matched control. In terms of **single signal** (Table A1), using ProxEndTorque (on KK/TH) yields an accuracy of 97.26%, using Mom (on LMB/Hip) yields an

TABLE 6.2. **Results based on the data positions.**

| Data Source | Positions | More vs. Less affected sides | | Post-stroke vs. Age-matched control | | | |
| | | Right/Left[a] | | Right/Left[a] | | Center[b] | |
		With Force	Without Force	With Force	Without Force	With Force	Without Force
KK	UA	**100.0%** (0.57)	**99.21%** (0.51)	94.09% (0.22)	93.64% (0.26)	-	-
	FA	96.03% (0.62)	96.83% (0.66)	91.82% (0.27)	89.55% (0.33)	-	-
	FT	96.83% (0.41)	96.03% (0.49)	**95.91%** (0.16)	**95.45%** (0.23)	-	-
	MF	93.65% (0.4)	96.03% (0.49)	**95.91%** (0.09)	**95.45%** (0.12)	-	-
	Model	-	-	-	-	88.18% (1.43)	88.18% (1.43)
	PV	-	-	-	-	**100.0%** (0.45)	**98.18%** (0.37)
	SK	98.41% (0.44)	98.41% (0.45)	94.06% (0.24)	93.15% (0.34)	-	-
	TA	-	-	-	-	99.09% (0.39)	97.27% (0.29)
	TH	99.21% (0.47)	**99.21%** (0.48)	95.43% (0.33)	94.52% (0.39)	-	-
LMB	Ankle	**99.21%** (0.62)	**97.62%** (1.17)	96.36% (0.54)	**96.36%** (0.9)	-	-
	Elbow	94.44% (0.91)	94.44% (0.91)	93.64% (0.58)	93.64% (0.58)	-	-
	Hip	**99.21%** (0.46)	**97.62%** (0.6)	96.36% (0.43)	91.82% (0.81)	-	-
	Knee	98.41% (0.75)	88.89% (1.35)	**96.82%** (0.48)	94.55% (0.45)	-	-
	Pelvis-wrtLab	-	-	-	-	**91.82%** (0.48)	**91.82%** (0.48)
	Shoulder	93.65% (0.39)	93.65% (0.39)	85.78% (0.85)	85.78% (0.85)	-	-
	Trunk-wrtLab	-	-	-	-	**91.82%** (1.26)	**91.82%** (1.26)
	Trunk-wrtPelvis	-	-	-	-	89.09% (0.81)	89.09% (0.81)
FP	GR	97.62% (0.72)	-	95.0% (0.64)	-	-	-

accuracy of 94.55%, and using COFP and FORCE (on FP/GR) yields an accuracy of 90.91%. In terms of **data positions** (Table 6.2), using KK (on FT and MF) yields an accuracy of 95.91%. LMB (of Knee) yields the highest accuracy, of 96.82%, and FP (on GR) yields an accuracy of 95%. In terms of **data types** (Table 6.3), using KK (on ProxEndTorque) yields an accuracy of 96.82%, using LMB (on Mom) yields an accuracy of 95%, and using FP (on COFP and FORCE) yields an accuracy of 90.91%. In terms of **data sources** (Table 6.4), using a collection of KK+LMB+FP data yields an accuracy of 96.36%. The deep ensemble model achieves an accuracy of 95.91% when

TABLE 6.3. **Results based on the data types.**

Data Source	Types	More vs. Less affected sides		Post-stroke vs. Age-matched control			
		Right/Left[a]		Right/Left[a]		Center[b]	
		With Force	Without Force	With Force	Without Force	With Force	Without Force
KK	AngAcc	-	96.83% (0.67)	-	93.18% (0.48)	-	85.45% (1.07)
	AngVel	-	96.03% (0.64)	-	93.18% (0.39)	-	90.91% (1.18)
	CGAcc	-	97.62% (0.99)	-	92.73% (0.25)	-	92.73% (0.58)
	CGPos	-	96.83% (0.72)	-	**97.73%** (0.32)	-	90.0% (0.51)
	CGVel	-	97.62% (0.37)	-	94.09% (0.23)	-	**93.64%** (0.84)
	COM	-	-	-	-	-	88.18% (1.43)
	DistEndPos	-	96.03% (0.53)	-	96.82% (0.33)	-	90.0% (0.8)
	DistEndVel	-	96.83% (0.36)	-	92.73% (0.35)	-	91.82% (0.59)
	ProxEndForce*	**98.41%** (0.37)	-	93.18% (0.58)	-	**93.64%** (1.3)	-
	ProxEndPos	-	**98.41%** (0.55)	-	96.36% (0.42)	-	88.18% (1.38)
	ProxEndTorque	96.03% (0.44)	-	**96.82%** (0.25)	-	90.0% (1.3)	-
	ProxEndVel	-	**98.41%** (0.51)	-	93.18% (0.26)	-	90.91% (0.89)
LMB	Ang	-	**100.0%** (0.77)	-	**99.55%** (0.45)	-	**95.45%** (0.44)
	AngVel	-	96.03% (0.48)	-	93.64% (0.37)	-	94.55% (0.86)
	Force*	**98.41%** (0.53)	-	93.18% (0.58)	-	-	-
	Mom*	97.62% (0.78)	-	**95.0%** (0.63)	-	-	-
	Pow*	94.44% (0.31)	-	93.18% (0.48)	-	-	-
FP	COFP*	92.86% (0.92)	-	**90.91%** (0.63)	-	-	-
	FORCE*	**96.03%** (1.3)	-	**90.91%** (1.09)	-	-	-
	FREEMOMENT*	91.27% (1.04)	-	90.0% (1.13)	-	-	-

applied to KK and 95% when applied to FP, respectively, Using LMB yields the highest accuracy of 96.82%. Furthermore, calculating KK+LMB+FP results with 100% accuracy (w and w.o.GF).

6.4.2. More vs. Less Affected Sides. In terms of **single signal** (Table A1), using Prox-EndForce (on KK/FT and KK/SK) yields an accuracy of 98.41%, using Force (on LMB/Ankle) yields an accuracy of 99.21%, and using FORCE (on FP/GR) yields an accuracy of 96.03%. In

Data Source	Move vs. Less affected sides		Post-stroke vs. Age-matched control			
	Right/Left[a]		Right/Left[a]		Center[b]	
	With Force	Without Force	With Force	Without Force	With Force	Without Force
KK	99.21% (0.56)	**99.21%** (0.36)	95.91% (0.32)	95.91% (0.38)	99.09% (0.58)	98.18% (0.82)
LMB	99.21% (0.0)	97.62% (0.7)	**96.82%** (0.22)	**96.82%** (0.53)	99.09% (1.51)	99.09% (1.51)
FP	97.62% (1.4)	-	95.0% (0.91)	-	-	-
KK+LMB +FP	**100.0%** (0.36)	**99.21%** (0.36)	96.36% (0.41)	95.91% (0.38)	**100.0%** (0.45)	**100.0%** (0.64)

Note: the symbol "-" means that there is no data corresponding to the data source, position, or type.

[a] Based on the central mass of the body.

[b] Based on the right and left sides of the body.

* Force signal.

terms of **data positions** (Table 6.2), using KK (on UA) yields an accuracy of 100%, and LMB (either Ankle or Hip) yields the highest accuracy of 99.21%, using FP (on GR) yields an accuracy of 97.62%. In terms of **data types** (Table 6.3), while using KK (on ProxEndForce) and LMB (on Force) yields an accuracy of 98.41%, FP (on FORCE) yields an accuracy of 96.03%. In terms of **data sources** (Table 6.4), the deep ensemble model achieves an accuracy of 99.21% when applied to extracted gait data KK or LMB while obtaining an accuracy of 97.62% when applied to FP data. Furthermore, using a collection of KK+LMB+FP data yields the highest accuracy of 100%, resulting in the best fusion scheme for this task.

6.4.3. With and Without Ground Reaction Force Data. In all tables, we compare "With Force" and "Without Force" for each task. They indicate whether signals from the GF data are used or not. In Table A1 and 6.3, a single signal or signal type is either generated from the GF data or not. Therefore, the results are either with GF data or without. In these two tables, we observe that the signals generated from GF data are, in general, as good as, if not better, than signals from GF data.

In Table 6.2, we aggregate multiple biomechanical data from a single **positions** (say UA). In this case, "No Force" uses a subset of the signals in "With Force." We, in general, observe negligible

performance differences for the best results (bolded ones). In terms of the **post-stroke vs. age-matched control**, the biggest difference comes from PV (on KK), where the accuracy dropped from 100% to 98.18%. Furthermore, the highest accuracy of 100% achieved by UA (on KK) slightly dropped to 99.21% when removing GF data. Similarly, the highest accuracy of 96.82% achieved by Knee (on LMB) slightly dropped to 94.55% when removing GF data, indicating that the usage of LMB can compensate for GF data but with a trade-off. In terms of the **central mass of the body**, we can observe the PV (on KK) outperforms the highest accuracy of Pelvis-wrtLab and Trunk-wrtLab (on LMB), while in the case of **post-stroke vs. age-matched control** (right/left), the Anke (on LMB) exceeds the FT and MF (on KK) by 0.91% (Tables-6.2). This indicates that no single biomechanical signal can work very well in and of itself due to the characteristics of each task (as also shown in Table A1). These results indicate that the usage of KK/LMB data can largely compensate for GF data for the respective tasks.

In Table 6.4, we, in general, observe negligible performance differences. Regarding **post-stroke vs. age-matched control**, removing GF data, the highest accuracy of 96.82% in the LMB does not change. This indicates that the integration of KK+LMB can largely compensate for GF data, at least for this task. In terms of the **more vs. less affected sides**, removing GF data, the highest accuracy of 100% in the integration of KK+LMB+FP dropped slightly to 99.21%.

6.5. Discussions

This work serves as a first step toward gait analysis automation using deep ensemble learning. We note that the tasks studied in this work are relatively simple, i.e., identification of post-stroke vs. age-matched control (Q1) and more vs. less affected sides (Q2). Our current results are promising, over 99% identification accuracy was achieved with and without GF data. It inspires us to study more sophisticated tasks in the future, such as the quantification of gait anomaly degrees. Our study on the GF data has the same implication: we have only evaluated the impact of GF on the two simple tasks, Q1 and Q2. We hope to further study the impact of GF on more sophisticated tasks in the future.

Furthermore, the identification performance can be improved by selectively integrating the most relevant gait signals. We will study different feature selection algorithms to achieve this goal. It

can further evaluate our hypothesis that the complementarity between the most pertinent biomechanical data can distinguish between different gait patterns and compensate for the GF data. Furthermore, in the current ensemble learning, we use the majority voting scheme to aggregate information. Additional aggregation schemes can be evaluated in the future so that different signals may be weighted differently based on their importance. The results of both approaches can potentially provide useful and quantitative information to clinicians in terms of the importance of different signals on gait abnormality. In the future, we can also evaluate additional machine learning techniques, such as transfer learning (use models trained in one biomechanical signal as the pre-trained model for another signal), multitask learning (complete multi goals at the same time), and self-supervised learning (aggregate all signals to pre-train the network architecture by predicting, say signals in the next time step) in this context. The objective is to explore opportunities to improve the prediction performance and possibly using less signals.

6.6. Conclusion

Accurate recognition and assessment of gait abnormality play an important role in making effective rehabilitation treatment strategies for post-stroke patients. This research proposes a deep ensemble learning approach to extensively analyze the voluminous biomechanical signals generated in motion labs (205 signals, 615 time series, per subject). It answers four questions, including the identification of post-stroke vs. age-matched control, more vs. less affected sides, identification and quantification of the most revealing signals, and quantifying the impact of GF data. The work demonstrates strong identification results, and the identification performance improves as more biomechanical signals are used. Using deep ensemble learning, the best biomechanical sets produce over 99% identification accuracy in both tasks with and without force place data. This is a promising step towards gait analysis automation better understand the impact of stroke on mobility.

Walk4Me: Telehealth Community Mobility Assessment, An Automated System for Early Diagnosis and Disease Progression

7.1. Introduction

The increasingly rapid growth of mobile and carry-on devices and the consequent enormous community data production opened the door to the opportunity of exploring out-of-the-lab data and obtaining a more realistic understanding of how the treatments and rehabilitations affect the patient's daily life out-of-the controlled environmental conditions. This promoted the development of advanced tools to facilitate manipulating and constructing the produced data.

However, this opportunity has not been exploited effectively, even though several existing commercial systems target specific aspects such as disease detection, athlete assessments, etc. These systems are costly and require to be installed in hospitals or labs. There is a lack of a system that is publicly available, free, easy to use, and scalable to upgrade by researchers, developers, and clinicians. However, building such a system (that can handle big data such as steam of data, e.g., time-series data collected from sensors) which contains visualizers, recommenders, and identifiers for clinical purposes, is not a trivial task: it involves a large number of subtle decisions/features that require experience and close collaboration from the expertise, e.g., clinicians, doctors, developer, and data scientists.

In this paper, we present Walk4me, a telehealth community mobility assessment, an automated system for early diagnosis, severity, and progression identification. This system is a tool that presents visualization to facilitate the exploration process. Walk4Me system enables both open-ended and focused exploration. It provides classical machine learning (CML) and deep learning (DL) algorithms [**70**, **91**] to help researchers engage in disease identification based on time-series data. It aims to allow researchers to answer more specific questions using ML compared to a traditional specification tool. The system automatically extracts CFs directly for the raw signals,

such as the number of steps, distance, and step length. It also provides several visualizations, such as the 3d shape of the signal and the correlation between different clinical features. It also provides a feature for lining up and synchronizing the raw signal with video.

We conducted two studies using our tool for DMD patients [**94**, **95**] and post-stroke survivors to analyze some aspects of the disease characteristics and related data exploration, extract clinical features, utilize AI for early diagnosis purposes, and identify the most related biomechanical signals. We collected participants' data (from patients and healthy subjects) to test and develop our system and enhance our approaches. As a result, we propose a method to help facilitate the task of clinicians, researchers, and developers and help early diagnose, measure the treatment effectiveness, and better estimate the disease progression to improve the life of patients in this domain.

7.2. System Design

7.2.1. System Architecture. The Walk4Me system is composed of two main components: the server-side and the client-side, as depicted in Figure 7.1. The server-side technologies are used for data management, processing, and analysis, while the client-side technologies are used for data collection and transmission.

The system utilizes the following technologies:

- Server-side
    - Webserver: Apache
    - Programming languages: PHP, JavaScript, and Python
    - Data storage formats: CSV and JSON
    - Computational resources: Amazon Web Services (AWS) EC2 instances
- Client-side
    - Devices: iPhone with an accelerometer, gyroscope, pedometer, and GPS sensors
    - Programming language: Swift
    - Data storage formats: CSV and JSON

The server-side technologies enable data processing and analysis, such as feature extraction, machine learning model training, and statistical analysis. The client-side technologies enable real-time data collection and transmission, allowing the system to collect and store data from multiple participants.
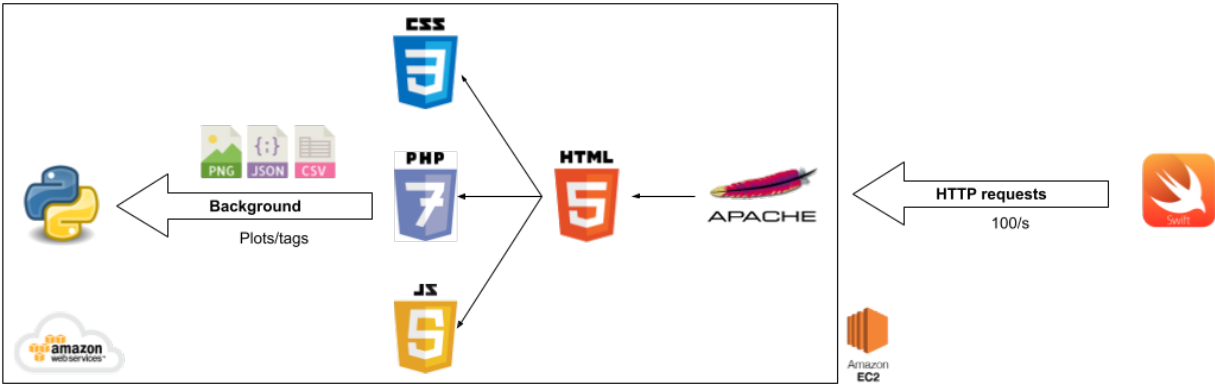

FIGURE 7.1. Project-specific login form for the Walk4Me system, enabling separate accounts for each project

7.2.2. Client side. We developed a smartphone app to be used by the participants to collect and stream sensory data from a smartphone that is attached to the participant's body (i.e., in the DMD study, the smartphone was placed on the central mass of the body). The smartphone app has 3 phases: The off, the Ready, and the Streaming phases. In the Off phase, the app is in an off mode where no data comes in or out. In the Ready phase, the app is on standby, waiting for a signal from the server to start streaming. During the Streaming phase, the app actively streams the sensory data online to the server.

7.2.3. Server side. We developed a web portal application to facilitate access, exploration, data entry, and data upload for researchers working with the Walk4Me system. The application's server offers several essential functions, including authentication to ensure only authorized users can access the portal, initiating data collection through activating the streaming mode using a record button, browsing and visualizing collected data, synchronizing video and sensor data, marking activities performed by participants using time and demographic information, analyzing and processing data, and running different machine learning models for various tasks such as classification and regression using AI-based algorithms. Figure 7.3 shows the web portal interface.

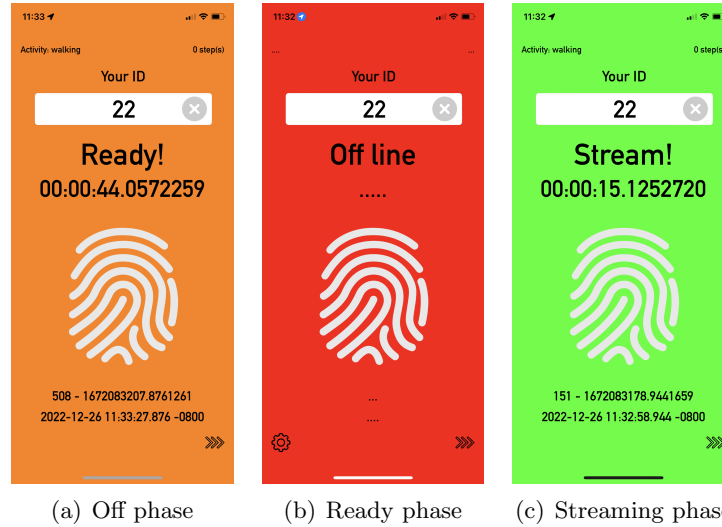(a) Off phase   (b) Ready phase   (c) Streaming phase

FIGURE 7.2. This figure shows the Graphical User Interface (GUI) for the Walk4Me smartphone app, which has three main phases: (a) the "Off" phase, where the app is ready to stream but is not yet active; (b) "Ready" phase, where the app is inactive and waiting for the user to initiate streaming; (c) "Streaming" phase, where the app sends sensor data via the internet to the server. The GUI provides a user-friendly interface for streaming data and allows the user to monitor the status of the data transmission. The figure can help users understand the functionality of the Walk4Me app and its features.

7.2.3.1. *Authentication login.* Figure 7.4 shows the login page of the web application, which allows researchers to access their project data securely. The system provides separate logins for each project, enabling multiple groups of researchers to work in their own private space without accessing data from other projects.

7.3. Feature Analysis and Implementation

The Walk4Me system provides a range of tools for exploring and analyzing sensory data. These tools include feature extraction, data visualization, and statistical analysis. By using these tools, researchers can gain valuable insights into the gait patterns of patients and develop more effective treatments for mobility impairments. Additionally, the system offers the ability to implement these features in machine learning models for automated classification and prediction.

7.3.1. Feature Extraction. The system provides a feature extraction tool that can extract a variety of gait features from the accelerometer raw signal. These features include both clinical
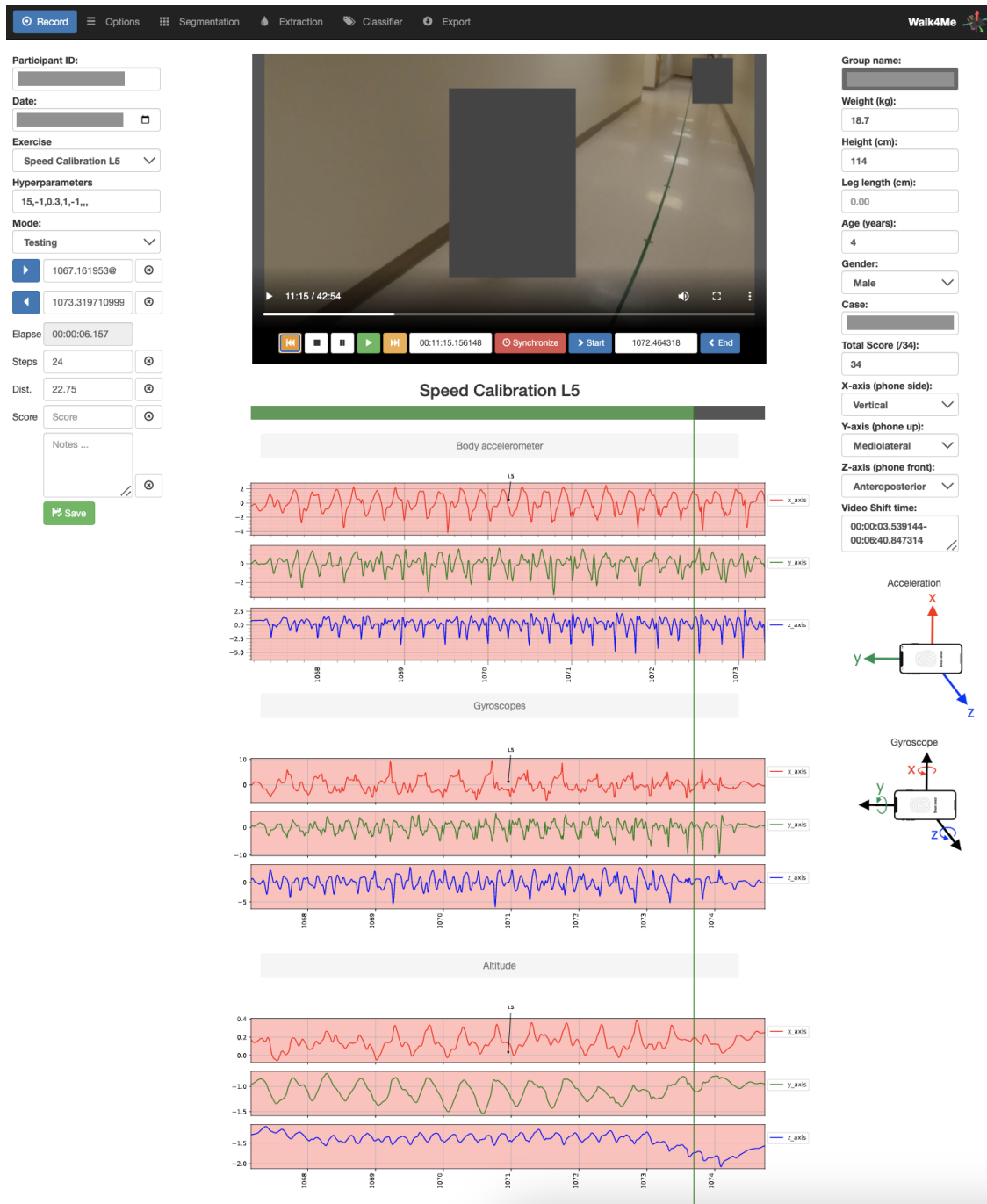
FIGURE 7.3. The web portal of the Walk4Me system. The window displays a patient's profile and the raw signal of one of the gait activities

FIGURE 7.4. This figure shows the login form of the Walk4Me system, which allows for the creation of separate accounts for each project.

and computational features, such as the number of steps, total distance, step lengths, average step length, step duration, total duration, step frequency, and average speed. These features are useful for data analysis and can be used to train classical machine learning classifiers.

FIGURE 7.5. Phone orientation during streaming data.

7.3.2. Orientation.

7.3.3. Dashboard. The clinical dashboard is another tool provided by our system that enables you to illustrate the measurements obtained using signal processing. For instance, you can create a visualization for the number of steps at different step velocities, a distribution histogram of the number of steps at different step velocities, step length per forward acceleration peaks, and the distance traveled by step length.

7.3.4. Signal overlap. A signal overlap is a tool provided to the researchers to compare two signals either in the same patient or across different participants. One of the benefits of this tool is comparing a gait cycle of a control participant vs. a patient with a mobility disorder. Figure 7.9 shows the tool's user interface. Nothing to mention that this tool also allowed researchers to mark

103
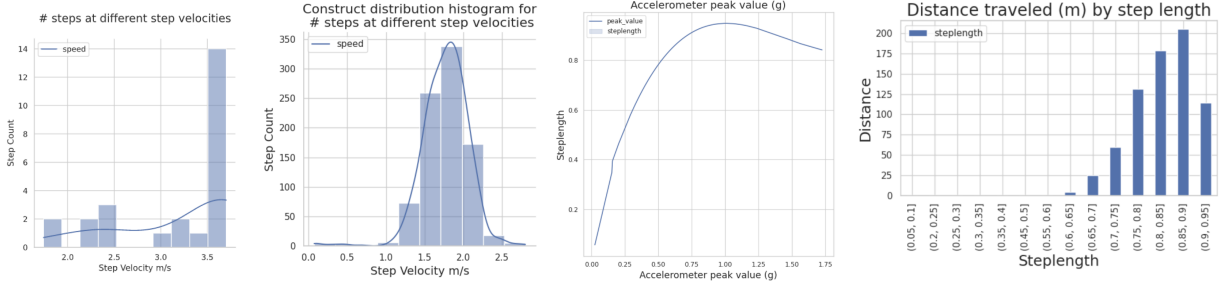
FIGURE 7.6. This figure shows visualizations provided by the system, including the number of steps at different step velocities, distribution histogram of the number of steps at different step velocities, step length per forward acceleration peaks, and the distance traveled by step length. These visualizations are part of the data analysis process that the system provides to users.
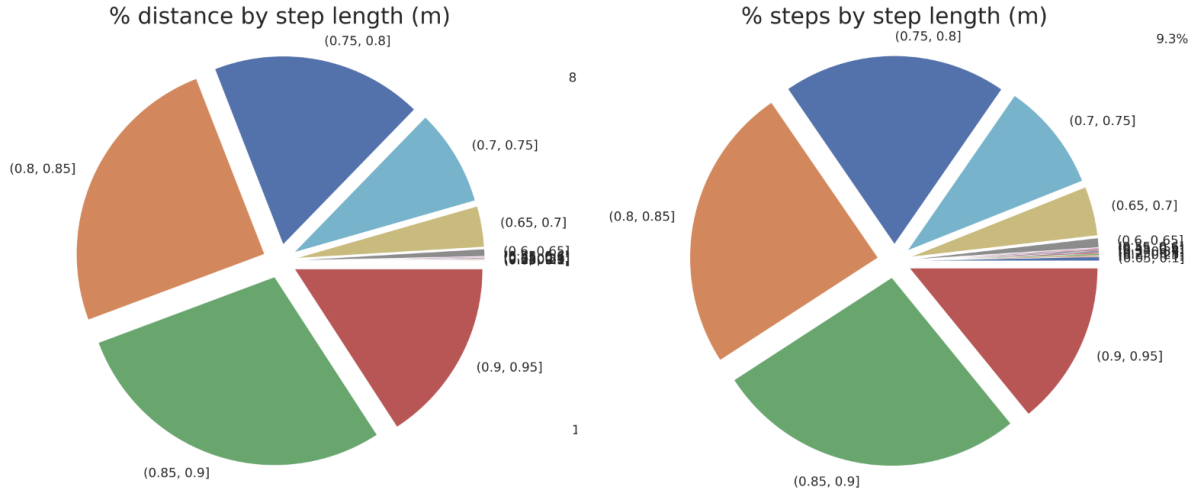
FIGURE 7.7. This figure shows the percentage of distance and the number of steps per step length.

a specific time event. For example, in walking analysis, researchers can mark different gait events such as Initial contact, Opposite Toe off, Heel rise, Opposite initial contact, Toe off, Feet adjacent, Tibia vertical, and Next initial contact.

7.4. Data Processing

The system provides three stages of data processing: feature extraction, raw data, and classification with dimensionality reduction.

104
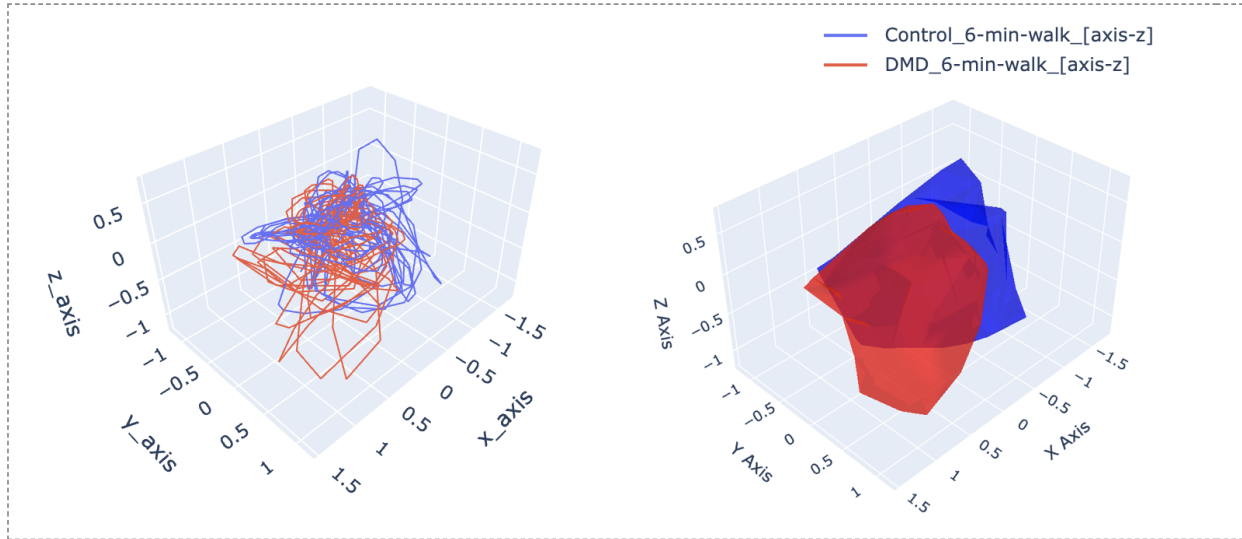
FIGURE 7.8. 3D visualization of forward acceleration for a participant with DMD compared to a TD peer.

7.5. Machine Learning Tool

The system includes a patient profile for each participant, which contains demographic and characteristic data. Additionally, researchers can label patients to indicate their class for use in the ML classifier. There is no limit on the number of classes that can be used. In binary classification, patients are labeled with two classes. For example, in the DMD study, patients were labeled as either DMD or control. In the stroke study, patients were labeled as either Stroke vs. control, or More vs. Less affected side.

Our system provides a machine learning (ML) tool to train the classification model. The ML tool includes two types of ML training methods: classical machine learning using the scikit-learn library, and DL using TensorFlow. With this tool, users can choose the best machine learning model and method for their specific dataset and classification task, enabling high accuracy and robustness in their machine learning models.

7.5.1. Classical Machine Learning. In classical machine learning, our system utilizes the generated classification features to train a variety of different classifiers. Specifically, we provide support for six different classifiers: AdaBoost, Random Forest, Bagging (ensemble), Gradient Boosting, Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbors, Gaussian Naive Bayes, and
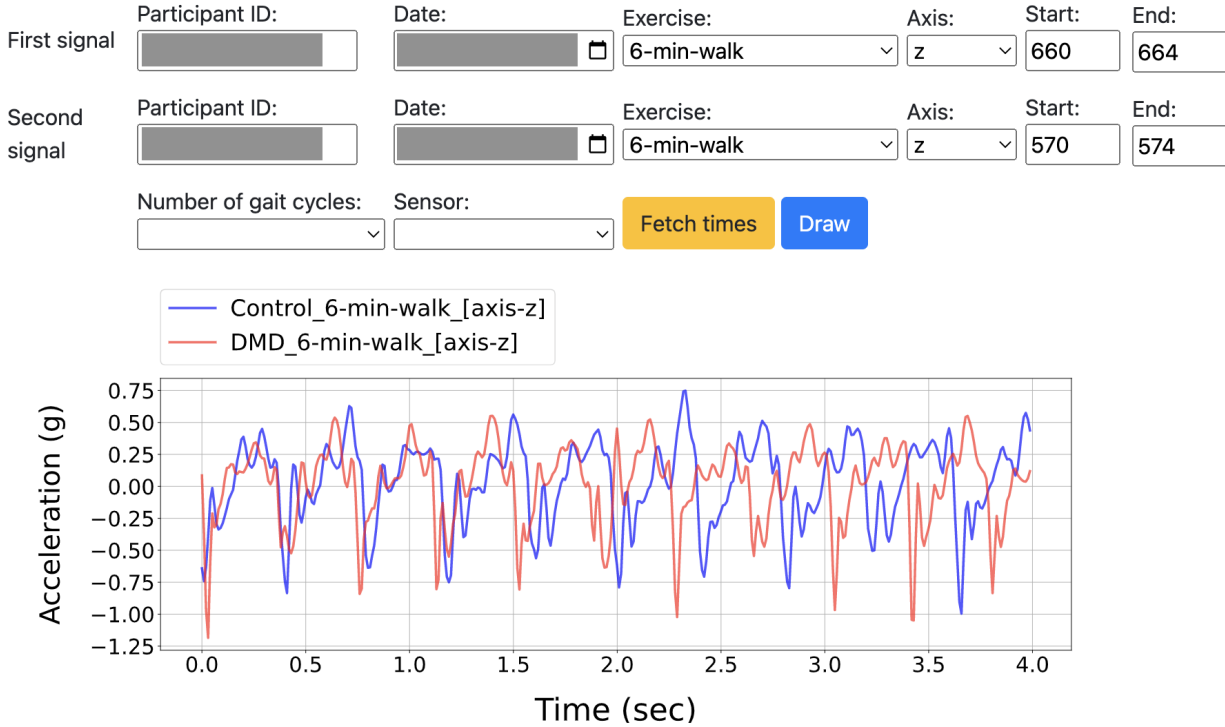
FIGURE 7.9. Comparison of two raw accelerometer signals recorded during the 6-minute walk test (6MWT) performed by a participant with Duchenne muscular dystrophy (DMD) and typically developing peer (TD). The graph shows the acceleration data from the forward movement overlapped to visualize the differences in gait patterns between the two groups.

Logistic Regression. This enables users to choose the best classifier for their particular dataset and classification task, and achieve high accuracy and robustness in their machine learning models.

7.5.1.1. *Feature correlation.* The system provides a visual representation of the CF as shown in Figure 7.10

7.5.1.2. *Dimensionality Reduction.* Our system implements dimensionality reduction as an optional step before the classifier to reduce the feature space and improve model performance. Two methods of dimensionality reduction are provided: Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA). Figure 7.11 shows an example of implementing LDA and PCA with an AdaBoost binary classifier on a participant. This step can greatly improve the efficiency and accuracy of the classification process, particularly when dealing with high-dimensional datasets.

Features Correlation
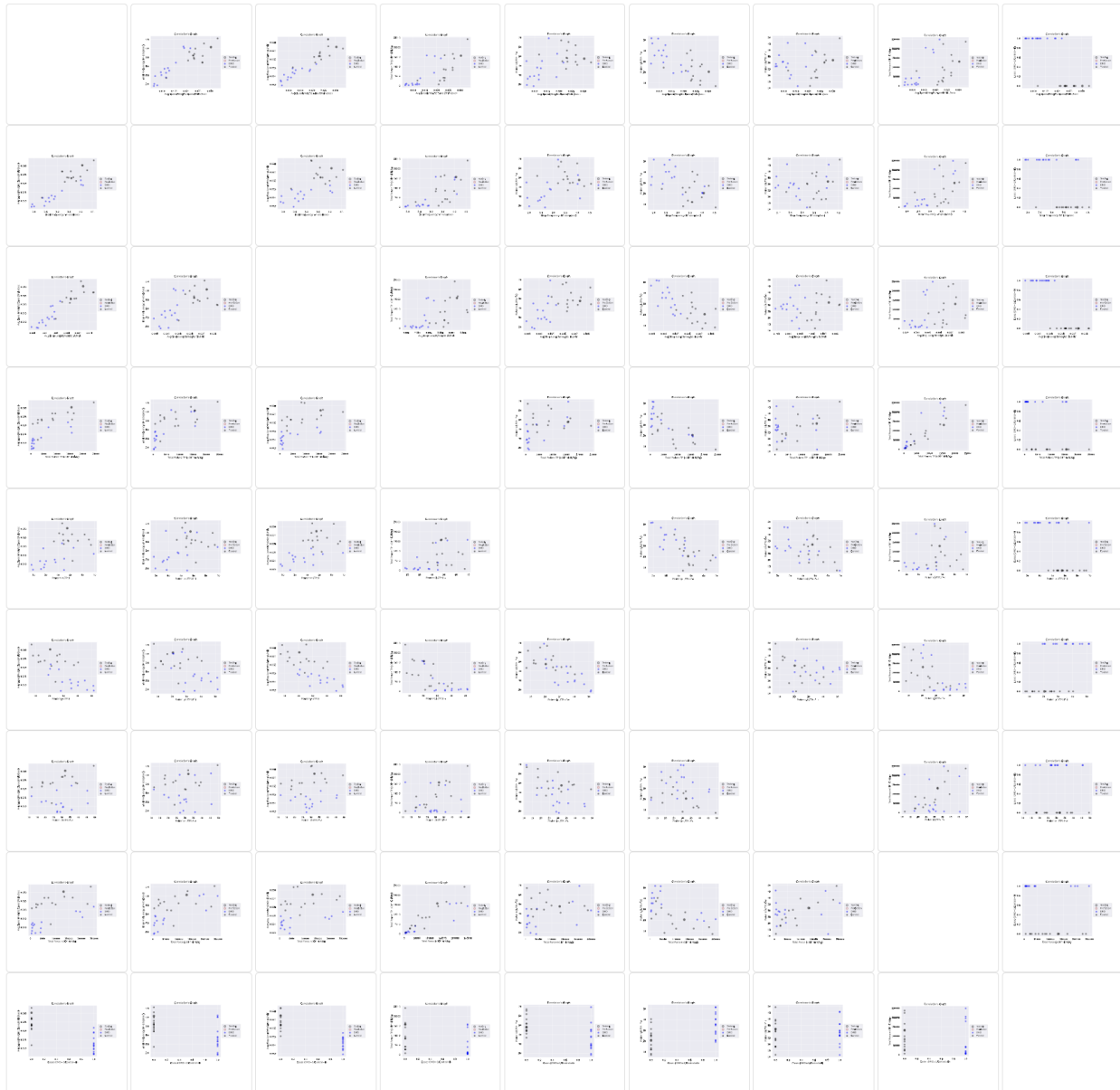
FIGURE 7.10. This figure shows the correlation between features across 8 of the clinical features (CF) in the DMD study. The figure provides insight into how these features may be related and can help guide data analysis and feature selection for machine learning models.
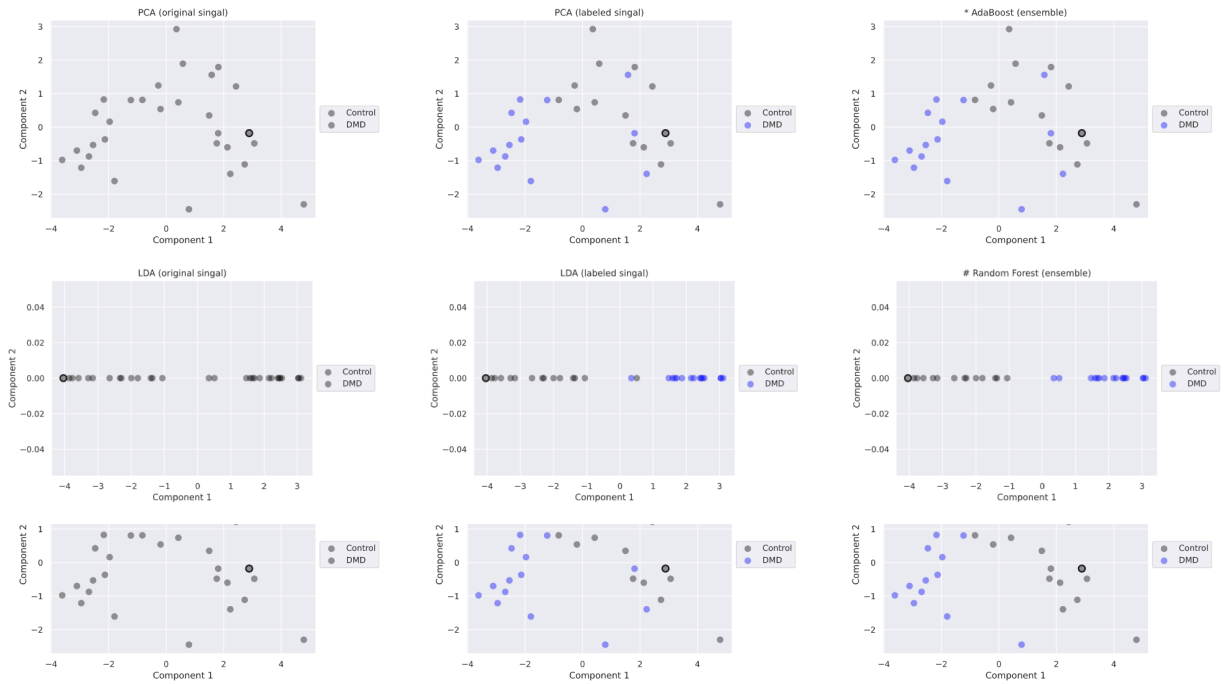
FIGURE 7.11. This figure illustrates the PCA and LDA representation of patient data. It includes the original data points, ground truth labels, and the corresponding classification results obtained by the system (multiple classifiers are used, and this is an example). The visualization provides valuable insights into the classification performance of the system.

7.5.2. Deep Learning for Participant Classification. Our system offers a powerful tool for training DL models to classify participants. The interface allows for flexible tuning of hyper-parameters, including the number of epochs and the use of leave-one-out and ensemble learning techniques. This enables users to optimize their models and achieve high accuracy in the participant classification task.

7.6. Conclusion

This work presents Walk4me, a telehealth system that uses Artificial Intelligence (AI) to detect gait characteristics in patients and typically developing peers. The system collects data remotely and in real-time from device sensors, extracts gait characteristics and raw data signal characteristics, and uses machine learning techniques to identify patterns that can facilitate early diagnosis, identify early indicators of clinical severity, and track disease progression across the ambulatory phase of the

disease. The authors have identified several machine learning techniques that differentiate between patients and typically-developing subjects with 100% accuracy across the age range studied and have identified corresponding temporal/spatial gait characteristics associated with each group. The system has the potential to inform early clinical diagnosis, treatment decision-making, and monitor disease progression.

CHAPTER 8

Conclusion

This work explores the intersection of wearable sensors and AI techniques in healthcare applications, with a focus on human activity recognition. The studies cover a wide range of topics, including gait analysis, non-verbal communication, and stroke-related gait analysis, and demonstrate the effectiveness and accuracy of wearable sensor-based systems in measuring gait parameters and identifying abnormalities that can assist in clinical diagnosis and rehabilitation treatment.

Advanced AI techniques, including deep learning and ensemble learning, play a vital role in improving the accuracy and reliability of these systems. The work in this collection offers significant implications for healthcare advancement and the application of AI in improving patient outcomes.

The first work, Overview of human activity recognition using wearable sensors, presents an overview of the essential components of human activity recognition (HAR) systems and their potential applications in healthcare. It is gaining popularity by applying wearable sensors to recognize and analyze human activities for the healthcare domain. For example, we leverage HAR systems to recognizing patients' early mobility activities in ICU and to analyzing the symptoms of DMD patients. This overview work covers the system design of HAR systems based on wearable sensors, focusing on healthcare applications. We emphasize the essential components of HAR systems, including sensor factors, data segmentation, feature extraction and selection, and AI model comparison. We also highlight the challenges and opportunities of HAR systems for healthcare.

The second work, Automated Detection of Gait Events and Travel Distance Using Waist-worn Accelerometers Across a Typical Range of Walking and Running Speeds, proposes a novel method for estimating gait characteristics accurately and reliably using a single accelerometer attached to the central body mass, which has significant implications for advancing the assessment of gait in both research and clinical settings. This study presents a novel method for estimating CFs during various gait speeds of walking, running, and jogging. By leveraging signal processing and ML techniques, this method estimates the number of steps, distance, and step length accurately.

Notably, the model was calibrated based on the individual's gait style, making the results more personalized and precise. Our findings demonstrate that using a single accelerometer attached to the central body mass is a more accurate and reliable method for CF estimation than a pedometer. Importantly, this method can be applied to both healthy individuals and those with muscle disorders without the need for GRF measurements. To our knowledge, this is the first study to propose a novel method that can extract reliable CFs from raw accelerometer data across different gait speeds and participants that can be used clinically. Overall, our results hold significant implications for advancing the assessment of gait in both research and clinical settings.

The third work, Gait Characterization in Duchenne Muscular Dystrophy (DMD) Using a Single-Sensor Accelerometer: Classical Machine Learning and Deep Learning Approaches, demonstrates the ability to measure gait parameters in Duchenne Muscular Dystrophy (DMD) using single-sensor accelerometers and employing both classical machine learning (CML) and deep learning (DL) models, with promising results for improving predictive accuracy and identifying clinically useful parameters indicating typical growth and development, gait impairment, and disease progression. Use of ubiquitous and widely available mobile devices with single accelerometers to remotely measure differences in common clinical gait parameters represents an opportunity to expand the study of temporospatial gait characteristics into the community setting. Our initial laboratory-based studies demonstrate ability to measure selected gait parameters across a range of typical ambulatory velocities in DMD and TD children to detect significant differences in temporospatial gait CFs that are consistent with previous studies, as well as to detect differences in proportional power of accelerations in vertical, mediolateral and anteroposterior axes. By using these clinical gait parameters and raw data and employing both CML and DL models, we are able to correctly predict whether sensor data is derived from children with DMD up to 100% of the time at self-selected-walk pace (SC-L3). By combining these approaches, we anticipate that through ongoing studies, we will be able to improve predictive accuracy and identify additional clinically-useful parameters indicating typical growth and development, gait impairment and disease progression across a wide range of individuals with neuromuscular disease.

The fourth work, BWCNN: Blink to Word, a Real-Time Convolutional Neural Network Approach, introduces a deep learning CNN architecture for non-verbal communication that converts

eye-blinks to words, achieving high accuracy and low latency with an InceptionV3 model. In this work, we designed an AI system for non-verbal communication that converts eye-blinks to words using a deep learning CNN architecture. The system predicts the state of the eyes of the patient and finds the blinking pattern. We compared several CNN architectures and hyperparameter selections in model training. For the evaluation, we tested our system using 16,979 facial images and found that our proposed prediction model was efficient and effective. Results demonstrate that overall prediction accuracy is 99.20% and the average prediction time is 94ms. We trained different architectures with different hyperparameters to identify parameter combinations that lead to high accuracy and low latency. For the sake of conducting a clear comparative analysis, we compare the results of each architecture with batch sizes of 8, 16, and 32. SqueezeNet received the lowest accuracy with the fewest parameters. The DenseNet, ResNet, and InceptionV3 acquired accuracies in the range of 99.20% and above. As InceptionV3 had the lowest latency, we chose this architecture. We introduced transfer learning, which improved the convergence when compared to random initialization, to otherwise similar accuracy and latency in the response. Future work includes the training on a more generalized training dataset, the application of hybrid technologies that fuse computer vision techniques such as the one presented together with other natural language processing methods, including recurrent neural networks, to introduce memory in the actions and system.

The fifth work,A Deep Ensemble Learning Approach to Identify Stroke-related Features of Gait, proposes a deep ensemble learning approach to extensively analyze the voluminous biomechanical signals generated in motion labs for post-stroke patients, identifying the most revealing signals and demonstrating strong identification results using the best biomechanical sets. These findings represent a promising step towards gait analysis automation and a better understanding of the impact of stroke on mobility. Accurate recognition and assessment of gait abnormality play an important role in making effective rehabilitation treatment strategies for post-stroke patients. This research proposes a deep ensemble learning approach to extensively analyze the voluminous biomechanical signals generated in motion labs (205 signals, 615 time series, per subject). It answers four questions, including the identification of post-stroke vs. age-matched control, more vs. less affected sides, identification and quantification of the most revealing signals, and quantifying the

impact of GF data. The work demonstrates strong identification results, and the identification performance improves as more biomechanical signals are used. Using deep ensemble learning, the best biomechanical sets produce over 99% identification accuracy in both tasks with and without force place data. This is a promising step towards gait analysis automation better understands the impact of stroke on mobility.

Finally, the sixth work present Walk4Me, a telehealth system that uses Artificial Intelligence (AI) to detect gait characteristics in patients and typically developing peers. The system collects data remotely and in real-time from device sensors, extracts gait characteristics and raw data signal characteristics, and uses machine learning techniques to identify patterns that can facilitate early diagnosis, identify early indicators of clinical severity, and track disease progression across the ambulatory phase of the disease. We have identified several machine learning techniques that differentiate between patients and typically-developing subjects with 100% accuracy across the age range studied and have identified corresponding temporal/spatial gait characteristics associated with each group. The system has the potential to inform early clinical diagnosis, treatment decision-making, and monitor disease progression.

Overall, these works demonstrate the potential of wearable sensors and AI techniques in health-care applications, particularly in the field of human activity recognition. They offer significant implications for healthcare advancement and the application of AI in improving patient outcomes. With further research and development, these systems have the potential to transform the clinical diagnosis, treatment decision-making, and disease progression monitoring.

8.1. Futuer Work

8.1.1. Gait disorder on Community-based. HAR for healthcare has many potential use cases, including moving gait diagnoses from expensive motion labs to the community with inexpensive hardware. Gait analysis can also be used in many other healthcare applications, such as stroke detection, gait modification (to prevent failing), and early detection of diseases. We plan to dig deeper into our current approach to identify the severity and the progression of DMD in addition to detection and expand our techniques to be suitable for community use with commodity hardware. The expansion includes not only collecting DMD data for a wide range of patients but

also a set of other diseases that can benefit from our system. We propose to investigate and build diagnostic tools for the following diseases: Becker muscular dystrophy, Limb-Girdle Muscular Dystrophy, Charcot-Marie-Tooth Disease, Chronic Kidney Disease, and Stroke. Our current system will be used to gather data from multiple different disease/age groups in order to more rigorously test AI (CML and DL) techniques for separating phenotypes.

8.1.2. Identifying stroke from gait based on the multi-camera system and force plates. The multi-Camera-based system is a unique system to transfer the body movement to a skeleton movement, which helps visualize the movement in 3D using a simulation system. The set of collected features can be extracted from gait measurements directly from the virtual skeleton along with using the force plate (force plates are mechanical sensing systems designed to measure the ground reaction forces and moments involved in human movements).

Using force plates is challenging to adopt in community-based experiments since the area covered is small to observe enough gait characteristics. It is also limited to just a few steps only, which is not enough to run a deep gait analysis. To overcome this problem, we propose to use ML models to extract features from the skeleton and find an alternative gait feature to replace the force plate. Measuring the stroke severity. Currently, the stroke severity is measured by directly observing the symptoms and thus lacks a fine step measurement. We propose to use ML-based systems to categorize the patient on a scale based on the severity measured by extracting the quantitative gait features.

8.1.3. Identifying acute stroke based on wearable sensors in real-time. In case of a stroke, the current protocol directs the Ambulance (with the patient) to the nearest hospital to diagnose the stroke and stroke's severity first, then directing the acute stroke patient to a specialized medical center, if needed. Time is a key factor here, and if the acute stroke patients were delivered directly to the specialized medical center instead of the hospital, that would help save lives.

Most diagnosis systems for stroke are image-based analyses (e.g., CT or MRI) that are either expensive or difficult to use for real-time diagnosis. **Proposed work**: Since the patients who suffer from stroke have abnormal movement behavior, Using low-cost wearable sensors such as ECG along with an accelerometer/gyroscope could help capture this abnormal behavior. By analyzing these

sensor signals, ML models can diagnose stroke in real-time. As a result, Ambulances can decide either to go to the nearest hospital or directly go to the specialized medical center. The patient would be directed to the appropriate treatment quickly. This method can potentially be used for other diseases, such as heart disease. The challenge here is how to overcome the noise and distortion in sensor data caused by the Ambulance's movement that can interfere with the patient's movement.

8.1.4. Integrate sensor-based and camera-based to create multi-modal methods to boost performance. As mentioned previously, there are two main methods of applied AI in healthcare: sensor-based and camera-based, applied separately. Even though we plan to focus in-depth into the sensor-based method to solve the time-series problem in real-world healthcare applications, we also plan to explore the challenges and opportunities of using multimodality by combining both the sensor-based and camera-based methods that can be jointly optimized to boost the performance of our system.

APPENDIX-A

TABLE A1. **Results based on a single biomechanical signal.**

Data Source	Positions	Types	More vs. Less affected sides		Post-stroke vs. Age-matched control			
			Right/Left[a]		Right/Left[a]		Center[b]	
			With Force	Without Force	With Force	Without Force	With Force	Without Force
KK	UA	AngAcc	-	**97.62%** (2.35)	-	87.73% (1.44)	-	-
		AngVel	-	92.86% (1.64)	-	85.0% (0.64)	-	-
		CGAcc	-	**97.62%** (1.63)	-	89.09% (0.94)	-	-
		CGPos	-	90.48% (0.89)	-	89.09% (0.9)	-	-
		CGVel	-	95.24% (2.07)	-	90.45% (0.71)	-	-
		DistEndPos	-	88.1% (1.87)	-	88.64% (0.45)	-	-
		DistEndVel	-	90.48% (1.98)	-	90.45% (0.57)	-	-
		ProxEndForce*	92.86% (1.67)	-	90.0% (1.14)	-	-	-
		ProxEndPos	-	90.48% (1.1)	-	90.0% (0.9)	-	-
		ProxEndTorque*	91.27% (1.34)	-	93.18% (0.47)	-	-	-
		ProxEndVel	-	95.24% (2.07)	-	91.36% (0.88)	-	-
	FA	AngAcc	-	92.86% (0.66)	-	85.91% (0.82)	-	-
		AngVel	-	91.27% (1.46)	-	86.82% (0.84)	-	-
		CGAcc	-	94.44% (2.95)	-	86.36% (0.24)	-	-
		CGPos	-	92.86% (1.45)	-	90.45% (1.31)	-	-
		CGVel	-	85.71% (1.08)	-	87.27% (0.58)	-	-
		DistEndPos	-	91.27% (1.72)	-	90.0% (0.58)	-	-
		DistEndVel	-	89.68% (1.69)	-	86.82% (0.76)	-	-
		ProxEndForce*	96.03% (1.63)	-	89.55% (0.89)	-	-	-
		ProxEndPos	-	88.1% (2.5)	-	88.18% (0.77)	-	-
		ProxEndTorque*	91.27% (1.39)	-	93.18% (1.1)	-	-	-
		ProxEndVel	-	90.48% (1.52)	-	89.55% (0.41)	-	-

FT	AngAcc	-	88.89% (0.79)	-	90.91% (0.76)	-	-
	AngVel	-	89.68% (2.03)	-	92.73% (0.94)	-	-
	CGAcc	-	92.86% (1.74)	-	91.36% (0.48)	-	-
	CGPos	-	92.86% (1.67)	-	95.0% (0.4)	-	-
	CGVel	-	95.24% (0.82)	-	91.82% (0.97)	-	-
	DistEndPos	-	93.65% (1.87)	-	95.0% (0.84)	-	-
	DistEndVel	-	94.44% (0.66)	-	91.36% (0.47)	-	-
	ProxEndForce*	**98.41%** (0.51)	-	91.82% (1.4)	-	-	-
	ProxEndPos	-	92.86% (1.34)	-	95.0% (0.49)	-	-
	ProxEndTorque*	95.24% (1.2)	-	89.55% (0.97)	-	-	-
	ProxEndVel	-	94.44% (1.87)	-	94.09% (0.69)	-	-
MF	AngAcc	-	84.92% (1.16)	-	90.0% (0.72)	-	-
	AngVel	-	89.68% (1.84)	-	91.82% (0.81)	-	-
	CGAcc	-	87.3% (2.47)	-	92.27% (1.26)	-	-
	CGPos	-	94.44% (1.11)	-	95.0% (1.05)	-	-
	CGVel	-	93.65% (0.95)	-	93.64% (1.08)	-	-
	DistEndPos	-	89.68% (1.51)	-	94.55% (0.54)	-	-
	DistEndVel	-	92.06% (1.57)	-	90.0% (0.29)	-	-
	ProxEndForce*	50.0% (0.0)	-	57.27% (0.29)	-	-	-
	ProxEndPos	-	92.06% (0.93)	-	**95.91%** (0.76)	-	-
	ProxEndTorque*	50.0% (0.0)	-	57.27% (0.6)	-	-	-
	ProxEndVel	-	93.65% (0.51)	-	94.55% (0.38)	-	-

Model	COM	-	-	-	-	-	88.18% (1.43)
PV	AngAcc	-	-	-	-	-	86.36% (1.93)
	AngVel	-	-	-	-	-	87.27% (1.09)
	CGAcc	-	-	-	-	-	91.82% (0.9)
	CGPos	-	-	-	-	-	85.45% (1.36)
	CGVel	-	-	-	-	-	90.91% (1.36)
	DistEndPos	-	-	-	-	-	89.09% (0.95)
	DistEndVel	-	-	-	-	-	**92.73%** (0.82)
	ProxEndForce*	-	-	-	-	91.82% (1.37)	-
	ProxEndPos	-	-	-	-	-	87.27% (1.93)
	ProxEndTorque*	-	-	-	-	91.82% (1.86)	-
	ProxEndVel	-	-	-	-	-	**92.73%** (1.18)
SK	AngAcc	-	90.48% (1.3)	-	91.78% (1.17)	-	-
	AngVel	-	88.89% (0.79)	-	93.15% (0.75)	-	-
	CGAcc	-	96.03% (0.89)	-	89.95% (1.26)	-	-
	CGPos	-	89.68% (2.24)	-	95.89% (0.48)	-	-
	CGVel	-	92.86% (1.45)	-	91.32% (0.54)	-	-
	DistEndPos	-	92.06% (1.19)	-	94.98% (0.6)	-	-
	DistEndVel	-	95.24% (0.83)	-	93.61% (1.05)	-	-
	ProxEndForce*	**98.41%** (0.36)	-	91.78% (1.09)	-	-	-
	ProxEndPos	-	89.68% (2.03)	-	94.06% (1.19)	-	-
	ProxEndTorque*	92.06% (0.72)	-	93.61% (0.68)	-	-	-
	ProxEndVel	-	96.03% (1.51)	-	90.87% (1.0)	-	-

TA	AngAcc	-	-	-	-	-	85.45% (1.45)
	AngVel	-	-	-	-	-	**92.73%** (1.51)
	CGAcc	-	-	-	-	-	90.91% (1.07)
	CGPos	-	-	-	-	-	91.82% (0.82)
	CGVel	-	-	-	-	-	93.64% (1.13)
	DistEndPos	-	-	-	-	-	91.82% (1.02)
	DistEndVel	-	-	-	-	-	90.91% (0.97)
	ProxEndForce*	-	-	-	-	**95.45%** (1.91)	-
	ProxEndPos	-	-	-	-	-	91.82% (1.6)
	ProxEndTorque*	-	-	-	-	90.0% (0.97)	-
	ProxEndVel	-	-	-	-	-	90.91% (1.6)
TH	AngAcc	-	95.24% (1.61)	-	90.87% (1.22)		
	AngVel	-	95.24% (1.1)	-	91.32% (0.54)	-	-
	CGAcc	-	91.27% (0.92)	-	87.67% (0.96)	-	-
	CGPos	-	93.65% (1.03)	-	89.04% (0.55)	-	-
	CGVel	-	89.68% (1.79)	-	92.24% (0.87)	-	-
	DistEndPos	-	91.27% (0.79)	-	94.98% (0.62)	-	-
	DistEndVel	-	96.83% (0.93)	-	90.41% (1.08)	-	-
	ProxEndForce*	96.83% (0.73)	-	94.52% (0.68)	-	-	-
	ProxEndPos	-	89.68% (1.98)	-	84.93% (0.73)	-	-
	ProxEndTorque*	97.62% (1.69)	-	**97.26%** (0.97)	-	-	-
	ProxEndVel	-	95.24% (1.74)	-	89.95% (0.87)	-	-

LMB	Ankle	Ang	-	92.06% (1.57)	-	96.36% (0.8)	-	-
		AngVel	-	**96.83%** (1.4)	-	92.73% (1.08)	-	-
		Force*	**99.21%** (1.12)	-	92.27% (0.72)	-	-	-
		Mom*	95.24% (1.59)	-	91.36% (1.18)	-	-	-
		Pow*	90.48% (0.83)	-	90.91% (0.49)	-	-	-
	Elbow	Ang	-	93.65% (2.03)	-	95.91% (0.97)	-	-
		AngVel	-	92.86% (1.46)	-	89.09% (1.21)	-	-
	Hip	Ang	-	95.24% (1.26)	-	91.36% (1.08)	-	-
		AngVel	-	93.65% (1.4)	-	92.73% (0.67)	-	-
		Force*	97.62% (0.99)	-	93.64% (1.06)	-	-	-
		Mom*	97.62% (0.99)	-	**94.55%** (1.32)	-	-	-
		Pow*	91.27% (1.1)	-	91.82% (0.32)	-	-	-
	Knee	Ang	-	90.48% (1.89)	-	**97.73%** (0.63)	-	-
		AngVel	-	89.68% (1.4)	-	92.27% (0.68)	-	-
		Force*	97.62% (0.51)	-	92.27% (0.97)	-	-	-
		Mom*	92.06% (1.45)	-	94.52% (0.48)	-	-	-
		Pow*	89.68% (0.83)	-	89.95% (1.18)	-	-	-
	Pelvis-wrtLab	Ang	-	-	-	-	-	90.0% (1.02)
		AngVel	-	-	-	-	-	88.18% (0.64)
	Shoulder	Ang	-	90.48% (1.11)	-	94.04% (1.12)	-	-
		AngVel	-	93.65% (1.26)	-	86.24% (0.96)	-	-
	Trunk-wrtLab	Ang	-	-	-	-	-	**91.82%** (0.0)
		AngVel	-	-	-	-	-	90.0% (2.51)
	Trunk-wrtPelvis	Ang	-	-	-	-	-	90.0% (0.69)
		AngVel	-	-	-	-	-	85.45% (1.82)
FP	GR	COFP*	92.86% (0.92)	-	**90.91%** (0.63)	-	-	-
		FORCE*	**96.03%** (1.3)	-	**90.91%** (1.09)	-	-	-
		FREEMOMENT*	91.27% (1.04)	-	90.0% (1.13)	-	-	-

The symbol "-" means that there is no data corresponding to the data source, position, or type.
[a] Based on the central mass of the body.
[b] Based on the right and left sides of the body.
* Force signal.

Bibliography

[1] A. AARTSMA-RUS, I. GINJAAR, AND K. BUSHBY, *The importance of genetic diagnosis for duchenne muscular dystrophy*, Journal of Medical Genetics, 53 (2016), pp. jmedgenet–2015.

[2] J. ADLER AND D. MALONE, *Early mobilization in the intensive care unit: a systematic review*, Cardiopulmonary Physical Therapy Journal, 23 (2012), pp. 5–13.

[3] L. N. ALFANO, N. F. MILLER, K. M. BERRY, H. YIN, K. E. ROLF, K. M. FLANIGAN, J. R. MENDELL, AND L. P. LOWES, *The 100-meter timed test: normative data in healthy males and comparative pilot outcome data for use in duchenne muscular dystrophy clinical trials*, Neuromuscular Disorders, 27 (2017), pp. 452–457.

[4] A. S. ALHARTHI, S. U. YUNAS, AND K. B. OZANYAN, *Deep learning for monitoring of human gait: A review*, IEEE Sensors Journal, 19 (2019), pp. 9575–9591.

[5] J. C. ALVAREZ, D. ÁLVAREZ, AND A. M. LÓPEZ, *Accelerometry-based distance estimation for ambulatory human motion analysis*, Sensors, 18 (2018), p. 4441.

[6] K. AMINIAN, P. ROBERT, E. JÉQUIER, AND Y. SCHUTZ, *Incline, speed, and distance assessment during unconstrained walking.*, Medicine and science in sports and exercise, 27 (1995), pp. 226–234.

[7] J. AN, Z. XIE, F. JIA, Z. WANG, Y. YUAN, J. ZHANG, AND J. FANG, *Quantitative coordination evaluation for screening children with duchenne muscular dystrophy*, Chaos, 30 (2020), p. 023116.

[8] I. BARTHÉLÉMY, E. BARREY, P. AGUILAR, A. URIARTE, M. CHEVOIR, J.-L. THIBAUD, T. VOIT, S. BLOT, AND J.-Y. HOGREL, *Longitudinal ambulatory measurements of gait abnormality in dystrophin-deficient dogs*, BMC musculoskeletal disorders, 12 (2011), p. 75.

[9] I. BARTHÉLÉMY, E. BARREY, J.-L. THIBAUD, A. URIARTE, T. VOIT, S. BLOT, AND J.-Y. HOGREL, *Gait analysis using accelerometry in dystrophin-deficient dogs*, Neuromuscular Disorders, 19 (2009), pp. 788–796.

[10] J. L. BOWDEN, J. L. TAYLOR, AND P. A. MCNULTY, *Voluntary activation is reduced in both the more-and less-affected upper limbs after unilateral stroke*, Frontiers in neurology, 5 (2014), p. 239.

[11] A. C. CASTRO-AVILA, P. SERON, E. FANG, M. GAETE, AND S. MICKAN, *Effect of early rehabilitation during intensive care unit stay on functional status: systematic review and meta-analysis*, PloS One, 10 (2015), pp. 1–21.

[12] T. CHAU, *A review of analytical techniques for gait data. Part 2: neural network and wavelet methods*, Gait & posture, 13 (2001), pp. 102–120.

[13] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, *Smote: synthetic minority over-sampling technique*, Journal of Artificial Intelligence Research, 16 (2002), pp. 321–357.

[14] G. Chen and C. Patten, *Joint moment work during the stance-to-swing transition in hemiparetic subjects*, Journal of biomechanics, 41 (2008), pp. 877–883.

[15] K. Chen, L. Yao, X. Wang, D. Zhang, T. Gu, Z. Yu, and Z. Yang, *Interpretable parallel recurrent neural networks with convolutional attentions for multi-modality activity modeling*, 2018. arXiv:1805.07233.

[16] S. Chen, J. Lach, B. Lo, and G.-Z. Yang, *Toward pervasive gait analysis with wearable sensors: A systematic review*, IEEE journal of biomedical and health informatics, 20 (2016), pp. 1521–1537.

[17] S. A. Chohan, P. K. Venkatesh, and C. H. How, *Long-term complications of stroke and secondary prevention: an overview for primary care physicians*, Singapore medical journal, 60 (2019), p. 616.

[18] S. A. Conger, L. P. Toth, C. Cretsinger, A. Raustorp, J. Mitáš, S. Inoue, and D. R. Bassett, *Time trends in physical activity using wearable devices: a systematic review and meta-analysis of studies from 1995 to 2017*, Medicine & Science in Sports & Exercise, 54 (2022), pp. 288–298.

[19] A. M. Connolly, J. M. Florence, M. M. Cradock, E. C. Malkus, J. R. Schierbecker, C. A. Siener, C. O. Wulf, P. Anand, P. T. Golumbek, C. M. Zaidman, et al., *Motor and cognitive assessment of infants and young boys with duchenne muscular dystrophy: results from the muscular dystrophy association dmd clinical research network*, Neuromuscular Disorders, 23 (2013), pp. 529–539.

[20] C. Cui, G.-B. Bian, Z.-G. Hou, J. Zhao, G. Su, H. Zhou, L. Peng, and W. Wang, *Simultaneous recognition and assessment of post-stroke hemiparetic gait by fusing kinematic, kinetic, and electrophysiological data*, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 26 (2018), pp. 856–864.

[21] Z. Cui, W. Chen, and Y. Chen, *Multi-scale convolutional neural networks for time series classification*, arXiv preprint arXiv:1603.06995, (2016).

[22] L. M. Dang, K. Min, H. Wang, M. J. Piran, C. H. Lee, and H. Moon, *Sensor-based and vision-based human activity recognition: a comprehensive survey*, Pattern Recognition, 108 (2020), pp. 1–41.

[23] R. De Sanctis, M. Pane, S. Sivo, V. Ricotti, G. Baranello, S. Frosini, E. Mazzone, F. Bianco, L. Fanelli, M. Main, et al., *Suitability of north star ambulatory assessment in young boys with duchenne muscular dystrophy*, Neuromuscular Disorders, 25 (2015), pp. 14–18.

[24] A. Dementyev and C. Holz, *Dualblink: a wearable device to continuously detect, track, and actuate blinking for alleviating dry eyes and computer vision syndrome*, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 1 (2017), pp. 1–19.

[25] W. T. Dempster, *Space requirements of the seated operator, geometrical, kinematic, and mechanical aspects of the body with special reference to the limbs*, tech. rep., Michigan State Univ East Lansing, 1955.

[26] S. Díaz, J. B. Stephenson, and M. A. Labrador, *Use of wearable sensor technology in gait, balance, and range of motion analysis*, Applied Sciences, 10 (2019), p. 234.

[27] J. Díaz-Manera, J. Llauger, E. Gallardo, and I. Illa, *Muscle mri in muscular dystrophies*, Acta Myologica, 34 (2015), pp. 95–108.

[28] M. G. D'Angelo, M. Berti, L. Piccinini, M. Romei, M. Guglieri, S. Bonato, A. Degrate, A. C. Turconi, and N. Bresolin, *Gait pattern in duchenne muscular dystrophy*, Gait & posture, 29 (2009), pp. 36–41.

[29] V. L. Feigin, R. V. Krishnamurthi, P. Parmar, B. Norrving, G. A. Mensah, D. A. Bennett, S. Barker-Collo, A. E. Moran, R. L. Sacco, T. Truelsen, et al., *Update on the global burden of ischemic and hemorrhagic stroke in 1990-2013: the GBD 2013 study*, Neuroepidemiology, 45 (2015), pp. 161–176.

[30] S. Feng and M. Duarte, *Few-shot learning-based human activity recognition*, Expert Systems with Applications, 138 (2019), pp. 1–12.

[31] J. Figueiredo, C. P. Santos, and J. C. Moreno, *Automatic recognition of gait patterns in human motor disorders using machine learning: A review*, Medical engineering & physics, 53 (2018), pp. 1–12.

[32] B. Filtjens, P. Ginis, A. Nieuwboer, P. Slaets, and B. Vanrumste, *Automated freezing of gait assessment with marker-based motion capture and multi-stage spatial-temporal graph convolutional neural networks*, Journal of NeuroEngineering and Rehabilitation, 19 (2022), pp. 1–14.

[33] C. Finn, P. Abbeel, and S. Levine, *Model-agnostic meta-learning for fast adaptation of deep networks*, in Proceedings of the 34th International Conference on Machine Learning, 2017, pp. 1126–1135.

[34] E. G. Fowler, L. A. Staudt, K. R. Heberer, S. E. Sienko, C. E. Buckon, A. M. Bagley, M. D. Sussman, and C. M. McDonald, *Longitudinal community walking activity in duchenne muscular dystrophy*, Muscle & nerve, 57 (2018), pp. 401–406.

[35] B. Fraysse, I. Barthélémy, E. M. Qannari, K. Rouger, C. Thorin, S. Blot, C. Le Guiner, Y. Cherel, and J.-Y. Hogrel, *Gait characterization in golden retriever muscular dystrophy dogs using linear discriminant analysis*, BMC Musculoskeletal Disorders, 18 (2017).

[36] R. Ganea, P.-Y. Jeannet, A. Paraschiv-Ionescu, N. M. Goemans, C. Piot, M. Van den Hauwe, and K. Aminian, *Gait assessment in children with duchenne muscular dystrophy during long-distance walking*, Journal of child neurology, 27 (2012), pp. 30–38.

[37] N. Gaudreault, D. Gravel, S. Nadeau, S. Houde, and D. Gagnon, *Gait patterns comparison of children with duchenne muscular dystrophy to those of control subjects considering the effect of gait velocity*, Gait & posture, 32 (2010), pp. 342–347.

[38] R. C. González, A. M. López, J. Rodriguez-Uría, D. Alvarez, and J. C. Alvarez, *Real-time gait event detection for normal subjects from lower trunk accelerations*, Gait & posture, 31 (2010), pp. 322–325.

[39] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, *Generative adversarial networks*, 2014. arXiv:1406.2661.

[40] K. Grauman, M. Betke, J. Gips, and G. R. Bradski, *Communication via eye blinks-detection and duration analysis in real time*, in Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, vol. 1, IEEE, 2001, pp. I–I.

[41] E. P. Hanavan Jr, *A mathematical model of the human body*, tech. rep., Air Force Aerospace Medical Research Lab Wright-patterson AFB OH, 1964.

[42] G. J. Hankey, *Stroke*, The Lancet, 389 (2017), pp. 641–654.

[43] K. He, X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition*, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[44] K. Heberer, E. Fowler, L. Staudt, S. Sienko, C. E. Buckon, A. Bagley, M. Sison-Williamson, C. M. McDonald, and M. D. Sussman, *Hip kinetics during gait are clinically meaningful outcomes in young boys with duchenne muscular dystrophy*, Gait & posture, 48 (2016), pp. 159–164.

[45] O. Henmi, Y. Shiba, T. Saito, H. Tsuruta, A. Takeuchi, M. Shirataka, S. Obuchi, M. Kojima, and N. Ikeda, *Spectral analysis of gait variability of stride interval time series: comparison of young, elderly and parkinson's disease patients*, Journal of Physical Therapy Science, 21 (2009), pp. 105–111.

[46] F. Horst, S. Lapuschkin, W. Samek, K.-R. Müller, and W. I. Schöllhorn, *Explaining the unique nature of individual gait patterns with deep learning*, Scientific reports, 9 (2019), pp. 1–13.

[47] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, *Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size*, arXiv preprint arXiv:1602.07360, (2016).

[48] Y. Iwasawa, K. Nakayama, I. Yairi, and Y. Matsuo, *Privacy issues regarding the application of dnns to activity-recognition using wearables and its countermeasures by use of adversarial training*, in International Joint Conference on Artificial Intelligence (IJCAI), 2017, pp. 1930–1936.

[49] D. Jarchi, J. Pope, T. K. Lee, L. Tamjidi, A. Mirzaei, and S. Sanei, *A review on accelerometry-based gait analysis and emerging clinical applications*, IEEE reviews in biomedical engineering, 11 (2018), pp. 177–194.

[50] D. Jiménez-Grande, S. F. Atashzar, V. Devecchi, E. Martinez-Valdes, and D. Falla, *A machine learning approach for the identification of kinematic biomarkers of chronic neck pain during single-and dual-task gait*, Gait & Posture, 96 (2022), pp. 81–86.

[51] C. O. Johnson, M. Nguyen, G. A. Roth, E. Nichols, T. Alam, D. Abate, F. Abd-Allah, A. Abdelalim, H. N. Abraha, N. M. Abu-Rmeileh, et al., *Global, regional, and national burden of stroke, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016*, The Lancet Neurology, 18 (2019), pp. 439–458.

[52] N. C. Joyce, B. Oskarsson, and L.-W. Jin, *Muscle biopsy evaluation in neuromuscular disorders*, Physical Medicine and Rehabilitation Clinics, 23 (2012), pp. 609–631.

[53] K. Kaczmarczyk, A. Wit, M. Krawczyk, and J. Zaborski, *Gait classification in post-stroke patients using artificial neural networks*, Gait & posture, 30 (2009), pp. 207–210.

[54] J. J. Kavanagh, S. Morrison, D. A. James, and R. Barrett, *Reliability of segmental accelerations measured using a new wireless gait analysis system*, Journal of biomechanics, 39 (2006), pp. 2863–2872.

[55] K. Keenan, P. Lovoi, and W. Smith, *The neurological examination improves cranial accelerometry large vessel occlusion prediction accuracy*, Neurocritical Care, (2020), pp. 1–10.

[56] S. Khalid, T. Khalil, and S. Nasreen, *A survey of feature selection and feature extraction techniques in machine learning*, in Science and Information Conference, 2014, pp. 372–378.

[57] S. Khandelwal and N. Wickström, *Evaluation of the performance of accelerometer-based gait event detection algorithms in different real-world scenarios using the marea gait database*, Gait & posture, 51 (2017), pp. 84–90.

[58] P. Khera and N. Kumar, *Role of machine learning in gait analysis: a review*, Journal of Medical Engineering & Technology, 44 (2020), pp. 441–467.

[59] E. Kim, *Interpretable and accurate convolutional neural networks for human activity recognition*, IEEE Transactions on Industrial Informatics, 16 (2020), pp. 7190–7198.

[60] R. J. Kuczmarski, C. L. Ogden, S. S. Guo, L. M. Grummer-Strawn, K. M. Flegal, Z. Mei, R. Wei, L. R. Curtin, A. F. Roche, and C. L. Johnson, *2000 cdc growth charts for the united states: methods and development*, Vital and health statistics. Series 11, Data from the National Health Survey, (2002), p. 1—190.

[61] K. V. Laerhoven, A. Schmidt, and H.-W. Gellersen, *Multi-sensor context aware clothing*, in International Symposium on Wearable Computers, 2002, pp. 1–8.

[62] O. D. Lara and M. A. Labrador, *A survey on human activity recognition using wearable sensors*, IEEE Communication Surveys & Tutorials, 15 (2013), pp. 1192–1209.

[63] H.-y. Lau, K.-y. Tong, and H. Zhu, *Support vector machine for classification of walking conditions of persons after stroke with dropped foot*, Human movement science, 28 (2009), pp. 504–514.

[64] A. Le Guennec, S. Malinowski, and R. Tavenard, *Data augmentation for time series classification using convolutional neural networks*, in ECML/PKDD workshop on advanced analytics and learning on temporal data, 2016.

[65] G. C. Le Masurier and C. Tudor-Locke, *Comparison of pedometer and accelerometer accuracy under controlled conditions*, Medicine & Science in Sports & Exercise, 35 (2003), pp. 867–871.

[66] Y. LeCun, Y. Bengio, and G. Hinton, *Deep learning*, Nature, 521 (2015), pp. 436–44.

[67] J. B. Lee, R. B. Mellifont, and B. J. Burkett, *The use of a single inertial sensor to identify stride, step, and stance durations of running gait*, Journal of Science and Medicine in Sport, 13 (2010), pp. 270–273.

[68] Y. Lei and Z. Wu, *Time series classification based on statistical features*, EURASIP Journal on Wireless Communications and Networking, 2020 (2020), pp. 1–13.

[69] R. Liu, S. A. Fazio, H. Zhang, A. A. Ramli, X. Liu, and J. Y. Adams, *Early mobility recognition for intensive care unit patients using accelerometers*, in KDD Workshop on Artificial Intelligence of Things (AIoT), 2021, pp. 1–6.

[70] R. Liu, A. A. Ramli, H. Zhang, E. Henricson, and X. Liu, *An overview of human activity recognition using wearable sensors: Healthcare and artificial intelligence*, in Internet of Things – ICIOT 2021, 2022, pp. 1–14.

[71] D. J. Lott, T. Taivassalo, C. R. Senesac, R. J. Willcocks, A. M. Harrington, K. Zilke, H. Cunkle, C. Powers, E. L. Finanger, W. D. Rooney, et al., *Walking activity in a large cohort of boys with duchenne muscular dystrophy*, Muscle & nerve, 63 (2021), pp. 192–198.

[72] V. Lugade, V. Lin, A. Farley, and L.-S. Chou, *An artificial neural network estimation of gait balance control in the elderly using clinical evaluations*, PloS one, 9 (2014), p. e97595.

[73] S. Mainali, M. E. Darsie, and K. S. Smetana, *Machine learning in action: Stroke diagnosis and outcome prediction*, Frontiers in neurology, 12 (2021), p. 2153.

[74] U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher, *Activity recognition and monitoring using multiple sensors on different body positions*, in International Workshop on Wearable and Implantable Body Sensor Networks (BSN), 2006, pp. 1–4.

[75] E. Mazzone, S. Messina, G. Vasco, M. Main, M. Eagle, A. D'Amico, L. Doglio, L. Politano, F. Cavallaro, S. Frosini, et al., *Reliability of the north star ambulatory assessment in a multicentric setting*, Neuromuscular Disorders, 19 (2009), pp. 458–461.

[76] C. M. Mcdonald, E. K. Henricson, R. T. Abresch, J. M. Florence, M. Eagle, E. Gappmaier, A. M. Glanzman, P.-G.-.-D. S. Group, R. Spiegel, J. Barth, et al., *The 6-minute walk test and other endpoints in duchenne muscular dystrophy: longitudinal natural history observations over 48 weeks from a multicenter study*, Muscle & nerve, 48 (2013), pp. 343–356.

[77] C. M. McDonald, L. M. Widman, D. D. Walsh, S. A. Walsh, and R. T. Abresch, *Use of step activity monitoring for continuous physical activity assessment in boys with duchenne muscular dystrophy*, Archives of physical medicine and rehabilitation, 86 (2005), pp. 802–808.

[78] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Arcas, *Communication-efficient learning of deep networks from decentralized data*, in International Conference on Artificial Intelligence and Statistics (AISTATS), 2017, pp. 1–10.

[79] J.-B. Mignardot, T. Deschamps, E. Barrey, B. Auvinet, G. Berrut, C. Cornu, T. Constans, and L. Dedecker, *Gait disturbances as specific predictive markers of the first fall onset in elderly people: a two-year prospective observational study*, Frontiers in aging neuroscience, 6 (2014), p. 22.

[80] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, *Deep learning for healthcare: review, opportunities and challenges*, Briefings in bioinformatics, 19 (2018), pp. 1236–1246.

[81] S. Mo and D. H. Chow, *Accuracy of three methods in gait event detection during overground running*, Gait & posture, 59 (2018), pp. 93–98.

[82] A. Mohammed and S. A. Anwer, *Efficient eye blink detection method for disabled-helping domain*, Eye, 10 (2014), p. P2.

[83] D. M. Mohan, A. H. Khandoker, S. A. Wasti, S. Ismail Ibrahim Ismail Alali, H. F. Jelinek, and K. Khalaf, *Assessment methods of post-stroke gait: A scoping review of technology-driven approaches to gait characterization and analysis*, Frontiers in Neurology, 12 (2021), p. 650024.

[84] M. Moro, G. Marchesi, F. Hesse, F. Odone, and M. Casadio, *Markerless vs. marker-based gait analysis: A proof of concept study*, Sensors, 22 (2022), p. 2011.

[85] K. Mukherjee and D. Chatterjee, *Augmentative and alternative communication device based on eye-blink detection and conversion to morse-code to aid paralyzed individuals*, in 2015 International Conference on Communication, Information & Computing Technology (ICCICT), IEEE, 2015, pp. 1–5.

[86] M. D. Nguyen, K.-R. Mun, D. Jung, J. Han, M. Park, J. Kim, and J. Kim, *IMU-based Spectrogram Approach with Deep Convolutional Neural Networks for Gait Classification*, in 2020 IEEE International Conference on Consumer Electronics (ICCE), 2020, pp. 1–6.

[87] J. Nonnekes, E. Růžička, T. Serranová, S. G. Reich, B. R. Bloem, and M. Hallett, *Functional gait disorders*, Neurology, 94 (2020), pp. 1093–1099.

[88] M. H. M. Noor, Z. Salcic, and K. I.-K. Wang, *Adaptive sliding window segmentation for physical activity recognition using a single tri-axial accelerometer*, Pervasive and Mobile Computing, 38 (2016), pp. 41–59.

[89] S. Pandian, K. N. Arya, and D. Kumar, *Effect of motor training involving the less-affected side (MTLA) in post-stroke subjects: a pilot randomized controlled trial*, Topics in stroke rehabilitation, 22 (2015), pp. 357–367.

[90] Physiopedia, *North start ambulatory assessment.* `https://www.physio-pedia.com/North_Star_Ambulatory_Assessment`. [accessed on 29-June-2021].

[91] A. A. Ramli, R. Liu, R. Krishnamoorthy, I. B. Vishal, X. Wang, I. Tagkopoulos, and X. Liu, *Bwcnn: Blink to word, a real-time convolutional neural network approach*, in Internet of Things - ICIOT 2020, Cham, 2020, Springer International Publishing, pp. 133–140.

[92] A. A. Ramli, X. Liu, and E. Henricson, *Walk4me system.* `https://albara.ramli.net/research/walk4me/`.

[93] A. A. Ramli, X. Liu, and E. K. Henricson, *Walk4me: Telehealth community mobility assessment, an automated system for early diagnosis and disease progression*, 2023.

[94] A. A. Ramli, A. Nicorici, P. Prasad, J. Hou, C. McDonald, X. Liu, and E. Henricson, *An automated system for early diagnosis, severity, and progression identification in duchenne muscular dystrophy: a machine learning and deep learning approach*, in Annual Human Genomics Symposium – University of California Davis Medical Center, 2020, pp. 12–12.

[95] A. A. Ramli, H. Zhang, J. Hou, R. Liu, X. Liu, A. Nicorici, D. Aranki, C. Owens, P. Prasad, C. McDonald, and E. Henricson, *Gait Characterization in Duchenne Muscular Dystrophy (DMD) Using a Single-Sensor Accelerometer: Classical Machine Learning and Deep Learning Approaches*, arXiv, 2021.

[96] A. A. Ramli, H. Zhang, J. Hou, R. Liu, X. Liu, A. Nicorici, D. Aranki, C. Owens, P. Prasad, C. McDonald, and E. Henricson, *Gait characterization in duchenne muscular dystrophy (dmd) using a single-sensor accelerometer: Classical machine learning and deep learning approaches*, 2021.

[97] K. M. Rashid and J. Louis, *Times-series data augmentation and deep learning for construction equipment activity recognition*, Advanced Engineering Informatics, 42 (2019), p. 100944.

[98] T. B. Rodrigues, D. P. Salgado, C. Ó. Catháin, N. E. O'Connor, and N. Murray, *Human gait assessment using a 3d marker-less multimodal motion capture system*, Multim. Tools Appl., 79 (2020), pp. 2629–2651.

[99] R. Romijnders, E. Warmerdam, C. Hansen, G. Schmidt, and W. Maetzler, *A deep learning approach for gait event detection from a single shank-worn imu: Validation in healthy and neurological cohorts*, Sensors, 22 (2022), p. 3859.

[100] M. Sarshar, S. Polturi, and L. Schega, *Gait phase estimation by using LSTM in IMU-based gait analysis—Proof of concept*, Sensors, 21 (2021), p. 5749.

[101] M. Sharifi Renani, A. M. Eustace, C. A. Myers, and C. W. Clary, *The use of synthetic imu signals in the training of deep learning models significantly improves the accuracy of joint kinematic predictions*, Sensors, 21 (2021), p. 5876.

[102] M. Sharifi Renani, C. A. Myers, R. Zandie, M. H. Mahoor, B. S. Davidson, and C. W. Clary, *Deep learning in gait parameter prediction for OA and TKA patients wearing IMU sensors*, Sensors, 20 (2020), p. 5553.

[103] H. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, *Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning*, IEEE Transactions on Medical Imaging, 35 (2016), pp. 1285–1298.

[104] H. Singh and J. Singh, *Real-time eye blink and wink detection for object selection in hci systems*, Journal on Multimodal User Interfaces, 12 (2018), pp. 55–65.

[105] W. Staab, R. Hottowitz, C. Sohns, J. M. Sohns, F. Gilbert, J. Menke, A. Niklas, and J. Lotz, *Accelerometer and gyroscope based gait analysis using spectral analysis of patients with osteoarthritis of the knee*, Journal of physical therapy science, 26 (2014), pp. 997–1002.

[106] M. Su, C. Yeh, S. Lin, P. Wang, and S. Hou, *An implementation of an eye-blink-based communication aid for people with severe disabilities*, in 2008 International Conference on Audio, Language and Image Processing, IEEE, 2008, pp. 351–356.

[107] D. H. Sutherland, R. Olshen, L. Cooper, M. Wyatt, J. Leach, S. Mubarak, and P. Schultz, *The pathomechanics of gait in duchenne muscular dystrophy*, Developmental Medicine & Child Neurology, 23 (1981), pp. 3–22.

[108] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, *Rethinking the inception architecture for computer vision*, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2818–2826.

[109] R. Tanawongsuwan and A. Bobick, *Performance analysis of time-distance gait parameters under different speeds*, in International Conference on Audio-and Video-Based Biometric Person Authentication, Springer, 2003, pp. 715–724.

[110] D. Trabassi, M. Serrao, T. Varrecchia, A. Ranavolo, G. Coppola, R. De Icco, C. Tassorelli, and S. F. Castiglia, *Machine learning approach to support the detection of Parkinson's disease in IMU-based Gait analysis*, Sensors, 22 (2022), p. 3700.

[111] S. Vafadar, W. Skalli, A. Bonnet-Lebrun, A. Assi, and L. Gajny, *Assessment of a novel deep learning-based marker-less motion capture system for gait study*, Gait & Posture, 94 (2022), pp. 138–143.

[112] F.-C. Wang, S.-F. Chen, C.-H. Lin, C.-J. Shih, A.-C. Lin, W. Yuan, Y.-C. Li, and T.-Y. Kuo, *Detection and classification of stroke gaits by deep neural networks employing inertial measurement units*, Sensors, 21 (2021), p. 1864.

[113] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, *Deep learning for sensor-based activity recognition: a survey*, Pattern Recognition Letters, 119 (2019), pp. 3–11.

[114] A. P. Wibawa, A. B. P. Utama, H. Elmunsyah, U. Pujianto, F. A. Dwiyanto, and L. Hernandez, *Time-series analysis with smoothed Convolutional Neural Network*, Journal of big Data, 9 (2022), pp. 1–18.

[115] A. Wijekoon and N. Wiratunga, *Learning-to-learn personalised human activity recognition models*, 2020. arXiv:2006.07472.

[116] G. Williams, D. Lai, A. Schache, and M. E. Morris, *Classification of gait disorders following traumatic brain injury*, Journal of head trauma rehabilitation, 30 (2015), pp. E13–E23.

[117] E. M. Yiu and A. J. Kornberg, *Duchenne muscular dystrophy*, Journal of Paediatrics and Child Health, (2015), pp. 759–764.

[118] Q. Zhang, Y. N. Wu, and S.-C. Zhu, *Interpretable convolutional neural networks*, in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 8827–8836.

[119] W. Zijlstra and A. L. Hof, *Assessment of spatio-temporal gait parameters from trunk accelerations during human walking*, Gait & posture, 18 (2003), pp. 1–10.