On the Impact and Detection of Biceps Muscle Fatigue in Wearable Sensors-Based Human Activity Recognition

Mohamed Elshafei

A Thesis

in

The Department

of

Computer Science and Software Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy (Software Engineering) at

Concordia University

Montréal, Québec, Canada

June 2022

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Mohamed Elshafei**

Entitled: **On the Impact and Detection of Biceps Muscle Fatigue in Wearable Sensors-Based Human Activity Recognition**

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Software Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

_____ Chair
Dr. Anjan Bhowmick

_____ External Examiner
Dr. Olga Baysal

_____ Examiner
Dr. Marta Kersten

_____ Examiner
Dr. Tristan Glatard

_____ Examiner
Dr. Roch Glitho

_____ Supervisor
Dr. Emad Shihab

Approved by _____
Dr. Lata Narayanan, Chair
Department of Computer Science and Software Engineering

29 June 2022 _____
Dr. Mourad Debbabi, Dean
Gina Cody School of Engineering and Computer Science

Abstract

On the Impact and Detection of Biceps Muscle Fatigue in Wearable Sensors-Based Human Activity Recognition

Mohamed Elshafei, Ph.D.

Concordia University, 2022

Nowadays, modern sport and athletic training are very interested in wearable-based Human Activity Recognition (HAR) systems due to their cost-efficiency, portability, and convenience. However, this leads the developers to compete in developing the various HAR applications with little attention to HAR's-related problems such as fatigue. In this thesis, we select the bicep curls as an example of a HAR activity to study the fatigue problem in wearable-based HAR. We approach the fatigue problem through three studies: first, we study the impact of fatigue in wearable-based HAR. Second, we detect the presence of fatigue during human activity, e.g., biceps curls exercise. Third, we improve the performance of fatigue detection models while reducing the test's data consumption. Throughout our studies, we use our dataset, which consists of 3,750 repetitions of biceps curls from twenty-five volunteers between 20–46 years and with body mass index (BMI) between 24–46.

Our first study on the impact of fatigue in wearable-based HAR shows that fatigue often occurs in later sets of biceps curls. During fatigue, the completion time of later sets extends by up to 31%, while muscular endurance decreases by 4.1%. Also, our study shows that changes in data patterns often occur during fatigue, turning some features to be statistically insignificant. This can lead to a substantial decrease in performance in both subject-specific and cross-subject models. In addition, muscle fatigue can lead to various

injuries such as muscle strain and tendons rupture, which may require up to 22 weeks of treatment. Therefore, it is essential to be aware of fatigue during human activity, which we address in our second study.

The second study proposes a wearable-based approach to detect fatigue in biceps curls. We provide a set of 16 most fatigue representative features from 33 extracted features. Then, we employ these features in five models to detect fatigue in biceps curls. Our study shows that a two-layer FNN achieves the highest accuracy of 98% and 88% for subject-specific and cross-subject models, respectively. We observe that the cross-subject models are preferable for a large crowd since these models can utilize crowd data. However, we observe that inter-subject data variability is usually high in the large crowd due to the physical differences among the individuals, resulting in different data patterns for the same activities. As a result, researchers may suggest using subject-specific models for each user in the crowd to achieve higher performance. Still, such a performance comes with a higher data cost of the user's subject-specific model; therefore, improving fatigue detection in cross-subject models is essential, which is the goal of our third study.

In the third study, we propose a personalization approach as a solution to improve the cross-subject models' performance by utilizing data from the crowd based on similarities between the test subject and users from the crowd. We extract 11 hand-crafted features to measure the similarities between the test subject and the individuals in the crowd. Then, we employ these similarities to prioritize and select the training data from the crowd for two cross-subject models. Our study shows that the personalization approach improves the performance of the cross-subject models in terms of precision by up to 7.25%, recall by up to 5.69%, accuracy by up to 6.67%, and F1-measure by up to 6.52%. Furthermore, adding 20% of the test subject's data into the training dataset of the personalized cross-subject models can produce accurate results closer to the ones from subject-specific models.

Related Publications

The following publication are related to the materials presented in this thesis:

- **Elshafei, M.**, Costa, D. E., & Shihab, E. (2021). On the Impact of Biceps Muscle Fatigue in Human Activity Recognition. Sensors, 21(4), 1070.

- **Elshafei, M.**,& Shihab, E. (2021). Towards detecting biceps muscle fatigue in gym activity using wearables. Sensors, 21(3), 759.

- **Elshafei, M.**, Costa, D. E., & Shihab, E. (2022). Toward the Personalization of Biceps Fatigue Detection Model for Gym Activity: An Approach to Utilize Wearables' Data from the Crowd. Sensors, 22(4), 1454.

The following publications are not directly related to the material presented in this thesis but were conducted as parallel work to the research presented in this thesis.

- Abdellatif, A., Zeng, Y., **Elshafei, M.**, Shihab, E., & Shang, W. (2020). Simplifying the search of npm packages. Information and Software Technology, 126, 106365.

Dedication

To my mother, in loving memory.

Acknowledgments

First and foremost, I thank Almighty Allah for providing me with the strength and the ability to pursue my Ph.D. thesis.

I want to express my gratitude and appreciation to my supervisor Dr. Emad Shihab for his informative guidance, infinite support, and endless patience during my Ph.D. journey. Thanks for believing in me and for the immense effort you have put into my success. Also, I would like to thank my committee members, Dr. Marta Kersten, Dr. Tristan Glatard, Dr. Roch Glitho, and Dr. Olga Baysal. Their support and insightful comments have enhanced this thesis. I have also had the privilege to collaborate and discuss my research with great researchers, Drs. Rabe Abdalkareem and Diego Costa. Thanks for sharing your insights and advice.

My appreciation extends to my colleagues, Suhaib Mujahid, Ahmad Abdellatif, Mahmoud Alfadel, Giancarlo Sierra, Hosein Nourani, Atique Reza, Xiaowei Chen, Mouafak Mkhallalati, and everyone else in the Data-driven Analysis of Software (DAS) Lab.

To my parents, rest in peace, for I have fulfilled your will. Dad, thanks for your long-lasting wise words and guidance. Mom, thank you for literally everything.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Research Statement

Human Activity Recognition (HAR) is one of the active research areas in pervasive computing that monitors the human body's movements or gestures via sensors to detect human activities (Golestani & Moghaddam, 2020). HAR utilizes learning algorithms and data analysis techniques to comprehend human activities from input data sources, such as sensors and multimedia devices (Dang, Hassan, Im, & Moon, 2019). As a result, HAR plays an essential role in ubiquitous computing, which may involve direct or indirect interactions between humans and smart devices (Dang et al., 2020). For example, HAR is often the primary and innovative comportment in surveillance systems (Jalal, Kim, Kim, Kamal, & Kim, 2017; X. Ji, Cheng, Feng, & Tao, 2018), behavior analysis (Batchuluun, Kim, Hong, Kang, & Park, 2017), gesture recognition (Pigou, Van Den Oord, Dieleman, Van Herreweghe, & Dambre, 2018), and various healthcare systems (Aviles-Cruz, Rodriguez-Martinez, Villegas-Cortez, & Ferreyra-Ramirez, 2019; Qi, Yang, Hanneghan, Tang, & Zhou, 2018).

Nowadays, sensor technology has achieved exceptional development in multiple perspectives, including computational power, size, accuracy, and manufacturing costs (Liu,

Nie, Liu, & Rosenblum, 2016). These advancements enable the integrations of various sensors in smart wearables, resulting in more convenient, cost-efficient, and portable wearable-based HARs (Dang et al., 2020). Furthermore, many of today's handheld devices (e.g., smartphones and watches) contain inertia sensors like accelerometers, magnetometers, and gyroscopes, which boost users' views in favor of wearable-based HARs (Ignatov, 2018; Ramanujam, Perumal, & Padmavathi, 2021). Recent surveys show that scientists and physicians in sports science often utilize wearable-based HARs in their examinations or research due to their performance, portability, and power-efficient (Dang et al., 2020; Demrozi, Pravadelli, Bihorac, & Rashidi, 2020; Fu, Damer, Kirchbuchner, & Kuijper, 2020; Nweke, Teh, Al-Garadi, & Alo, 2018). However, they also present some obstacles of their works in the form of challenges in wearable-based HARs, such as transfer learning, lack of datasets, and subject exhaustion, also known as fatigue (K. Chen et al., 2021; Nweke et al., 2018). Despite the plethora of works in the literature about wearable-based HAR systems in sports and daily lives, little is known about the impact of HAR challenges, specifically fatigue, on HAR's performance (Enoka & Duchateau, 2016; Ramasamy Ramamurthy & Roy, 2018). Also, a prior work shows that despite such challenges can affect HAR's performance, developers are more likely to compete in developing more cost-effective HAR systems rather than addressing HAR challenges (Demrozi et al., 2020).

In this thesis, we focus on addressing the muscle fatigue challenge in the wearable-based HAR systems and demonstrate how such a challenge can affect body movements to the point where altering sensory data patterns may impair the performance of HAR systems (Elshafei, Costa, & Shihab, 2021). There are three reasons of interest in this particular challenge, which are:

(1) Muscle fatigue is one of the most recurring HAR challenges, especially in recent wearable-based HAR applications (Biagetti, Crippa, Falaschetti, Orcioni, & Turchetti, 2017).

(2) Muscle fatigue often occurs during over-training or intensive physical activity, which push muscles into a vulnerable state, where muscles are prone to fatigue injuries (Kellmann, 2010; Opar, Williams, & Shield, 2012).

(3) Fatigue-induced muscle injuries pose a devastating threat to muscles to the extent of losing muscle strength and flexibility permanently (Thalman, Lam, Nguyen, Sridar, & Polygerinos, 2018).

After considering the above three reasons, we realize the necessity to address such a challenge for the wearable-based HAR systems. We choose biceps concentration curls, or biceps curls for short, as an example of a HAR activity because it involves flexing one of the most active skeletal muscles, namely biceps, at the elbow joint countless times to pick, lift, and pull objects (Steffen et al., 2006; Troiano et al., 2008). Also, fatigue-induced biceps injuries may delay athletes' training schedules for weeks, forcing their immediate withdrawal from competitions sometimes (Hopkins, Marshall, Quarrie, & Hume, 2007).

1.1 Research Statement

Motivated by the challenge above, we propose three studies that together form the goal of this thesis: detect biceps muscle fatigue and study its impact on the wearable-based HAR systems. We hypothesize that muscle fatigue impacts the collected data by altering its patterns, leading to a snowball effect, hindering the extracted features and HAR models' performance. Also, detecting biceps muscle fatigue during activity, e.g., bicep concentration curls, can help individuals to avoid fatigue-induced injuries and reduce the risk of long-term muscle injuries. Therefore, we state our research statement as follows:

> "Given the literature's abundance of works on wearable-based HAR systems and the devastating effects of muscle fatigue-induced injuries, we believe it is

time to address the muscle fatigue challenge in these systems. Therefore, we conduct three studies: 1) study the impact of muscle fatigue in wearable-based HAR system, 2) detect the presence of muscle fatigue during exercises, and 3) propose an approach to improve fatigue detection and reduce data consumption in HAR systems."

1.2 Thesis Overview

In this section, we provide an overview of the works presented in this thesis and highlight the main results of each work.

Chapter 2: Background and Related Work

Before diving into fatigue detection in bicep muscles, we first present a background of muscle fatigue, biceps muscles, and related fatigue injuries. Then, we present studies related to fatigue detection in the literature where we discuss three approaches for fatigue detection. After that, we present the studies related to fatigue's impact in HAR to highlight the importance of addressing such a challenge.

Chapter 3: Dataset Collection, Processing, and Challenges

To carry out our experiments throughout the three proposed studies, we need to have a high-quality dataset with sufficient data entries. Our dataset consists of 3,750 bicep curl repetitions from 25 volunteers ranging between 20 and 46 years old. We provide the volunteers with a 4.5 kg weight dumbbell and attach a 50 Hz Neblina inertial measurement unit (IMU) to their wrist and an Apple Watch Series 4 to their opposite wrist during the exercise. We explain and provide the Borg's scale to each volunteer, which allows them to express their fatigue level through the rate of perceived exertion (RPE).

Also, we discuss the applied solutions to two challenges encountered during our data collection procedure. The first challenge is the dumbbell weight, where we have to select a suitable dumbbell so that volunteers do not reach a fatigued state quickly, resulting in few data points, nor do they last long during the exercise and produce many redundant data points. The second challenge is the familiarity of Borg's scale, where we try to introduce the concept of fatigue self-evaluation to the volunteers so that they do not misevaluate their fatigue by over or under-estimating their perceived exertion rates.

Chapter 4: On the Impact of Biceps Muscle Fatigue in Human Activity Recognition

With the rapid development of wearables and smart devices, the number of HAR applications has been increasing proportionally. In light of this development, the developer took it upon themselves to compete in developing the various HAR applications with little attention to the side/hidden problems such as fatigue (Demrozi et al., 2020). There is no work so far that focuses on studying the impacts of fatigue on HAR systems; instead, fatigue is just presented in online discussions or informal literature and interviews as unwanted noise in datasets. Therefore, we use the biceps data we collected as an example of a HAR activity to observe the impact of fatigue in HAR. Then, we quantify the fatigue share in our dataset and locate its presence. Also, we study the possible data pattern changes during fatigue presence and their impact on significant features. Finally, we examine the decrease in performance in both subject-specific and cross-subject models during fatigue presence.

Chapter 5: Towards Detecting Biceps Muscle Fatigue in Gym Activity Using Wearables

After realizing the impact of fatigue on HAR systems, the question of whether one should be aware of fatigue presence is an important complementary decision to improve HAR systems and prevent fatigue-induced injuries. Although there have been several works on fatigue detection in the literature, the rapid development of wearables has shown promising results in monitoring fatigue levels. Therefore, we adopt a wearable approach to detect biceps muscle fatigue during a bicep concentration curl exercise as an example of gym activity. We use the aforementioned dataset to extract fatigues' significant features. These features are the most overall representative and correlated with bicep curl movement, yet they are fatigue-specific features. Then, we utilize these features in five fatigue detection models, including subject-specific and cross-subject models.

Chapter 6: The Personalization of Biceps Fatigue Detection Model For Gym Activity: An Approach To Utilize Wearables' Data From The Crowd

We investigate an approach that can help us to improve the performance of HAR systems which tends to degrade the most in the case of cross-subject models. Although subject-specific models tend to outperform the cross-subject ones, those models have a higher demand for subjective data. This is an obstacle for developing high-performance fatigue detection models. On the other hand, if we look at cross-subject models, we find that these models suffer from lower performance but can utilize crowds' data. Therefore, we plan to boost cross-subject models' performance while reducing the demand for data from the test subject. We propose a personalized model to detect biceps muscle fatigue that uses data from the crowd for training; in addition, we inject a small portion of the test's

data to boost its performance while training. We measure the similarities between the test subject and the subjects in the crowd. Then, we rank the crowd's data according to their similarities factor. Thus, subjects similar to the test subject will have more shares and effects in training the personalized model. Then, we will evaluate the personalized model's performance in compression with both subject-specific and cross-subject models.

1.3 Thesis Contributions

The main contributions of the thesis are:

(1) Addressing bicep muscle fatigue challenge, one of the most recurring HAR challenges in recent wearable-based HAR applications.

(2) Studying the impact of bicep muscle fatigue using the wearable-based HAR approach instead of the clinical approaches.

(3) Detecting bicep muscle fatigue during exercise to avoid fatigue-induced injuries.

(4) Providing a more practical and feasible approach for fatigue detection in daily life compared to early approaches.

(5) Propose a personalization approach to reduce the demand for test subjects' data and improve the performance of cross-subject models by utilizing wearables' data from the crowd.

(6) Utilize the similarities between test subjects and individuals from the crowd to reduce inter-subject variability in the training datasets for the fatigue detection models.

(7) Publishing a biceps muscle fatigue dataset in the form of a concentration curl exercise. The columns in the dataset show sensory data from the 3-axis of the accelerometer, gyroscope, and magnetometer separately, while the raws present a point in time

where the data are sampled. (bicep fatigue dataset—https://zenodo.org/record/3698242#.XmFZ5qhKguU).

1.4 Thesis Organization

This thesis is organized as follows: Chapter 2 provides general background on HAR and presents the literature review. In Chapter 3, we provide the details of the data collection process and discuss the related challenges. Then, we present our first study on the impact of bicep muscle fatigue in wearable-based HAR systems in Chapter 4. Chapter 5 presents our second study on biceps muscle fatigue detection in bicep curls using a wearable approach. Last but not least, in Chapter 6, we present our third study on the personalization of bicep muscles fatigue detection and the utilization of wearables' data from the crowd. Chapter 7 summarizes the thesis, lists the work limitation, and proposes future work.

Chapter 2

Background and Related Work

This chapter provides an overview of the relevant background to our research, including a brief description of muscle fatigue, biceps muscles, and related injuries. Also, we present reasons of interest in muscle fatigue as a challenge in the wearable-based HAR systems. Then, we list the related works to fatigue detection and the impact of fatigue in the literature.

2.1 Background

Nowadays, HAR systems have become a task of high interest within the data science and sports field, where physical activities often describe any bodily movement produced by skeletal muscles result in energy expenditure above resting level (Hsu, Yang, Chang, & Lai, 2018). Fatigue is a natural phenomenon that describes physiological impairments or lack of energy often caused by prolonged activities (Enoka & Duchateau, 2008). A previous study classifies fatigue into two types, subjective and objective fatigues (Enoka & Duchateau, 2016). The subjective fatigue causes a decline in alertness and mental concentration due to intense mental tasks (De, 1984). In comparison, objective fatigue, also known in the literature as muscle fatigue, causes a decrease in the capability to exert mechanical work

due to intense physical activities (Gruet et al., 2013). During muscle fatigue, the capability to exert physical activities decreases while the risk of fatigue-induced injury increases. Such injuries may require up to 22 weeks of treatment, while tendons rupture may result in a substantial loss in muscle' strength permanently (Enoka & Duchateau, 2016; Ma, Li, Cao, Wang, & Wu, 2014; Nesterenko, Domire, Morrey, & Sanchez-Sotelo, 2010).

Biceps is a muscle in the anterior compartment of the upper arm, along with the brachialis muscle and the coracobrachialis muscle (Bogart & Bogart, 2007). Commonly, biceps represents an attribute of strength within a variety of worldwide cultures (Mueller-Wohlfahrt et al., 2013). Biceps is one of the most active functional skeletal muscle where, it flexes the arm at the elbow joint countless times each day for picking, lifting, and pulling objects (Steffen et al., 2006; Troiano et al., 2008). Also, biceps collaborate between brachialis and brachioradialis for flexion at the elbow joint, as well as utilizes shoulders and back muscles as stabilizers. Although it seems that biceps are essential for upper limb skeletal muscle movements, it has limited repetitive movements due to its placement in the anterior compartment of the upper arm. These repetitive movements stress the biceps' structures (e.g., muscle tissues, tendons, or joints) over time, leading to muscle fatigue. Unfortunately, fatigue reduces muscle exertion gradually, until it exceeds the structure's stress tolerance, where overuse injuries occur (Nesterenko et al., 2010).

Previous studies on muscle injuries indicate that muscle fatigue often occurs priorly, making the muscles vulnerable to fatigue injuries (Mueller-Wohlfahrt et al., 2013; Opar et al., 2012). Biceps muscle fatigue injuries include muscle tear, contusion, and tendons rupture. In extreme cases, muscle fatigue may cause rhabdomyolysis, a potentially life-threatening condition resulting from the breakdown of skeletal muscle fibers and leakage of muscle contents into the circulation (Garrett Jr, 1996; Mair, Seaber, Glisson, & Garrett JR, 1996; Nesterenko et al., 2010). Also, biceps injuries can escalate to a series of intricacies in

the lower-back and upper-limb that can hinder pronation and supination movements (Thalman et al., 2018). Economically, fatigue-induced injuries can result in a total cost of $190 billion and over 1.1 million lost days of work yearly (LEIGH, 2011; of Labor Statistics, 2016). Hence, it is essential to detect muscle fatigue priorly to prevent fatigue-induced injuries; however, the proposed approaches in the literature are often too complicated to be practically integrated into a person's daily life.

2.2 Related Work

In this section, we present the works most related to this thesis. We divide the prior works into two main areas; work related to fatigue detection in the literature and impact of fatigue in the literature.

2.2.1 Fatigue Detection in the Literature

Muscle fatigue is a complex and multifaceted phenomenon with various definitions. A common definition for muscle fatigue is the failure to maintain the required force to continue performing a task (Robergs, Ghiasvand, & Parker, 2004). While, a quantifying definition for muscle fatigue is the decline in the maximal force or the power capacity of the muscles after performing a prolonged task (Enoka & Duchateau, 2008). Nowadays, literature is abundant on fatigue detection approaches to monitor fatigue and reduce the risk of fatigue-induced injuries. There are three main categories for fatigue detection approaches: the invasive approach, the cardio-respiratory approach, and the wearable approach (Abbood, Al-Nuaimy, Al-Ataby, Salem, & AlZubi, 2014; Halson, 2014).

The Invasive Approach

The approach is one of the earliest methods used to detect fatigue. This approach usually requires instruments to puncture the skin or contact with the mucosa. A previous work uses blood lactate concentration test to muscle endurance in athletes to reduce the risk of over-training and injuries (Bosquet, Léger, & Legros, 2001). The work's findings show that blood lactate analysis provides high accuracy (up to 97%) in evaluating muscle endurance; however, it often requires several blood samples from the swimmers during and after progressive incremental swimming. Another previous work measures, the lactic acid in the bloodstream to determine the maximal muscle effort that a person can maintain without risking fatigue injuries (Stoudemire et al., 1996). The work's findings indicate that blood lactate concentration levels are significantly different (P <0.05) in muscles during moderate to fatigue intensities. Another previous work measures lactate and creatine kinase levels in the bloodstream to assess the risk of skeletal muscle injuries during a marathon run (Kobayashi, Takeuchi, Hosoi, Yoshizaki, & Loeppky, 2005). The work's findings indicate that high levels of lactate and creatine kinase can indicate insufficient oxygen intake to the muscles, causing fatigue. In some extreme levels of creatine kinase, the injuries become inevitable. A less painful method was presented in a previous work that measures rectal temperature to predict exercise duration until fatigue occurs in different environmental conditions (Crewe, Tucker, & Noakes, 2008). The work's findings show that rectal temperature increased linearly throughout exercise trials and correlated significantly (r = 0.92) with exerted force to predict the duration of exercise to fatigue. Although the aforementioned approach provides an accurate estimation of fatigue, a practical drawback is that it requires several blood samples during different incremental exercise stages.

The Cardio-Respiratory Approach

The approach is based on a person's metabolic system and it requires a face mask to measure the rate of oxygen intake during an exercise. Some studies refer to this approach as VO_2 max, which stands for the maximum volume of oxygen consumption measured during incremental exercise. A previous work measures the circulatory and respiratory systems' ability to supply oxygen (O_2) to skeletal muscles during sustained physical activity without risking fatigue injuries (Robson-Ansley, Gleeson, & Ansley, 2009).The work's findings indicate that Oxygen consumption provides an accurate indication (>94%) of exercise intensity, yet this requires costly equipment and technical expertise, which outweighs its usefulness for quantifying load during routine training. Another previous work measures the volume of oxygen consumption (VO_2) to determine the time to reach fatigue between various runners (Billat & Koralsztein, 1996). The work's findings indicate that the runner's body requires far more oxygen consumption at the maximum speed, which cannot be satisfied; therefore, the oxygen debt continuously increases, causing the body to slow down—also known as fatigue. Another previous work measures the volume of oxygen consumption (VO_2) to study the development of muscle fatigue in healthy humans during incremental cycling (Kobayashi et al., 2005). The work's findings indicate that the rate of oxygen consumption increases as resistance increases to postpone the development of muscle fatigue evoked by incremental cycling. Another previous work shows that a reduction in work efficiency, also known as fatigue, results from an additional energy cost and oxygen requirement during high-intensity exercise (Cannon, White, Andriano, Kolkhorst, & Rossiter, 2011). The work's findings show that muscle fatigue can be detected while performing the (VO_2) max test, as the body cannot maintain maximum (VO_2) values until it fully recovers from fatigue. Although the aforementioned approach provides an accurate estimation of fatigue, a practical drawback is that the required setup of equipment is too complex to be operated singularly.

The Wearable Approach

The approach is a promising approach which recently developed to overcome such previous drawbacks. A previous work (Seshadri et al., 2019) uses mouth-guard and Galvanic Skin Response (GSR) biosensors to monitor athletes' metabolites from saliva and eccrine sweat continuously. The work's findings show that monitoring biomarkers from saliva or sweat allows us to detect up to 95% of over-training during incremental training accurately. Another previous work uses an Inertial Measurement Unit (IMU) to collect data from outdoor marathon runners and analyzes the data using Machine Learning to predict fatigue (Op De Beéck, Meert, Schütte, Vanwanseele, & Davis, 2018). Another previous work used wearable Electromyography (EMG) to evaluate workers' muscle fatigue as a means of assessing their physical stress on construction sites (Jebelli & Lee, 2019).

2.2.2 Impact of Fatigue in the Literature

There are a plethora of works on the impact of fatigue in the literature; however, most of these works focus on the physiological implications. Outwardly, fatigue impacts human performance through degradation of exerted force, while internally, it impacts heart rate, blood pressure, and core temperature, which can be measured using the appropriate tools. Some studies adopt the clinical approach to study human fatigue, where they provide an in-depth definition, characterization, and examination of fatigue. Clinically, the impact of fatigue is not limited to short periods of exhaustion only, but it often has a prolonged dormant effect. For example, before the pre-season, professional Basketball athletes always go under long training sessions, which may result in an accumulation of perceptual and performance fatigue (Edwards et al., 2018). In this case, fatigue's prolonged impact is measured through the creatine kinase enzyme test, where muscle cells often release the enzyme into the blood reflecting after heavy exercise, reflecting the severity of muscle damage. Several studies on the impact of fatigue on the human body show that the impairment

of muscle contraction is the superficial remanence of fatigue while other impacts of fatigue remain at the cellular level (E. Debold, 2015; E. P. Debold, Walcott, Woodward, & Turner, 2013; Karatzaferi, Franks-Skiba, & Cooke, 2008; Theofilidis, Bogdanis, Koutedakis, & Karatzaferi, 2018). Such impacts often derive from either : (a) alterations in excitability of the muscle fiber, (b) accumulation of metabolic by-products, (c) production of reactive oxygen species, and (d) Ca2+ movements in the fiber compartments.

Our literature survey shows that previous studies usually cover the impact of fatigue on human performance and internal body changes. Also, these studies often focus on clinical approaches that measure levels of lactate, creatine kinase, and VO_2 max (Orizio, Gobbo, Diemont, Esposito, & Veicsteinas, 2003; Sadoyama & Miyano, 1981; Smith & Newham, 2007). On the other hand, in this thesis, we study the impact of fatigue on detection models using the recent wearable approach instead of the clinical ones. Furthermore, we focus on how fatigue impacts the collected data, extracted features,and performance of detection models rather than focusing on human performance and internal body changes as in the clinical approaches. Fatigue may naturally occur in any human activity, but it poses a bigger challenge for HAR models when identifying physically demanding activities, such as gym activities. For this reason, we focus on studying the biceps concentration curls exercise, which involves flexing one of the most active skeletal muscles at the elbow joint.

2.3 Summary

This chapter provides background about muscle fatigue, biceps muscles, and their relevant definitions. Then, it surveys previous works on fatigue detection and the impact of fatigue. In the next chapter, we describe our dataset and related challenges.

Chapter 3

Dataset: Collection, Processing, and Challenges

This chapter details our data collection and processing. Also, it presents possible solutions to overcome such challenges to provide a high-quality dataset for the thesis. The main goal in this chapter is to present the process of building a dataset with a sufficient number of data entries and the least number of defects for three reasons:

(1) To have enough data entries that capture bicep muscle fatigue.

(2) To observe the variations of fatigue across the volunteers during the exercise.

(3) To capture the kinetic changes that occur due to fatigue during the exercise.

Our dataset consists of 3,750 concentration curl repetitions from 25 male volunteers who are diverse in age, ranging between 20 and 46. Previous studies show that the selected age covers three distinct stages of athletes' performance: early, middle, and late, where athletes usually notice physical declines (Adirim & Cheng, 2003; Burt & Overpeck, 2001). Also, the volunteers are diverse in weight and height, ranging between 69–127 and 165–190, respectively. It is essential to have such diversity because physical characteristics such

as weight and height may affect arm movements and the severity of injuries, including fatigue-induced ones (Green & Pizzari, 2017). We calculated the Body Mass Index (BMI) for the twenty-five volunteers using the following formula Prentice and Jebb (2001):

$$\text{BMI} = \frac{\text{Weight (kg)}}{\text{Height(m)}^2}$$

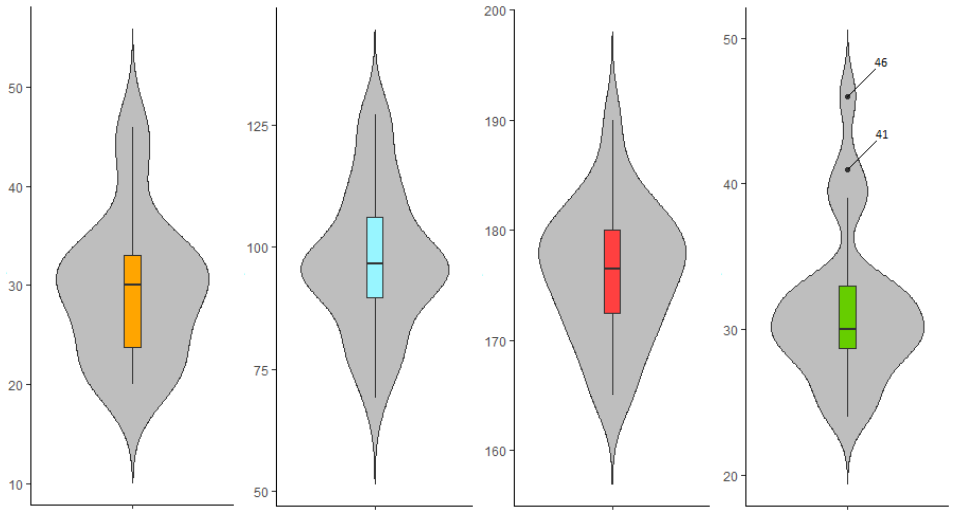


Figure 3.1: Boxplots to display the distribution of volunteers' age, weight, height, and BMI in our dataset. (**a**) Age (years). (**b**) Weight (kg). (**c**) Height (cm). (**d**) BMI (kg/m^2).

Although our dataset is too small to represent nationwide, the volunteers in this dataset are diverse in BMI, ranging between 24 and 46, to include normal weight ($18.5 \leq \text{BMI} \leq 24.9$), overweight ($25 \leq \text{BMI} \leq 29.9$), and obesity ($30 < \text{BMI}$) (Borga et al., 2018). The two black dots in Figure 3.1d are for two volunteers with outlier BMI values of 41 kg/m^2 and 46 kg/m^2 who are considered extremely obese (Burkhauser & Cawley, 2008). Previous studies show a correlation between BMI and risks of overexertion injuries, where trainees with the lowest BMIs exhibit the highest injury risks for both genders and across all fitness levels (Janssen, Katzmarzyk, & Ross, 2002; Jones et al., 2017). In addition, the volunteers have no chronic diseases, no muscle or bone surgeries, and have been gym-goers for at

least 1 year. Moreover, the volunteers are not on prescribed drugs or substances expected to affect their physical performance.

3.1 Dataset Collection

First of all, our dataset collection process meets the guidelines of the Declaration of Helsinki and is approved by Concordia University's Faculty Research & Ethics Advisory Committee (summary protocol form number: 30008716). Also, all volunteers must sign an informed consent before participating in the data collection process. We ask twenty-five volunteers to perform concentration curls while we use the following tools to construct our dataset: (1) We attach a 50 Hz Neblina inertial measurement unit (IMU) to the volunteer's wrist to measure its acceleration and calculate the linear and angular velocities. Previous studies show velocity loss as an early indicator of muscle fatigue during resistance training, especially when blood lactate and ammonia accumulate in muscle tissues (Apriantono, Nunome, Ikegami, & Sano, 2006; Coelho et al., 2015; Sanchez-Medina & González-Badillo, 2011). (2) We attach an Apple Watch Series 4 to the volunteer's opposite wrist to measure their heart rate during the exercise. (3) We provide the volunteers with a 4.5 kg weight dumbbell to perform concentration curls. (4) We provide the volunteers with Borg's scale presented in Table 3.1 to express their fatigue levels. Such a scale is often used as a subjective method to estimate the rate of perceived exertion (RPE), which expresses the fatigue intensity during an exercise. During data collection sessions, we ask each volunteer to complete 5 warm-up repetitions followed by 15 repetitions per set for a total of 5 sets per hand, as shown in Figure 3.2. Moreover, the volunteers report their RPE after each set, including the warm-up, yielding 6 RPE values per hand.

Table 3.1: Borg G.A. Psychophysical bases of perceived exertion (G. A. Borg, 1982).

Perceived Exertion	Borg Rating	Examples
None	6	Reading a book, watching television
Very, very light	7 to 8	Tying shoes
Very light	9 to 10	Chores like folding clothes with little effort
Fairly light	11 to 12	Slow Walking (without speeding breath)
Somewhat hard	13 to 14	Brisk walking (with effort and speeding breath)
Hard	15 to 16	Bicycling (high effort and heart pounding)
Very hard	17 to 18	Intense activity but can be sustained
Very, very hard	19 to 20	Very intense activity that can't be sustained

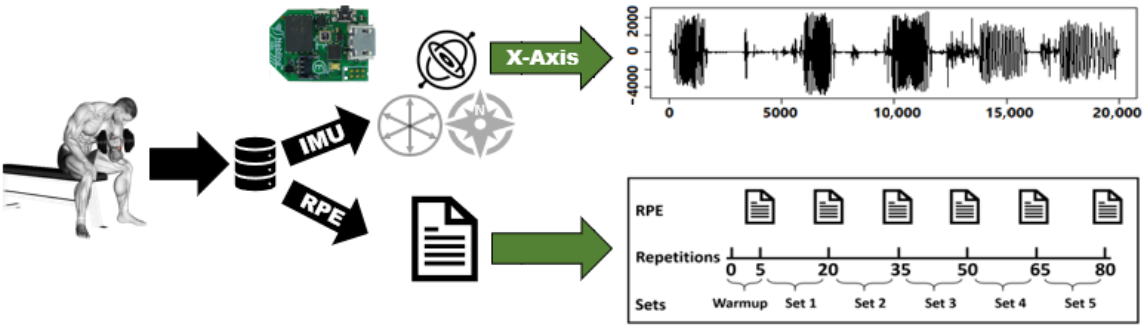


Figure 3.2: Visualization of data acquisition sessions of biceps concentration curl exercise. Rating of Perceived Exertion (RPE).

Figure 3.2 illustrates a data collection session for a concentration bicep curl exercise where each volunteer starts with five repetitions to warm-up, followed by 15 repetitions per set for a total of five sets. The volunteer reports their RPE values for each set, including the warm-up, yielding six RPE values. Then, we ask the volunteer to repeat the exercise using the other arm. We explain the exercise to each volunteer as the following:

(1) The volunteer should sit down on a flat bench with one dumbbell placed between their legs.

(2) The volunteer should be in a release position by using their right arm to pick the dumbbell up. Then, place the back of their right upper arm on the top of the inner

right thigh. The volunteer should rotate the palm of their hand until it is facing forward away from the thigh. Once, their arm is extended with the dumbbell above the floor then; the volunteer is in the correct release position.

(3) While the volunteer is holding the left arm stationary, they curls the weights forward while contracting the biceps as they breathe out. With the forearms movement only, the volunteer continues until their biceps are fully contracted, and the dumbbells are at shoulder level. The volunteer should hold the contracted position for a second.

(4) The volunteer should slowly begin to bring the dumbbells back to the release position as their breathe in. Avoid swinging motions at any time.

(5) Repeat for 15 repetitions. Then repeat the exercise the left arm to carry the dumbbell.

(6) The volunteers are equally allowed to rest for 2 min between sets.

It is essential to explain the rationale behind the tools used in the data collection sessions, such as the Borg's scale, the Apple Watch, and the 4.5 kg dumbbell. We use the Borg's scale in our work because we think that RPE is an appropriate marker of fatigue as previous studies within sport science have proven that RPE is capable of modeling a person's performance better in the real world compared to only heart rate monitoring (G. Borg, 1998; G. A. Borg, 1982; Crewe et al., 2008). We use the Apple Watch in our work as a way to strengthen the validity of the reported RPE by each volunteer. Borg's scale ranges from 6 to 20, where by multiplying these values by ten, we can estimate the volunteer's heart rate during the exercise. For example, if a volunteer reports 13 on Borg's scale, we should expect to measure their heart rate around 130 to 140 by the Apple Watch. However, in the rare cases of dissimilarity between the Borg scale and the measured heart rates, we average the reported RPE with the measured heart rate converted to RPE, as similarly performed in previous work (Yoo, Ackad, Heywood, & Kay, 2017). In total, there are 88 cases of dissimilarity, which count for 2.35% of our dataset. We use the 4.5 kg weight dumbbell in

our work because of three reasons. The first reason is that several previous works have used medium-weight dumbbells ranging between 3.5 kg and 5.5 kg to study muscular strength and fatigue (Bergquist, Iversen, Mork, & Fimland, 2018; Hwang, Chung, Song, Lim, & Kim, 2016; Liao et al., 2021). The second reason is that medium-weight dumbbells are often reported as the most commonly used dumbbells across gym-goers (Reis et al., 2017). The third reason is that the 4.5 kg weight dumbbell provides the best trade-off between number data points recorded in data sessions and time to reach fatigue during an exercise, as we explain in Section 3.3.

To summarize, each volunteer completed 5 warm-up repetitions followed by 15 repetitions per set for a total of 5 sets per hand, as shown in Figure 3.2. In total, our dataset consists of 3750 concentration bicep curl repetitions from 25 volunteers, where each repetition required approximately 2 seconds to complete. Given the fact that we use a 50 Hz IMU unit, a single repetition is captured in approximately 100 data samples. As a result, this allows us to distinctly capture signs of fatigue as muscle performance and exerted force decline overtime.

3.2 Dataset Processing

This section describes how we label our dataset through data processing as well as extracting fatigued and non-fatigued biceps repetitions from the collected data. Our collected data consists of an accelerometer, gyroscope, and magnetometer modules, where each module provides three-dimensional Cartesian coordinates (x, y, z). In addition, we have the RPE values reported by the volunteers for each bicep curls set. So far, our data contains a total of nine signals readings: The 3-axis of accelerometer, gyroscope, and magnetometer, along with their RPE values.

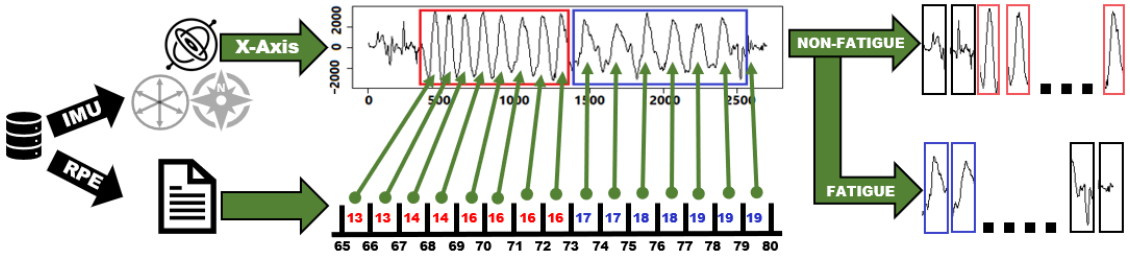
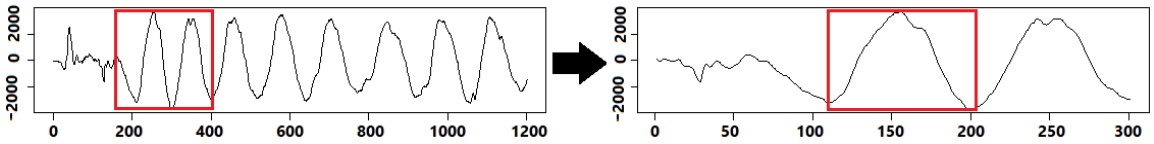
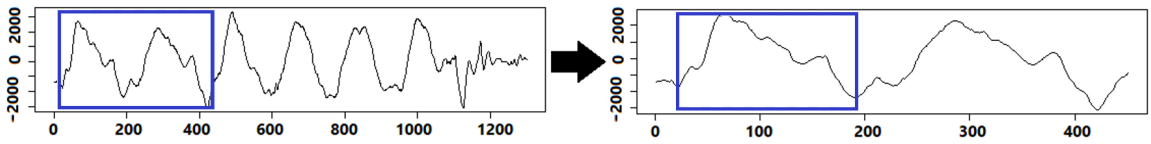


Figure 3.3: An example of extracting and labeling repetitions of the fifth set from the gyroscope's x-axis.

In Figure 3.3, we present an example of the fifth set of repetitions in its raw data form. In this example, the raw data is extracted from the gyroscope's x-axis along with their corresponding RPE values reported by the volunteer. A previous study on quantifying muscle fatigue suggests an RPE value of 16 as the threshold of true fatigue to estimate the declines in muscle strength during tasks (Whittaker, Sonne, & Potvin, 2019). Therefore, we extract and label each repetition manually according to the RPE values reported for the set, where we label repetitions with reported RPE values larger than 16 as fatigue and others as non-fatigue repetitions. The Figure shows two distinct groups of non-fatigue and fatigue repetitions extracted from the set. The non-fatigue repetitions are highlighted by the red border, while the blue border highlights the fatigue ones.



(a) Extracting a non-fatigue bicep curl repetition from the fifth set



(b) Extracting a fatigue bicep curl repetition from the fifth set.

Figure 3.4: A zoom-in at the bicep's repetitions data from the x-axis of the gyroscope.

Figure 3.4a shows a zoomed-in look on eight non-fatigue repetitions from the fifth

set, where we extract a complete non-fatigue repetition between two troughs. Similarly in Figure 3.4b, we extract a complete fatigue repetition between two troughs. However, we observe that non-fatigue repetitions are relatively symmetrical while the fatigue repetitions are often skewed positively, as shown in Figure 3.4. The troughs indicate full extension of the volunteer's arm at which the volunteer reaches the release position. In contrast, peaks indicate full flexion at which the volunteer reaches the contraction position.

It's important to mention that IT's modules record the data signal synchronously, which eases the data process task for each volunteer. In other words, we only need to fully extract and label all five sets recorded by one axis per volunteer, e.g., gyroscope's x-axis. Then, we use the same timestamps from the gyroscope's x-axis to extract and label repetitions for all the remaining signals (other axes of gyroscope, accelerometer, and magnetometer). In total, our dataset consists of 3,750 repetitions recorded from nine signals readings: The 3-axis of accelerometer, gyroscope, and magnetometer, along with their RPE values.

3.3 Dataset Challenges

This section discusses the applied solutions to two major challenges encountered during our data collection procedure. We presumed that selecting a certain dumbbell weight for collecting data from different volunteers may induce a loose variance in their fatigue measures, such as the number of repetitions or completion time per set. In addition, using subjective measures, such as RPE, we could introduce a dependency between the correctness of the selected Borg rating and the volunteers' body awareness.

3.3.1 Dumbbell Suitability

We believe that selecting a certain dumbbell weight for collecting data from different volunteers may cause volunteers to get exhausted quickly, hindering us from capturing

fatigue over time. This could cause a steady decline in muscle performance and its exerted force to plummet, which is less likely to occur naturally. To address this challenge, we provided all volunteers with three groups of dumbbells: Light-weight which includes 1.1 kg and 2.2 kg dumbbells, medium-weight which includes 4.5 kg dumbbells, and heavy-weight which includes 9 kg dumbbells, as shown in Table 3.2. Then, we asked each volunteer to perform at least 2 sets of bicep concentration curl repetitions until they felt fatigued. As expected, when volunteers used light-weight dumbbells, they were able to perform a high number of repetitions per set but fewer sets in total (see row 1 and 2 in Table 3.2). This resulted in long recording sessions with a lot of similar data entries until volunteers reached fatigue.

Table 3.2: The nominated dumbbells weights for the data collection process, the values reported are averages.

Weight (kg)	**Repetitions**	**Sets**	**Repetitions/Set**	**Completion Time**
1.1 kg	2960	60	42.0	2 Min, 30 S
2.2 kg	2417	79	31.0	1 Min, 45 S
4.5 kg	1580	100	16.0	1 Min, 17 S
9 kg	860	60	9.0	1 Min, 10 S

On the other hand, when we look at the results obtained with a heavy-weight dumbbell (9 kg), volunteers were able to accomplish fewer repetitions and fewer sets in total (see row 4 of Table 3.2). This resulted in short recording sessions with fewer data entries however, momentum changes were not captured clearly throughout the exercise because volunteers reached fatigue quickly. We found the results obtained with medium-weight dumbbells (4.5 kg) to be the best compromise between the recording time length and the momentum changes as volunteers reached fatigue more gradually. Volunteers were able to perform 16 repetitions per set, which each set taking on average 1 min and 17 s to complete (see row 3 in Table 3.2).

3.3.2 RPE's Subjectivity and Familiarity

We believe that using subjective measures, such as RPE, might introduce a dependency between the correctness of selected Borg rating and volunteers' awareness. Additionally, introducing the RPE to a volunteer for the first time may cause them to misevaluate their perceived exertion rate. To address both of these challenges, we apply a min-max normalization to the RPE value based on the current set to account for subjective differences in RPE. For illustration, we set the minimum value based on the RPE reported after the warm-up, which often ranges from 10 to 12. Then, we set the maximum value to the highest RPE on the Borg scale, which is 20. We use such a fixed value as the maximum RPE because if we use the values reported from the set, it will cause the current label to depend on future data, which is not methodologically sound. The longitudinal nature of the data acquisition sessions also helped participants to become more familiar with the scale as they performed more sets. Therefore their use of the RPE potentially evolved across consecutive sets.

3.4 Summary

This chapter illustrates how we collect and process our dataset. Also, it lists the possible solutions for the data challenges in our research. In the next chapter, we present our work on the impact of biceps muscle fatigue in wearable-based HAR systems. Also, we show that muscle fatigue may affect the bicep's movements, resulting in data patterns changes. Please note that we use our dataset in chapters 4 and 5 while it contains data processed from 20 volunteers. Later, we extend our dataset to reach 25 volunteers whom we utilize all of their data in chapter 6.

Chapter 4

On the Impact of Biceps Muscle Fatigue in Human Activity Recognition

This chapter introduces how we set up our experiment to evaluate the impact of fatigue in wearable-based HAR. Our main goal is to study the fatigue's direct and indirect impact on HAR systems, including the changes in collected data, extracted features, and models' performance during the presence of fatigue.

4.1 Introduction

Fatigue is an inevitable consequence when it comes to athletics and incremental exercises (Enoka & Duchateau, 2016). Nowadays, the literature floods with works on fatigue detection; however, there is no work on the impacts of fatigue in HAR systems. Instead, there are several works that study the impact of fatigue on the human body and performance. To study the impact of muscles fatigue in HAR systems, we have to examine the collected data, the extracted features, and the models' performance. Figure 4.1 shows an overview of the methodology used in this chapter.

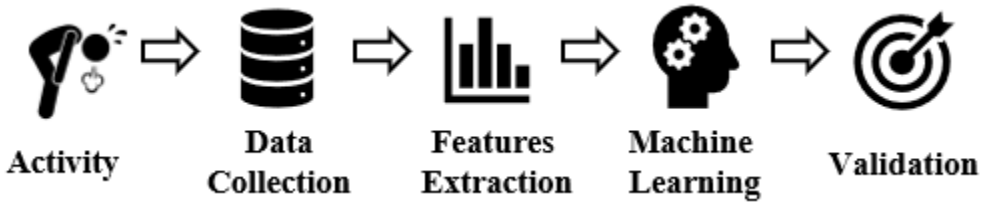


Figure 4.1: Overview of the wearable approach based human activity recognition system.

First, we use the collected biceps data from section 3.1 to extract a non-fatigue subset. We group all the non-fatigue repetitions labeled in section 3.2 into the subset. Then, we extract two sets of features: 1) a feature set from the complete dataset, 2) a feature set from the non-fatigue subset. After that, we compare the two sets of extracted features to find whether fatigue affects the number of significant features in detecting biceps curls. Next, we train five models to detect biceps concentration curls to observe the performance with and without the presence of fatigue. The five models are from comparative works which provide high-performance rates in detecting human activities using wearable IMU on the wrist (Kuhn et al., 2008; Min, Htay, & Oo, 2020; Moradi, Aghapour, & Shirbandi, 2019). The first model is the Generalized Linear Models (GLM) which has been adopted to analyze and count the number of walking steps in a previous study (Zhou, Ogihara, Nishimura, & Jin, 2017). The second model is the Logistic Regression (LR) which has been used to analyze and detect human activities (Alsheikh et al., 2016). The third model is Random Forest (RF), which has been used to detect and classify human actions using wearable motion sensor networks (Xu, Yang, Cao, & Li, 2017). The fourth model is the Decision Trees (DT) which has been used to count and classify ambulatory activities using eight plantar pressure sensors within smart shoes in a previous study (Jeong, Truong, & Choi, 2017). The fifth model is the Feedforward Neural Network (FNN) which has been used to detect and count repetitions for complex physical exercises (Soro, Brunner, Tanner, & Wattenhofer, 2019). We use the default settings of hyperparameter optimization for all the models. Furthermore, we consider two variations for each model: cross-subject and

subject-specific models. In cross-subject models, we train the models using data from different individuals while testing on unseen data from newcomers. While in subject-specific models, we train the models using data from an individual while testing on unseen data from the same individual.

Our main goal in this chapter is to study the impact of fatigue on the collected data, the number of significant features, and the models' performance. In addition, we evaluate the impact of fatigue on subject-specific and cross-subject models. Specifically, we address the following research questions:

- RQ1: How does fatigue impact the collected data?
- RQ2: What impact can fatigue impose on the extracted features?
- RQ3: What is the impact of fatigue on subject-specific biceps repetitions models?
- RQ4: What is the impact of fatigue on cross-subject biceps repetitions models?

Next, we detail the motivation, approach, and the findings for each research question. In section

4.2 The Impact of Fatigue on the Collected Data

In this section, we investigate and examine the data changes in the presence of fatigue to answer RQ1: How does fatigue impact the collected data? **Motivation:** We hypothesize that fatigue impacts the collected data by changing its patterns, leading to a snowball effect, affecting the extracted features and HAR models' performance. Hence, we want first to capture the data pattern changes, which might occur during the data collection process.

Approach: To fulfill our motivation, first, we quantify the fatigue in our data and then capture the data pattern changes. We calculate the average of fatigue repetition share per volunteer, as shown in Figure 4.2. The volunteers did not report any fatigue repetitions at

the warmup set thus, the share of fatigue for this set is 0%. However, in the first set, 19 volunteers reported the last repetition as fatigue, which represents 1 out of 15 repetitions (6.6%), while 1 volunteer reported no fatigue (0%). Hence, the average share of fatigue repetitions in the first set was reported at 6.2%. As we see in Figure 4.2, the average share of fatigue repetitions increases at each subsequent set as volunteers become more progressively tired, reaching an average of 56.0% of fatigue repetitions per volunteer in the fifth set.

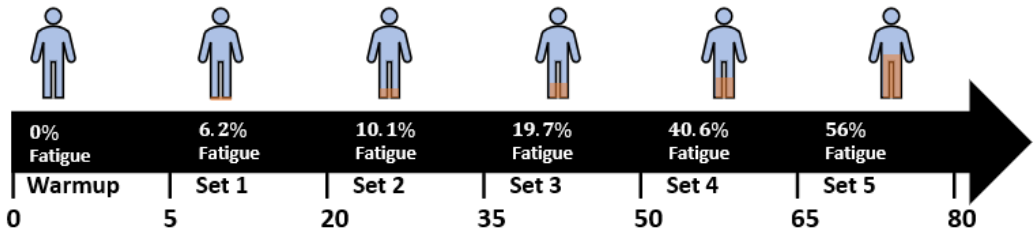


Figure 4.2: A visualization of the weighted average of fatigue shares per exercise sets in the collected data.

For capturing data pattern changes, we start with looking into the data provided by the IMU that contains the 3-axis gyroscope and accelerometer. We excluded the magnetometer for simplicity as it did not show any significant changes in the magnetic field regarding direction or strength during the exercise. We started with a visualization of the impact of fatigue on collected data to evaluate the data pattern changes. Figure 4.3 shows an example of the five sets of biceps repetitions using the gyroscope and the accelerometer signals. The X-axis represents the vertical displacement, which is the distance between the highest and lowest positions of the volunteer's hand during bicep extension and flexion. The Y-axis represents the horizontal displacement, which is the sideways vibration of the volunteer's hand during bicep extension and flexion. The Z-axis represents the depth displacement, which is the farthest and nearest positions of the volunteer's hand from their body during bicep extension and flexion. We select the X-axis from the gyroscope and Y-axis from the accelerometer because they provide the best visualization for the angular velocity and

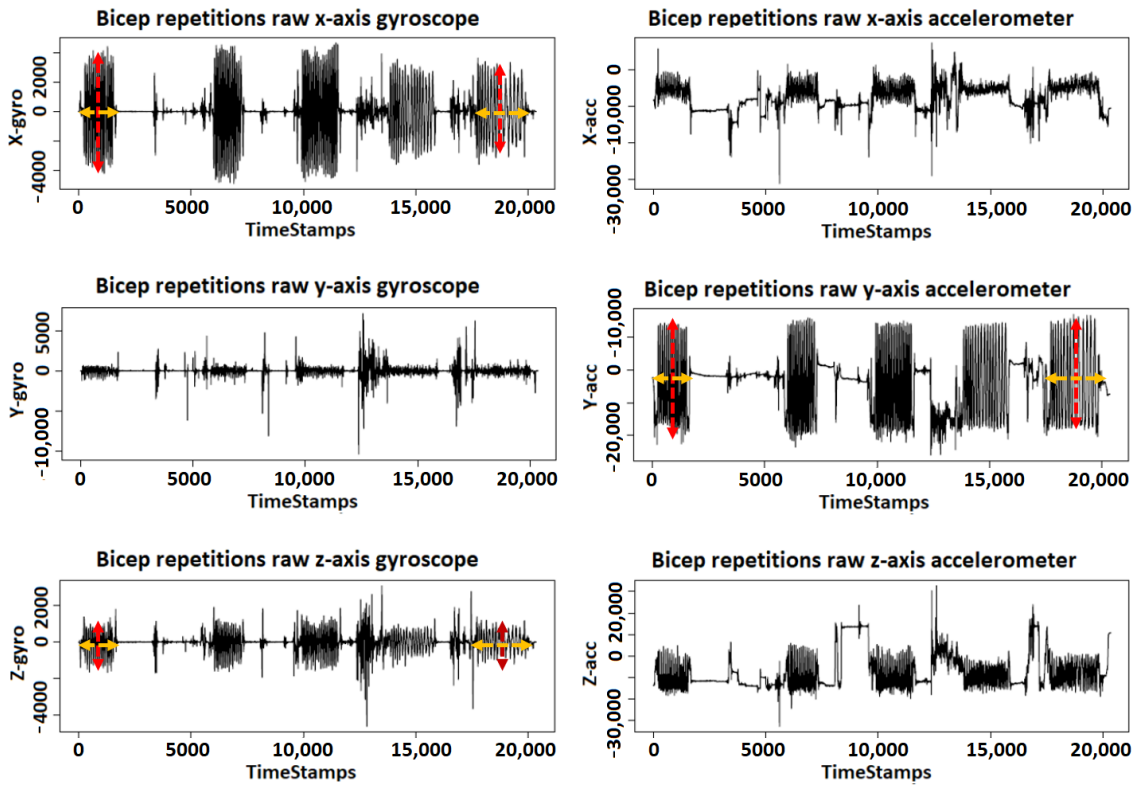


Figure 4.3: A visualization of the impact of fatigue on collected data.

sideways vibration of the volunteer's hand. Similarly, we showcase the Z-axis from the gyroscope as it provides the best visualization of the farthest and nearest positions of the volunteer's hand. We use the first set of bicep repetition as a reference set to comparatively measure the data pattern changes. The rationale behind this is that the first set usually contains the least number of fatigue repetitions. Therefore, as the fatigue accumulates in later sets, we would be able to differentiate the changes in the data patterns. The first set also always contains 15 biceps repetitions for all of the 20 volunteers.

We look for data pattern changes along the horizontal axis which indicates the changes in completion time whereas the vertical axis indicates the changes in angular velocity according to muscular endurance (Lee, Eun, Kim, Park, & Jee, 2017). To analyze the data pattern changes in the horizontal axis, we measured the time required to complete the first set of biceps repetitions for each volunteer, which is the time interval from the 1st repetition

until the end of the 15th repetition. We repeated the same approach to measure the completion time of the remaining sets separately. Then, we calculate the difference in completion time between each set compared to the first set. To analyze the data pattern changes in the vertical axis, we measured the absolute magnitude of each repetition in the first set to calculate the muscular endurance (Lee et al., 2017) during the first set of biceps repetitions. We repeated the same approach to measure the muscular endurance for the remaining sets separately. Then, we calculate the drop in muscular endurance between each set compared to the first set.

Findings: Table 4.1 shows the increase in completion time for each set in relation to the first set. In the 2nd set, the average increase in completion time is 1.7%, which is considerably a small change to the 1st set. The reason is that the 2nd set is usually the introductory stage of fatigue, where fatigue occurs for the first time at the last 1 or 2 repetitions. When we look at the 3rd set, we found the average increase in completion time to have increased to 8.1%. At the 4th set, volunteers take on average 14.3% more time to finish their exercises, compared to the time they took in the 1st set. Comparing the 4th set to the 3rd set, the 4th set contains almost twice the number of fatigue repetitions than the 3rd set, resulting on substantial increase in the time to complete metric, from 8% to 14%. At the 5th set, we found that the average increase in completion time is 31.0%, more than twice the increase observed in the 4th set. The reason is that the 5th set contains at least eight fatigue repetitions, which indicates that fatigue impacts later sets to a much larger extent, slowing down bicep movements and increasing the time completion for the set. As a result, fixed-size and non-overlapping windows will no longer be suitable to capture full repetitions because of its narrow fit, especially, at the 4th and 5th sets.

In Table 4.2, we present the changes of muscular endurance for each of the five sets as the fatigue accumulates during repetitions in the later sets. It is possible to measure the muscular endurance using the angular velocity from the gyroscope (Lee et al., 2017) rather

Table 4.1: The increase in the time to complete a set compared to the 1st set.

Axis-Sensor	**2nd Set**	**3rd Set**	**4th Set**	**5th Set**	**Avg.**
X-Gyroscope	+2.0%	+6.0%	+17.0%	+33.0%	+14.5%
Z-Gyroscope	+1.7%	+11.0%	+15.0%	+45.0%	+18.2%
Y-Accelerometer	+1.5%	+7.4%	+11.0%	+15.0%	+8.7%
Avg./set	+1.7%	+8.1%	+14.3%	+31.0%	

Table 4.2: The change in muscular endurance represented in vertical shrinks, compared to the 1st set.

Axis-Sensor	**2nd Set**	**3rd Set**	**4th Set**	**5th Set**	**Avg.**
X-Gyroscope	+0.7%	+1.2%	−6.3%	−5.2%	−2.4%
Z-Gyroscope	+0.5%	+1.7%	−10.4%	−7.5%	−3.9%
Y-Accelerometer	+0.6%	+0.3%	+0.3%	+0.4%	0.4%
Avg./set	+0.6%	+1.1%	−5.5%	−4.1%	

than using the accelerometer. Therefore, we observe that fatigue decreases the muscular endurance according to the X- and Z-axes from the gyroscope by an average of -2.4% and -3.9%, respectively. However, we do not observe a substantial decrease on the muscular endurance using the Y-axis from the accelerometer, with a small average change of only 0.4%. Overall, we observe that the average muscular endurance decreases in the later sets as the fatigue accumulates in the repetitions. For example, the 2nd and 3rd sets maintain a muscular endurance similar to the compared 1st set. However, the muscular endurance decreases by an average of 5.5% in the 4th set, and 4.1% in the 5th set, as fatigue accumulates over time. This could negatively impact data filtering, especially in the case of peak filtering, because such a filter may exclude a complete bicep repetition if it did not meet the peak threshold, especially, at the 4th and 5th sets.

4.3 The Impact of Fatigue on the Extracted Features

In this section, we investigate the features' significance in the presence of fatigue to answer RQ2: What impact can fatigue impose on the extracted features? **Motivation:** We hypothesize that if fatigue affects the collected data, it may affect the extracted features from the same data. In other words, some features may appear to be significant in detecting biceps repetitions without fatigue, but become less significant at later sets, where fatigue often occurs. We think that a factor such as fatigue can deform the patterns of these features reducing their correlation hence, some extracted features may be more sensitive to fatigue than others.

Approach: To fulfill our motivation, we extract three main features: mean, mean absolute deviation (MAD), and standard deviation (SD). We split our extracted features into two groups: extracted features from the complete dataset and extracted features from the non-fatigue subset. Our goal is to investigate whether fatigue can affect the significance of the extracted features in HAR models. Our complete dataset and the non-fatigue subset contain 12 data signals. There are 9 data signals from the 3-axes of gyroscope, accelerometer, and magnetometer, and there are three representative data signals for the rotations on X-, Y-, and Z-axes which are roll, pitch, and yaw. We use Spearman's rank correlation coefficient with a significance allowance of 0.1 to show how these extracted features correlate with bicep repetitions (Hauke & Kossowski, 2011). Previous works show that these features can improve the detection of human activities and achieve better performance overall (Ferrari, Micucci, Mobilio, & Napoletano, 2020; Janidarmian, Roshan Fekr, Radecka, & Zilic, 2017; Op De Beéck et al., 2018). In addition, these features are fit to capture patterns in data changes and are often immune against data anomalies or disturbance, especially when it comes to periodic or repetitive activities such as bicep curls (Javaid, Rashid, Tiwana, & Anwar, 2018).

Table 4.3: Table of the significant (✓) and insignificant (×) features extracted from both none-fatigue subset and complete dataset; the changed features are in highlighted bold.

			Non-Fatigue subset			**Complete dataset**		
			Mean	**MAD**	**SD**	**Mean**	**MAD**	**SD**
Sensors and Axes	**Acc.**	X-axis	**✓**	**✓**	**✓**	**×**	**×**	**×**
		Y-axis	✓	✓	✓	✓	✓	✓
		Z-axis	✓	✓	✓	✓	✓	✓
	Mag.	X-axis	×	×	×	×	×	×
		Y-axis	×	**✓**	×	×	**×**	×
		Z-axis	×	×	✓	×	×	✓
	Gyro.	X-axis	**✓**	✓	**✓**	**×**	✓	**×**
		Y-axis	✓	✓	✓	✓	✓	✓
		Z-axis	✓	✓	✓	✓	✓	✓
		Roll	✓	✓	✓	✓	✓	✓
		Pitch	✓	✓	✓	✓	✓	✓
		Yaw	✓	**✓**	✓	✓	**×**	✓

Table 4.3 shows the features extracted from each datasets. In section 4.2, we show that the X-axis represents the vertical displacement, which has the largest angle of movement and linear acceleration. However, fatigue often affects acceleration greatly compared to the angle of movement due to the movement nature of the biceps muscle (Ghazal, Alhalabi, Fraiwan, Yaghi, & Alkhatib, 2019). Therefore, changes on the linear acceleration, measured by the accelerometer, affected the significance of its extract features. On the other hand, the angular velocity, measured by the gyroscope, remains relatively steady because of the fixed angle of movement of bicep muscles. Regarding the low number of significant features from the magnetometer, we think that fatigue does not impact these features.

Besides, the magnetometer is not an optimal sensor in fatigue detection as it does not reveal any significant characteristics in data readings about the magnetic field's direction or strength, as shown section 4.2. Therefore, most of the features extracted from the magnetometer are insignificant.

Findings: Our findings show that fatigue significantly impacts the extracted features, by hindering their correlation coefficient values to the extent of turning some significant features into insignificant ones. We were able to extract 9 mean, 10 MAD, and 10 SD features from the none-fatigue subset for a total of 29 significant features with a significance allowance of 0.1. However, once the fatigue was introduced in the data, we were only able to extract 7 mean, 7 MAD, and 8 SD features for a total of 22 significant features, as shown in Table 4.3. This indicates that fatigue, once introduced in the dataset, reduced the significance of 7 features (24% of total significant features).

4.4 The Impact of Fatigue on Subject-Specific Biceps Repetitions Models

In this section, we examine the subject-specific models' performance in the presence of fatigue to answer RQ3: What is the impact of fatigue on subject-specific biceps repetitions models? **Motivation:** After observing the impact of fatigue on the collected data and the extracted features, we investigate how it may affect the performance of biceps detection models, especially the subject-specific models. We hypothesize that the presence of fatigue may decrease the model's performance in recognizing human activities; therefore, we examine their performance in detecting biceps curls while fatigue progressively pervades in the dataset.

Approach: To fulfill our motivation, we use the five detection models in section 4.1. These models use the 22 significant features extracted from our complete dataset to eliminate weak features that turn to insignificant once fatigue occurs. We use these models to detect biceps repetitions in our dataset. Then, we calculate the accuracy using the confusion matrix shown in Table 4.4, where non-repetition represents an incomplete repetition or any random movement, and repetition represents a completed repetition, whether it contains fatigue or not.

Table 4.4: Confusion matrix for biceps repetitions

		Actual	
		Repetition	**Non-Repetition**
Detect	**Repetition**	TRUE Repeat	FALSE Repeat
	Non-Repetition	FALSE Non-Repeat	TRUE Non-Repeat

We perform six 10-fold cross-validation runs using the non-fatigue subset where we replace 10% of the non-fatigue repetitions in the subset with fatigue repetitions from our complete dataset per experiment. This allows us to gradually observe fatigue effects on the detection models while fatigue propagates in the dataset. Naturally, fatigue should exist in both training and testing datasets since it is a byproduct of physical activities. Therefore, we train and test all five models under fatigue levels similar to Figure 4.2 approximately. Also, we use the first run in the 10-fold cross-validation as a reference point because it contains no fatigue repetitions to hinder models' performance. Then, we calculate the accuracy (1), precision (2), recall (3), and F1 (4) per run. Table 4.5 shows the performance averages for the six 10-fold cross-validation runs over all participants per model. Each $\triangle$ F1* row shows the difference in the model's performance, compared to the performance of the first run.

$$Accuracy = \frac{True(Repeat + NonRepeat)}{True(Repeat + NonRepeat) + False(Repeat + NonRepeat)} \quad (1)$$

$$Precision = \frac{True(Repeat)}{True(Repeat) + False(Repeat)} \quad (2)$$

$$Recall = \frac{True(Repeat)}{True(Repeat) + False(NonRepeat)} \quad (3)$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4)$$

Findings: Our findings show that the more fatigue in the dataset, the steeper the five models' performance decline. In fact, Table 4.5 shows that replacing as little as 10% of the repetitions with fatigue repetitions can drop the GLM models' performance by 8%, and 11% for DT. If we replace an additional 10% of the repetitions with fatigue repetitions, all models' performance decrease by at least 21% (FNN and DT). The decrease in the performance can be as severe as 30% in the RF model. Such findings indicate that, for some models (e.g., GLM, LR, and RF), it only takes 20% of fatigued repetitions to decrease a model's performance by more than 20%. The impact in the model's performance is even more significant when we reach to 40% and 50% of fatigue repetitions. With half the repetitions containing fatigue, the models lose between 47% (DT) to 57% (GLM) of its original performance, which may compromise the reliability of HAR systems that do not take fatigue properly into account.

Table 4.5: The performance averages for subject-specific models to detect biceps repetitions throughout the incremental replacement of fatigue repetitions.

			% of fatigue repetitions in dataset					
			0%*	**10%**	**20%**	**30%**	**40%**	**50%**
Models	**GLM**	Precision	0.94	0.89	0.78	0.71	0.65	0.44
		Recall	0.91	0.81	0.61	0.60	0.52	0.37
		Accuracy	0.9	0.84	0.80	0.76	0.63	0.45
		F1	0.92	0.85	0.68	0.67	0.58	0.40
		%△ **F1***	**–**	**-8%**	**-26%**	**-30%**	**-38%**	**-57%**
	LR	Precision	0.9	0.85	0.71	0.63	0.49	0.40
		Recall	0.81	0.73	0.53	0.44	0.55	0.36
		Accuracy	0.88	0.83	0.76	0.62	0.57	0.49
		F1	0.85	0.79	0.61	0.52	0.52	0.38
		%△ **F1***	**–**	**-9%**	**-29%**	**-40%**	**-41%**	**-56%**
	RF	Precision	0.88	0.82	0.68	0.6	0.56	0.45
		Recall	0.78	0.7	0.52	0.5	0.43	0.39
		Accuracy	0.85	0.81	0.75	0.49	0.29	0.19
		F1	0.83	0.76	0.59	0.55	0.49	0.42
		%△ **F1***	**–**	**-10%**	**-30%**	**-34%**	**-42%**	**-50%**
	DT	Precision	0.86	0.75	0.66	0.57	0.46	0.44
		Recall	0.7	0.64	0.57	0.43	0.4	0.4
		Accuracy	0.81	0.77	0.73	0.61	0.46	0.24
		F1	0.77	0.69	0.61	0.49	0.43	0.42
		%△ **F1***	**–**	**-11%**	**-21%**	**-36%**	**-45%**	**-47%**
	FNN	Precision	0.98	0.89	0.76	0.68	0.58	0.5
		Recall	0.91	0.79	0.71	0.65	0.6	0.48
		Accuracy	0.99	0.91	0.86	0.8	0.73	0.67
		F1	0.94	0.84	0.73	0.66	0.59	0.49
		%△ **F1***	**–**	**-10%**	**-21%**	**-31%**	**-38%**	**-49%**

* Reference to the first set which does not include fatigue.

4.5 The Impact of Fatigue on Cross-Subject Biceps Repetitions Models

In this section, we examine the cross-subject models' performance in the presence of fatigue to answer RQ4: What is the impact of fatigue on cross-subject biceps repetitions models? **Motivation:** After observing the impact of fatigue on the subject-specific models, we investigate the generality of the models and the ability to detect biceps repetitions across different subjects. Also, we expect fatigue to impact the performance of the cross-subject models to a greater extent since these models tend to perform less than the subject-specific ones.

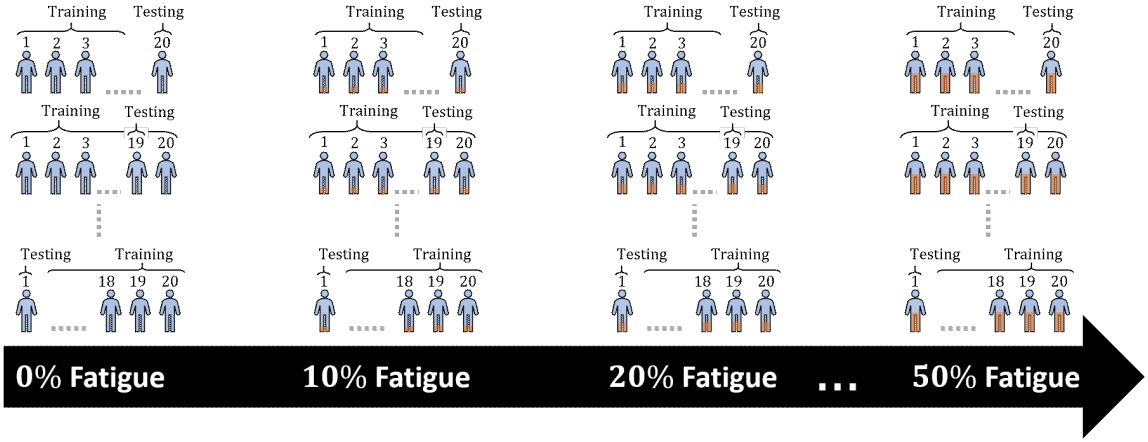


Figure 4.4: A partial representation of the six LOOCVs with K = 20 and different percentages of fatigue.

Approach: Similar to our approach in section 4.4, we use the five detection models in section 4.1 with the 22 significant features extracted from our complete dataset. We perform six Leave-One-Out Cross-Validations (LOOCV) runs where K equals the number of volunteers, K = 20. Figure 4.4 shows a partial representation of the six LOOCV runs with K = 20 for using the data of 20 volunteers individually. Each LOOCV run consists of 20 iterations, in the 1st iteration, we use the data from 19 volunteers to train our models then use the 20th volunteer's data to test the models. At last, in the 20th iteration, we

should have used all volunteers data for testing except the 1st volunteer therefore, we train the model using all the 19 volunteers dataset then, used the 1st volunteer's data for testing. We calculate the precision, recall, accuracy, and F1-score per iteration then report the averages. We repeat the LOOCV run after we replace 10% of each volunteer's data with fatigue repetitions from each individuals data, recursively. We use the first LOOCV run as a reference point because there are no fatigue repetitions in the individuals data to affect the models' performance. We train and test all five models under fatigue levels similar to our approach in section 4.4. Then, we calculate the accuracy (1), precision (2), recall (3), and F1 (4) for each run. Table 4.6 shows the performance averages for the six LOOCV runs per model. Each $\triangle$ F1* rows show the comparison of the model's performance against the performance obtained in the first run (no fatigue repetitions).

To further examine whether our approach adequately accounts for the impact of fatigue, we repeat RQ3 and RQ4 using all features. We use all the 36 extracted features, including the seven features that became insignificant due to the presence of fatigue. This allows us to observe the performance of detection models using all features versus models using only fatigue-resistant features. According to Table 4.7 and 4.8, fatigue impacts a model's performance to an even greater extent compared to the models based on 22 fatigue-resistant features presented in RQ3 and RQ4.

Findings: Our finding indicates fatigue impacts the performance in all five models significantly. Table 4.6 shows that replacing as little as 10% of the repetitions with fatigued ones can drop a model's performance by 6% for RF, and down to 13% for GLM and LR. If we replace an additional 10% of the repetitions with fatigued ones, the models' performance decrease by 20% for FNN and DT, and down to 25% for LR. Once the fatigue reaches 30% of repetitions, we see a sharp decrease in all models by at least 30%. This trend continues, as once the fatigue repetitions reach 50% of the dataset, the HAR models' performance decrease by at least 41%. We observe a negative linear effect in some

Table 4.6: The performance averages for cross-subject models to detect biceps repetitions throughout the incremental replacement of fatigue repetitions.

			% of fatigue repetitions in individuals data					
			0%*	**10%**	**20%**	**30%**	**40%**	**50%**
Models	**GLM**	Precision	0.85	0.73	0.60	0.52	0.45	0.41
		Recall	0.8	0.71	0.72	0.50	0.32	0.31
		Accuracy	0.87	0.71	0.66	0.52	0.41	0.33
		F1	0.82	0.72	0.65	0.51	0.37	0.35
		%△ **F1***	**–**	**-13%**	**-21%**	**-38%**	**-55%**	**-57%**
	LR	Precision	0.87	0.79	0.66	0.53	0.45	0.40
		Recall	0.78	0.65	0.58	0.41	0.36	0.30
		Accuracy	0.82	0.75	0.66	0.52	0.43	0.29
		F1	0.82	0.71	0.62	0.46	0.40	0.34
		%△ **F1***	**–**	**-13%**	**-25%**	**-44%**	**-51%**	**-58%**
	RF	Precision	0.78	0.71	0.64	0.58	0.5	0.46
		Recall	0.67	0.65	0.51	0.45	0.43	0.39
		Accuracy	0.79	0.73	0.58	0.43	0.39	0.21
		F1	0.72	0.68	0.57	0.51	0.46	0.42
		%△ **F1***	**–**	**-6%**	**-21%**	**-30%**	**-36%**	**-41%**
	DT	Precision	0.78	0.73	0.64	0.53	0.45	0.44
		Recall	0.71	0.63	0.55	0.41	0.41	0.32
		Accuracy	0.81	0.76	0.53	0.42	0.33	0.18
		F1	0.74	0.68	0.59	0.46	0.43	0.37
		%△ **F1***	**–**	**-9%**	**-20%**	**-38%**	**-42%**	**-50%**
	FNN	Precision	0.9	0.81	0.73	0.66	0.58	0.57
		Recall	0.84	0.75	0.66	0.55	0.48	0.46
		Accuracy	0.95	0.87	0.81	0.74	0.53	0.49
		F1	0.87	0.78	0.69	0.60	0.53	0.51
		%△ **F1***	**–**	**-10%**	**-20%**	**-31%**	**-40%**	**-41%**

* Reference to the first set which does not include fatigue.

Table 4.7: The performance averages for subject-specific models to detect biceps repetitions, using the 36 features, throughout the incremental replacement of fatigue repetitions.

			% of Fatigue Repetitions in Dataset					
			0% *	**10%**	**20%**	**30%**	**40%**	**50%**
Models	**GLM**	Precision	0.96	0.83	0.56	0.49	0.42	0.38
		Recall	0.91	0.81	0.67	0.47	0.30	0.29
		Accuracy	0.99	0.81	0.62	0.49	0.38	0.31
		$F1_{36}$	0.94	0.82	0.61	0.48	0.35	0.33
		$\%\triangle F1_{36}$*	–	-13%	-35%	-49%	-63%	-65%
	LR	Precision	0.95	0.86	0.60	0.47	0.37	0.33
		Recall	0.85	0.71	0.53	0.37	0.29	0.24
		Accuracy	0.89	0.82	0.60	0.46	0.35	0.24
		$F1_{36}$	0.90	0.78	0.56	0.41	0.33	0.28
		$\%\triangle F1_{36}$*	–	-13%	-37%	-54%	-64%	-69%
	RF	Precision	0.94	0.65	0.59	0.50	0.42	0.38
		Recall	0.80	0.60	0.47	0.39	0.36	0.32
		Accuracy	0.95	0.67	0.54	0.37	0.32	0.17
		$F1_{36}$	0.86	0.62	0.52	0.44	0.38	0.35
		$\%\triangle F1_{36}$*	–	-28%	-39%	-49%	-56%	-59%
	DT	Precision	0.90	0.78	0.58	0.42	0.38	0.37
		Recall	0.82	0.67	0.50	0.32	0.35	0.27
		Accuracy	0.93	0.81	0.48	0.33	0.28	0.15
		$F1_{36}$	0.86	0.72	0.54	0.37	0.36	0.31
		$\%\triangle F1_{36}$*	–	-15%	-37%	-57%	-57%	-63%
	FNN	Precision	0.99	0.97	0.63	0.57	0.50	0.49
		Recall	0.93	0.90	0.57	0.47	0.41	0.40
		Accuracy	0.99	0.96	0.70	0.64	0.46	0.42
		$F1_{36}$	0.96	0.93	0.60	0.52	0.45	0.44
		$\%\triangle F1_{36}$*	–	-3%	-38%	-46%	-53%	-54%

* Reference to the first set which does not include fatigue.

Table 4.8: The performance averages for cross-subject models to detect biceps repetitions, using the 36 features, throughout the incremental replacement of fatigue repetitions.

			% of Fatigue Repetitions in Dataset					
			0% *	**10%**	**20%**	**30%**	**40%**	**50%**
Models	**GLM**	Precision	0.92	0.80	0.63	0.58	0.53	0.36
		Recall	0.89	0.73	0.49	0.49	0.42	0.30
		Accuracy	0.88	0.76	0.65	0.62	0.51	0.36
		$F1_{36}$	0.91	0.76	0.56	0.53	0.47	0.33
		$\%\triangle F1_{36}$*	–	-16%	-39%	-42%	-48%	-64%
	LR	Precision	0.82	0.76	0.63	0.56	0.40	0.33
		Recall	0.74	0.65	0.47	0.39	0.45	0.29
		Accuracy	0.80	0.74	0.68	0.55	0.46	0.40
		$F1_{36}$	0.78	0.70	0.54	0.46	0.42	0.31
		$\%\triangle F1_{36}$*	–	-10%	-30%	-40%	-46%	-60%
	RF	Precision	0.85	0.77	0.59	0.52	0.46	0.37
		Recall	0.76	0.66	0.45	0.44	0.36	0.32
		Accuracy	0.82	0.76	0.65	0.43	0.24	0.16
		$F1_{36}$	0.80	0.71	0.51	0.47	0.40	0.35
		$\%\triangle F1_{36}$*	–	-11%	-36%	-41%	-50%	-57%
	DT	Precision	0.91	0.80	0.59	0.45	0.39	0.37
		Recall	0.74	0.69	0.51	0.34	0.34	0.34
		Accuracy	0.85	0.82	0.65	0.48	0.39	0.20
		$F1_{36}$	0.81	0.74	0.55	0.39	0.36	0.36
		$\%\triangle F1_{36}$*	–	-9%	-33%	-52%	-55%	-56%
	FNN	Precision	0.99	0.89	0.65	0.58	0.50	0.43
		Recall	0.92	0.79	0.61	0.56	0.52	0.41
		Accuracy	0.99	0.91	0.74	0.69	0.63	0.58
		$F1_{36}$	0.96	0.83	0.63	0.57	0.51	0.42
		$\%\triangle F1_{36}$*	–	-13%	-34%	-40%	-47%	-56%

* Reference to the first set which does not include fatigue.

models' performance as the fatigue increases. For instance, the performance of DT models decreases by an average of 10% for every 10% increase of fatigue in the dataset.

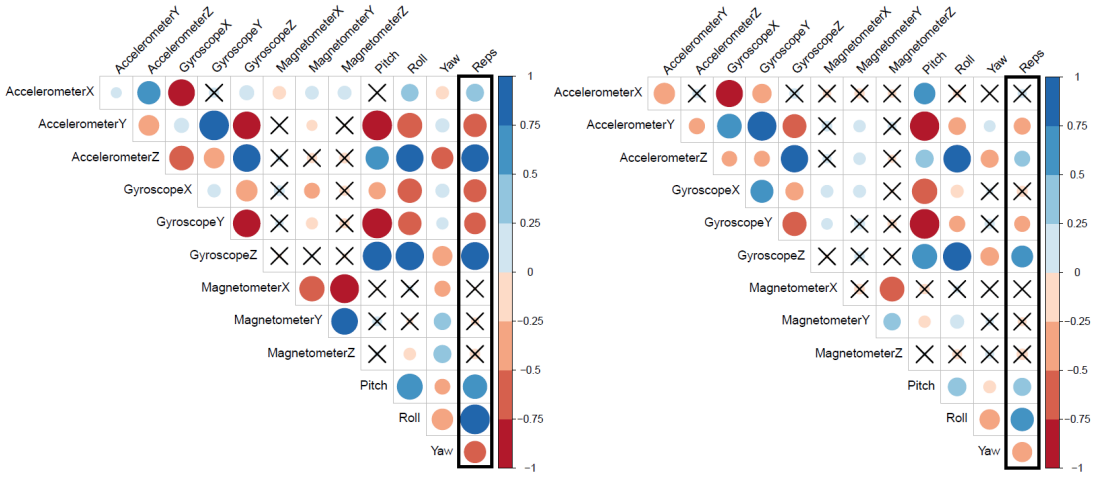
The results corroborate with our previous analyses, showing that the extraneous features are unlikely to contribute in detecting biceps concentration curls when fatigue is present. It is important to mention that models using all features do outperform the fatigue-resistant models when the presence of fatigue is very low in the dataset (no fatigue or fatigue data at 10%).

4.6 Discussion

In this section, we discuss the findings from our four research questions. In RQ1, our finding shows that fatigue can lead to changes in data patterns over time. A similar finding to ours is shown in a previous work, which suggests a decrease in the mean power frequency of the accelerometer readings trend with increasing biceps muscle fatigue (Mokaya, Lucas, Noh, & Zhang, 2016). Also, the work suggests that accelerometers should be used to sense skeletal muscle vibrations, which can reduce the error of estimating fatigue up to 50%. Therefore, we adopt a similar approach in RQ1, where we use a time series dataset collected using an inertial sensor that includes accelerometers to observe data pattern changes along the horizontal and vertical axes. In other words, this allows us to find a correlation between fatigue and data pattern changes that occur horizontally related to completion time and vertically associated with the muscular endurance and angular velocity.

In RQ2, we investigate the impact of fatigue in feature extraction. A previous work shows that muscle fatigue affects the collected biceps data signals from Electromyography (EMG) sensor by increasing the Root Mean Square Error (RMSEs) of the extracted features (Triwiyanto, Wahyunggoro, Nugroho, & Herianto, 2018). Similarly, our findings show that fatigue can hinder the correlation values of some of the extracted features to the extent of turning them into insignificant features. However, if we look at this problem from

another perspective, we can label the extracted features as fatigue-resistant. Meaning, although fatigue existed in the dataset, these features remain significant. As a result, we can develop a group of fatigue-resistant features that can counter data pattern changes due to fatigue and remain valuable to detect bicep activity such as biceps concentration curls. It is important to mention, however, that these features are still affected by fatigue as their correlation coefficient values drop.

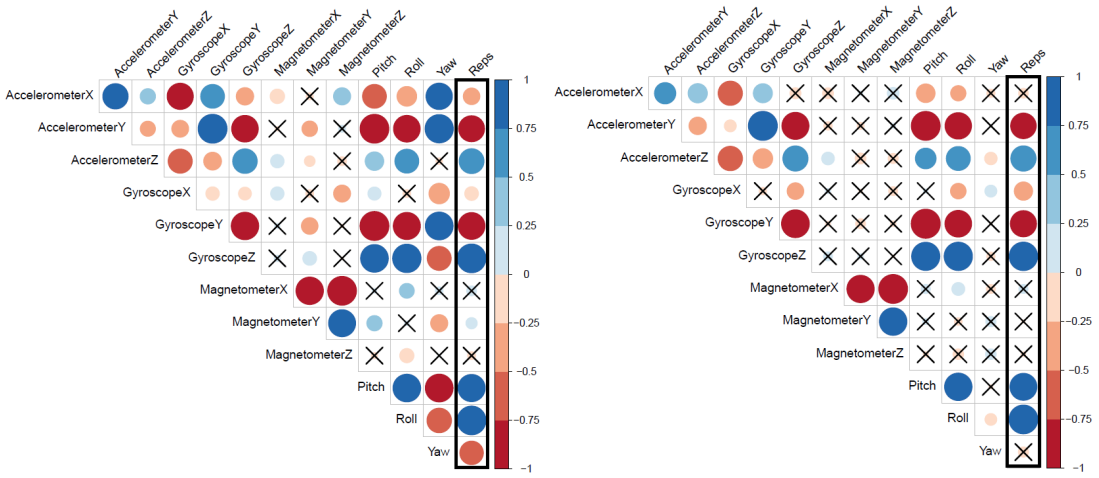


(a) (b)

Figure 4.5: Graphical display of the differences in the correlation matrices of the 12 extracted features (mean) and the bicep repetitions with and without fatigue. (**a**) Correlation matrix of the 12 mean features and the bicep repetitions in the non-fatigue subset. (**b**) Correlation matrix of the 12 mean features and the bicep repetitions in our complete dataset.

Figure 4.5 presents the correlation matrix for 12 mean features and the bicep repetitions. The positive correlations are displayed in blue and negative correlations are presented in red. The color intensity and the size of the circle are also proportional to the correlation coefficients whereas, the insignificant correlations are marked with $\times$. Figure 4.5a shows that bicep repetitions have significant correlations with 9 out of 12 mean features extracted from the non-fatigue subset. These 9 significant features are (X,Y,Z)-Accelerometer, (X,Y,Z)-Gyroscope, pitch, roll, and yaw. On the other hand, Figure 4.5b shows that bicep repetitions have significant correlations with 7 out of 12 mean features extracted from our

complete dataset where fatigue exists during the exercise. These 7 significant features are (Y,Z)-accelerometer, (Y,Z)-gyroscope, pitch, roll, and yaw. We can observe two impacts of fatigue on extracted features. First, some mean features correlations became insignificant to bicep repetitions such as X-accelerometer and X-gyroscope. Second, an overall drop in the correlation coefficient values for all mean features, as presented by the faint color intensity and the shrink of circle sizes.

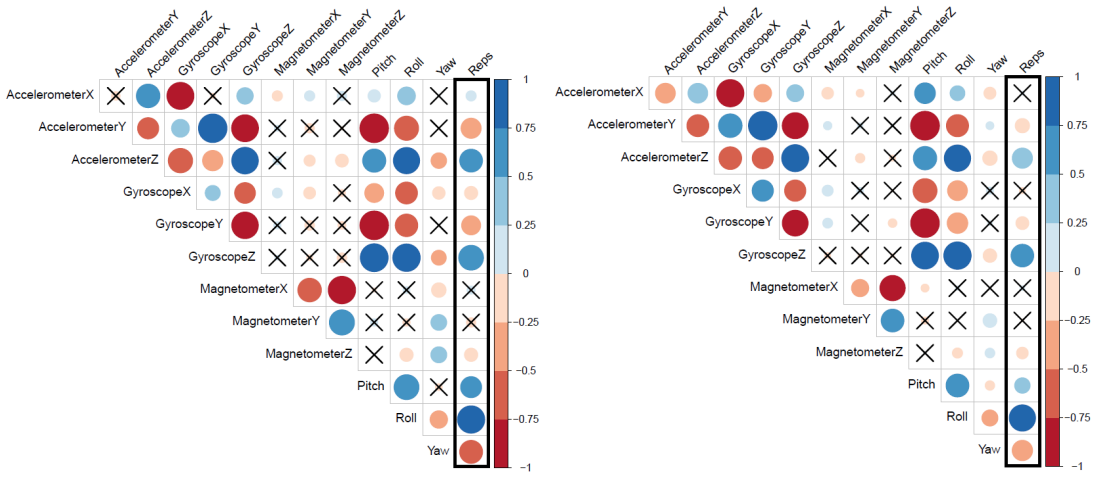


(a) (b)

Figure 4.6: Graphical display of the differences in the correlation matrices of the 12 extracted features (MAD: Mean Absolute Deviation) and the bicep repetitions with and without fatigue. (**a**) Correlation matrix of the 12 MAD features and the bicep repetitions in the non-fatigue subset. (**b**) Correlation matrix of the 12 MAD features and the bicep repetitions in our complete dataset.

To strengthen the evidence that points to fatigue as the potential cause of these impacts, we believe that similar observations should exist for MAD and SD features. Figure 4.6 presents the correlation matrix for 12 MAD features and the bicep repetitions in our complete dataset and the non-fatigue subset. Figure 4.6a shows that bicep repetitions have significant correlations with 10 out of 12 MAD features extracted from the non-fatigue subset. On the other hand, Figure 4.6b shows that bicep repetitions have significant correlations with seven out of 12 MAD features extracted from our complete dataset. Again, we encounter a similar effect to the aforementioned ones in extracted features (mean).

Some MAD features correlations became insignificant to bicep repetitions such as X-accelerometer, Y-magnetometer, and yaw. However, we did not observe a major drop in all MAD features' correlation coefficient values, only the newly three mentioned insignificant suffered from a drop in the correlation coefficient values.



(a) (b)

Figure 4.7: Graphical display of the differences in the correlation matrices of the 12 extracted features (SD: Standard Deviation) and the bicep repetitions with and without fatigue. (**a**) Correlation matrix of the 12 SD features and bicep repetitions in the non-fatigue subset. (**b**) Correlation matrix of the 12 SD features and the bicep repetitions in our complete dataset.

Figure 4.7 presents the correlation matrix for 12 SD features and the biceps repetitions in our complete dataset and the non-fatigue subset. Figure 4.7a shows that bicep repetitions have significant correlations with 10 out of 12 SD features extracted from the non-fatigue subset. On the other hand, Figure 4.7b shows that bicep repetitions have significant correlations with eight out of 12 SD features extracted from our complete dataset. Once more, some SD features correlations became insignificant to bicep repetitions such as the X-axis for both accelerometer and gyroscope. We also observe a slight drop in the correlation coefficient values for all SD features, as presented by the faint color intensity and the shrink of circle sizes. At this point, we clearly observe the same recurring effects when fatigue is introduced to the data, which indicates that fatigue impacts the significant features of a

HAR model.

In RQ3, our findings show that the more fatigue is added to the dataset, the steeper the decline in performance is on the five subject-specific models. Our findings show that the impact of fatigue can indeed disrupt the models' performance if not taken properly into account. From the evaluated models, our results indicate that FNN outperforms all other models in terms of precision, recall, accuracy, and F1-score in most cases. We did expect the highest performance from FNN compared to other models as this occurred in previous studies (Al-Mulla, Sepulveda, & Colley, 2011; González-Izal, Malanda, Gorostiaga, & Izquierdo, 2012; Lan, Feng, & Crago, 1994; Subasi & Kiymik, 2010). These studies show that neural networks have significantly better pattern recognition compared to other machine learning models especially, when it comes to periodic activities where extracted features inherit periodicity. Moreover, a popular reason for FNN performance supremacy is its robustness against small-to-moderate changes in the data. Other models, such as DT, has shown to be less robust to fatigue, as even smaller data pattern changes can cause a large change in the structure of the tree.

In RQ3 and RQ4 we compare the performances of the subject-specific and the cross-subject models. We observe a similar and significant performance loss in both models, with a loss of more than 20% if the dataset contains 20% or more of fatigue repetitions. Once again, the FNN has shown to be the most robust of the five evaluated models. Such a result is corroborated by another related work (Ghazal, Haeyeh, Abed, & Ghazal, 2018), which reported that FNN maintained the highest rate of accuracy in cross-subject experiments, when detecting fatigue in volunteers driving their vehicles.

4.7 Summary

Throughout this chapter, we introduce the impact of fatigue on HAR models for biceps concentration curls as an interesting and impactful data science problem. Specifically, its

significant challenges arise from analyzing the IMU collected data, selecting the suitable features, and evaluating the performance of HAR models in a dataset with realistic levels of fatigue. Throughout our study, we find that fatigue often occurs in later sets extending duration time up to 31% compared to the first set and decreasing the muscular endurance down to 4.1%. This leads to a change in data patterns, which causes a series of impacts such as hindering extracted features thus, decreasing models' performance.

As a result, the higher the presence of fatigue in the dataset, the steeper the performance of all models decline. Our findings show that FNN maintained the highest performance for cross-subject and subject-specific validations, respectively. Our results indicate that fatigue was a serious problem for machine learning models and we advise practitioners to take fatigue into consideration to develop and deploy accurate HAR systems. This chapter presents useful results and a solid start for enhancing real-world applications for HAR to overcome the inevitable impact of fatigue.

Chapter 5

Towards Detecting Biceps Muscle Fatigue in Gym Activity Using Wearables

In this chapter, we adopt a wearable approach to detect biceps muscle fatigue during a bicep concentration curl exercise as an example of a gym activity. This presents a solution to avoid fatigue-induced injuries by detecting fatigue levels in bicep muscles using wearable-based HAR models.

5.1 Introduction

Muscle fatigue is a complex and multifaceted phenomenon with various definitions; however, one of the most common definition of fatigue is ”failure to maintain the required force to continue performing a task” (Maughan, Maughan, & Gleeson, 2010; Robergs et al., 2004). Several publications in the literature have recently proposed fatigue detection approaches to avoid fatigue-induced injuries. Earlier studies usually propose the invasive approach, which requires measuring the lactic acid in the bloodstream to determine the

maximal muscle effort that a person can maintain (Stoudemire et al., 1996). While, other studies present the cardio-respiratory approach as the first non-invasive approach to detect fatigue; however, it requires a face mask to measure the circulatory and respiratory systems' ability to supply oxygen (O_2) to skeletal muscles during sustained physical exercise (Billat & Koralsztein, 1996). Recently, the wearable-based approach has gained momentum and interest after the rapid development of sensors technology. Nowadays, a wearable-based approach uses one or more wearable IMU to detect fatigue based on the rating of perceived exertion (RPE) (Op De Beéck et al., 2018).

In this chapter, we search for the most significant features to detect bicep muscle fatigue during bicep curls exercise, where participants have to train their muscles through incremental exercises. Since the bicep curls exercise has a repetitive nature, the collected data also have repetitive patterns. However, these patterns may not last for long in the presence of fatigue. Therefore, we have to find the changes in data patterns and extract the most significant features from these changes to detect fatigue once it kicks in. Also, we consider two variations for fatigue detection models: cross-subject and subject-specific models. In this work, we use our dataset from chapter 4 to extract fatigue detection features. We extract 16 significant features from a total of 33 features. Then, we employ these features to train and test five fatigue detection models.

5.2 Data Processing

Similar to section 3.2, we process the data collected from each volunteer by extracting their five sets of concentration curls. Then, for each set, we associate each concentration curl with the RPE values reported by the volunteers to identify whether a repetition contains fatigue or not. We extract and label each repetition manually, according to the RPE values reported for the set. In section 3.2, we mention the three-dimensional Cartesian coordinates (x, y, z) from gyroscope, magnetometer, and accelerometer; in this work, we add two more

computed signals from the accelerometer data, as the following:

- Total Acceleration: this is the vector sum of the tangential and centripetal accelerations, which makes it a place-independent signal, which means it does not rely on the exact attachment of the accelerometer because it combines x, y, and z acceleration signals at time t_i to compute a total acceleration, defined as: $\sqrt{a_{x_i}^2 + a_{y_i}^2 + a_{z_i}^2}$.
- Exerted Force: $F_{exerted} = m \times a$ is the exerted force by the a volunteer to lift the dumble. $F_{exerted}$ is calculated by multiplying the mass m of the lifted dumble by acceleration a.

Next, we detail the motivation, approach, and the findings for each research question. In addition, we evaluate the performance of the fatigue detection models and address the following research questions:

- RQ1: What are the most significant features to detect bicep muscles fatigue?
- RQ2: How accurately can we detect bicep muscles fatigue using subject-specific models?
- RQ3: How accurately can we detect bicep muscles fatigue using cross-subject models?

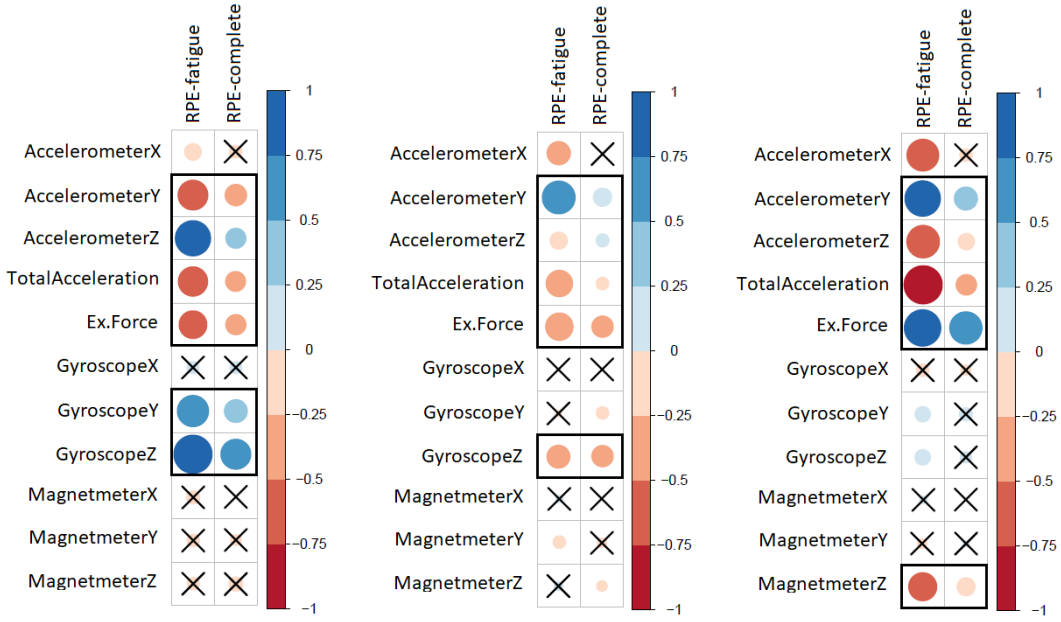
5.3 Significant Fatigue Detection Features for Bicep Curls

This section examines the significance of our extracted features and their correlation with the reported RPE values to answer RQ1: What are the most significant features to detect bicep muscles fatigue? **Motivation:** We hypothesize that the performance of detection models may degrade if we input non-relevant features. These features may increase the

uncertainty in the detection models, showing a decline in models' accuracy. Also, redundant features can introduce redundant information which may have no value to the model, hindering the model's performance (Suto, Oniga, & Sitar, 2017; M. Zhang & Sawchuk, 2011).

Approach: To fulfill our motivation, we group all the fatigue repetitions into a subset, so that we have a complete dataset and a fatigue subset. Then, we extract two sets of the 33 features, a set from the complete dataset and another set from the fatigue subset. Each set of features includes the mean, MAD, and SD of the three-dimensional Cartesian coordinates (x, y, z) from the gyroscope, magnetometer, and accelerometer; in addition, to the exerted force and total acceleration. Next, we apply a filter-based feature selection using Spearman's rank with a significance allowance of 0.1 on the features extracted from the fatigue subset to identify the most correlated features with the fatigue RPE values. Now, we repeat the previous step using the complete dataset. So, we apply a filter-based feature selection using Spearman's rank with a significance allowance of 0.1 on the features extracted from the complete dataset to identify the most correlated features with the overall reported RPE values. After that, we compare the two extracted sets features to select the overlapping features. These features are the most correlated with the reported RPE, yet are fatigue specific features.

Findings: Figure 5.1 presents the correlation matrices for 33 extracted features and the RPE values in the fatigue subset and our complete dataset. The positive and negative correlations are displayed in blue and red color, respectively. Additionally, the color intensity and the size of the circle are proportional to the correlation coefficients, whereas the insignificant correlations are marked with ($\times$). Figure 5.1a shows that the reported RPE values have overlapping significant correlations with 6 out of 11 mean features extracted from the fatigue subset and our complete dataset. These six significant features are (Y,Z)-Accelerometer, (Y,Z)-Gyroscope, total acceleration, and exerted force. In addition,



(a) Correlation matrix of the mean features and RPE.

(b) Correlation matrix of the SD features and RPE.

(c) Correlation matrix of the MAD features and RPE.

Figure 5.1: Graphical display of the differences in the correlation matrices of the 33 extracted features (mean, SD, MAD) and the RPE values in the fatigue subset and our complete dataset.

Figure 5.1b shows that the reported RPE values have overlapping significant correlations with 5 out of 11 SD features extracted from the fatigue subset and our complete dataset. These five significant features are (Y,Z)-Accelerometer, (Z)-Gyroscope, total acceleration, and exerted force. Furthermore, Figure 5.1c shows that the reported RPE values have overlapping significant correlations with 5 out of 11 MAD features extracted from the fatigue subset and our complete dataset. These five significant features are (Y,Z)-Accelerometer, (Z)-Magnetmeter, total acceleration, and exerted force. We can observe that the overlapped features remain significant and have higher correlation coefficient values to fatigue-reported RPE values, which indicate that these features are valuable to detect bicep muscle fatigue.

Table 5.1 summarizes the 33 extracted features and highlights the 16 overlapping features after applying spearman's rank. These features correlate most with the reported RPE and fatigue significantly.

Table 5.1: Table of the significant (✓) and insignificant (×) features extracted from both fatigue subset and complete dataset; the overlapping features are in highlighted bold.

			Fatigue subset			**Complete dataset**		
			Mean	**SD**	**MAD**	**Mean**	**SD**	**MAD**
Sensors and Axes	**Gyro.**	X-axis	×	×	×	×	×	×
		Y-axis	**✓**	×	✓	**✓**	✓	×
		Z-axis	**✓**	**✓**	✓	**✓**	**✓**	×
	Mag.	X-axis	×	×	×	×	×	×
		Y-axis	×	✓	×	×	×	×
		Z-axis	×	×	**✓**	×	✓	**✓**
	Acc.	X-axis	✓	✓	✓	×	×	×
		Y-axis	**✓**	**✓**	**✓**	**✓**	**✓**	**✓**
		Z-axis	**✓**	**✓**	**✓**	**✓**	**✓**	**✓**
		Total	**✓**	**✓**	**✓**	**✓**	**✓**	**✓**
		Ex.Force	**✓**	**✓**	**✓**	**✓**	**✓**	**✓**

5.4 Biceps Muscle Fatigue Detection Models Evaluation

After the feature extraction, we utilize the 16 overlapping features and the five models in section 4.1 to detect bicep muscle fatigue repetitions during biceps curls. Then, we evaluate the models' performance in the two variations: cross-subject and subject-specific models.

5.4.1 Performance Evaluation: Subject-Specific Models

This section estimates RPE values for each bicep curls repetition through subject-specific models to answer RQ2: How accurately can we detect bicep muscles fatigue using subject-specific models? **Motivation:** This section examines the performance of five subject-specific models in bicep muscle fatigue detection. Therefore, we measure the accuracy, precision, recall, and F1 for each model. For simplicity purposes, we start with subject-specificity, where we assess the reliability of our work and its ability to predict fatigue for a specific subject across different periods of time.

Table 5.2: Fatigue detection confusion matrix

		Actual	
		Fatigue ∈ [17,20]	**Non-Fatigue** ∈ **[6,16]**
Predict	**Fatigue** ∈ **[17,20]**	TRUE Fatigue	FALSE Fatigue
	Non-Fatigue ∈ **[6,16]**	FALSE Non-Fatigue	TRUE Non-Fatigue

Approach: To fulfill our motivation, we use these models to estimate each repetition's Borg rating (RPE) to determine whether it is fatigue or non-fatigue repetition. Then, we calculate the accuracy using the confusion matrix shown in Table 5.2, where non-fatigue repetition represents a Borg score from 6 to 16, and fatigue status represents a Borg score from 17 to 20. We calculate the accuracy using Equation (5), precision using Equation (6), recall using Equation (7), and F1 using Equation (8).

$$Accuracy = \frac{True(Fatigue + NonFatigue)}{True(Fatigue + NonFatigue) + False(Fatigue + NonFatigue)} \tag{5}$$

$$Precision = \frac{True(Fatigue)}{True(Fatigue) + False(Fatigue)} \tag{6}$$

$$Recall = \frac{True(Fatigue)}{True(Fatigue) + False(NonFatigue)} \tag{7}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{8}$$

Findings: In Table 5.3, we present the average subject-specific validations results using the five models. The two-layer Feedforward Neural Network seems to outperform all other models in terms of precision (95%), recall (93%), accuracy (94%), and F1-measure (94%). We did expect such high performance from FNN, compared to other models. Previous studies show that neural networks are often significantly better in pattern recognition, compared to other machine learning models; especially, when it comes to periodic activities such as bicep curls (Al-Mulla et al., 2011; González-Izal et al., 2012). Another reason for

the FNN's superior performance is that robustness against small-to-moderate changes in the data patterns. Whereas in DT, these changes can cause wide reformations in the tree's structure, causing instability. We observe that DT shows the lowest performance among the five models in terms of precision (66%), recall (61%), accuracy (58%), and F1-measure (63%). On the other hand, the GLM maintains the averaged performance across the five models in terms of precision (86%), recall (83%), accuracy (84%), and F1-measure (84%).

Table 5.3: Average precision, recall, and accuracy for subject-specific validations using the 16 extracted features to detect fatigue in biceps repetitions.

Models	Subject-Specific			
	Precision	Recall	Accuracy	F1
GLM	86%	83%	84%	84%
LR	81%	77%	79%	79%
RF	78%	76%	76%	77%
DT	66%	61%	58%	63%
FNN	**95%**	**93%**	**94%**	**94%**

5.4.2 Performance Evaluation: Cross-Subject Models

This section estimates RPE values for each bicep curls repetition through cross-subject models to answer RQ3: How accurately can we detect bicep muscles fatigue using cross-subject models? **Motivation:** This section examines the generality of the extracted features and the five models in bicep muscle fatigue detection across different subjects. Therefore, we measure the accuracy, precision, recall, and F1 for each model using leave-one-out cross validation (LOOCV).

Approach: To fulfill our motivation, we utilize the 16 extracted features with the five models, however, we use LOOCV to examine the performance of the five models. This is the same as a K-fold cross-validation, with K = 20 being equal to the number of volunteers. Figure 4.4 shows for a single model in the first iteration, we use 19 volunteers' datasets for training, excluding the 20th volunteer's dataset, which we use for testing. We

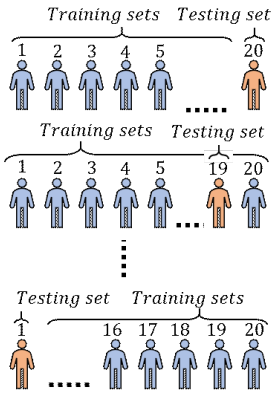


Figure 5.2: A representation of leave-one-out cross validation for a model using the 20 volunteers' datasets.

use the model to predict each repetition's Borg rating (RPE) to determine whether it is fatigue or non-fatigue repetition. Then, we calculate the precision, recall, and accuracy. Similarly, in the second iteration, we uses 19 volunteers' dataset for training, excluding the 19th volunteer's dataset, which we use for testing the model and calculating the precision, recall, and accuracy. In the 20th iteration, we should have used all volunteers' datasets for testing, except for the first volunteer's dataset; therefore, we train the model using all the 19 volunteers' datasets, then we use the first volunteer's dataset for testing the model and calculating the precision, recall, and accuracy. Finally, we compute the average values for precision, recall, and accuracy for the model using the same equations and the confusion matrix in section 5.4.1.

Findings: In Table 5.4, we present the average cross-subject validation results using the five models. The two-layers Feedforward Neural Network seems to outperforms all the other models in terms of precision (87%), recall (89%), accuracy (88%), and F1-measure (88%). In the case of cross-subject, we did expect FNN to maintain a superior performance compared to other models as this occurred in a previous study on detecting fatigue while driving where it did archive 95.8% of cross-validation accuracy (Ghazal et al., 2018). We observe that DT's performance drops significantly in terms of precision (47%), recall (49%), accuracy (43%), and F1-measure (48%). While, the GLM maintains the average

performance across the five models in terms of precision (78%), recall (71%), accuracy (75%), and F1-measure (74%).

Table 5.4: Average precision, recall, and accuracy for cross-subject validations using the 16 extracted features to detect fatigue in biceps repetitions.

Models	Cross-Subject			
	Precision	Recall	Accuracy	F1
GLM	78%	71%	75%	74%
LR	73%	74%	76%	73%
RF	69%	73%	70%	71%
DT	47%	49%	43%	48%
FNN	**87%**	**89%**	**88%**	**88%**

5.5 Discussion

Although RPE may pose a risk due to subjectivity due to individual differences such as physical fitness. The volunteers in this work are very close to athletic fitness. They are middle-age volunteers because athletes usually notice physical declines between 20 to 46 years with BMI ranging between 24 and 46 (Adirim & Cheng, 2003; Burt & Overpeck, 2001; Prentice & Jebb, 2001). Additionally, they have been gym-goers for at least 1 year. We construct a gold standard to counter RPE risks by combining heart rate and RPE value. The Borg rating ranges from 6 to 20, whereby multiplying the Borg rating by ten, we can estimate the person's heart rate during the activity. This serves as a way to strengthen the validity of the reported RPE for each volunteer. We employ a tolerance of ± 10 bpm to convert the measured heart rate to RPE before verifying the convergence between the Borg scale and the measured heart rate. We only found a small minority of repetitions where the heart rate metrics and the Borg scale diverge. In the worst case scenario, we found a volunteer that reported 5 out of 80 repetitions (6.2%) with a Borg scale dissimilar to the measured heart rate. To address these cases, we averaged between the measured heart

rate converted to RPE and the reported RPE, as done in similar work (Yoo et al., 2017). For example, if a volunteer reported a repetition of an RPE of 17, but we measured their heart rate as 145 bpm, we first converted the heart rate to RPE: 14.5. Then we averaged both metrics, $(17 + 14.5)/2 = 15.75$, rounded up to 16. The RPE of 16 is used for the labeling of this repetition (repetition without fatigue).

Our work adopts the wearable approach to detect bicep muscle fatigue using a wearable IMU and smartwatch. This allows us to overcome drawbacks from early approaches, such as complexity, discomfort, and invasion. First, regarding complexity, our work is simple compared to the works of early approaches. Our work requires only an IMU and a smartwatch as data acquisition devices, which are fairly easy to interact with and setup, whereas other approaches may require expert supervision, such as fatigue monitoring systems (Koutsos, Cretu, & Georgiou, 2016). Second, regarding discomfort, our work spins around portability and being light-weight compared to the works of early approaches. Our work used a Neblina IMU and an Apple Watch Series 4 that weigh 1.3 g and 40 g, respectively. Such light-weight devices do not hinder or interfere with the person's activity, whereas other approaches may require a face mask to measure oxygen consumption VO_2, which is inconvenient in public and often hinders a person's comfort (Billat & Koralsztein, 1996). Third, regarding invasion, our work was non-invasive compared to the works of early approaches. Our work does not introduce any instruments into a person's body or require a puncture of the skin. We simply attach the Neblina IMU and Apple Watch Series 4 on the person's wrist, unlike, the invasive approaches that often require blood lactate (Stoudemire et al., 1996), creatine kinase (Kobayashi et al., 2005), or rectal temperature (Crewe et al., 2008).

In this work, we present 16 overlapping features highlighted in Table 5.1, which are the most fatigue-specific and highly correlated with the reported RPE. However, if we look at the Table from another perspective, we would notice six non-overlapping but significant

features under the fatigue subset. This means fatigue may disturb data patterns over time, alternating features from significant into insignificant status or vice versa. A previous work suggests a decrease in the mean power frequency of the accelerometer readings trend with increasing bicep muscle fatigue, altering some of the extracted features from significant to insignificant (Mokaya et al., 2016). A fatigue implication can be viewed if we consider building a model to detect/count repetitions of biceps concentration curls and neglect fatigue's significant features. This may cause the model failure to in detecting/counting the fatigued biceps concentration curls as repetitions. The reason for such a dilemma is fatigue affects the collected data; its impacts will extend to the extracted features from the same data. A previous work (Triwiyanto et al., 2018) shows that muscle fatigue affects the electromyography (EMG) data signals collected from biceps by increasing the Root Mean Square Error (RMSEs), leading to the misclassification of some activities.

5.6 Summary

Throughout this chapter, we select 16 most fatigue representative features from a total of 33 features. Then, we employ these features in five fatigue detection models to detect fatigue in bicep curls. Our findings show that a two-layer FNN can achieve an accuracy of 98% and 88% for subject-specific and cross-subject models, respectively. Moreover, our methodology aims to detect fatigue for one of the most active skeletal muscles at the elbow joint, which is achievable according to our findings. Thus, we advise athletes to take fatigue into consideration to avoid fatigue-induced injuries. The results presented in this work are useful and represent a solid start for moving into real-world applications for detecting the fatigue level in bicep muscles using wearable sensors.

Chapter 6

The Personalization of Biceps Fatigue Detection Model For Gym Activity: An Approach To Utilize Wearables' Data From The Crowd

This chapter proposes a personalization approach that utilizes a portion of user-specific data to evaluate crowd data based on similarity metrics, resulting in less variability in training datasets and improving the cross-subject model's performance. Our motivation comes from observing the cross-subject models perform less than subject-specific ones when we detect bicep muscle fatigue by estimating the RPE during bicep curls (Elshafei & Shihab, 2021). Sometimes, using cross-subject models is preferable in serving a large crowd since these models utilize crowd data, thus, reducing the demand for user-specific data. So, given the advantage of using cross-subject for a large crowd and the detrimental effects of bicep fatigue injuries, such as muscle strain and tendon rupture (Nesterenko et al., 2010), we are eager to improve the performance of cross-subject models in fatigue detection.

6.1 Introduction

HAR applications in motion tracking and athletic training have become more prominent as wearable technology and machine learning techniques advance (Barshan & Yüksek, 2014; Shoaib, Bosch, Incel, Scholten, & Havinga, 2014). Yet, a common obstacle in such applications is having sufficient data to train the HAR models reliably (Barshan & Yurtman, 2016; Lockhart & Weiss, 2014). Training HAR models using insufficient data limits their performance and may even make them impractical for their user base. One way to work around this obstacle is to collect data from a large pool of users and train a cross-subject HAR model. However, this does not guarantee an accurate performance from the cross-subject HAR model because even with sufficient data from a large pool of users, individuals may perform the same activity differently. This increases the inter-subject data variability, which hinders the performance of HAR applications (Barshan & Yurtman, 2016). The inter-subject data variability is often high in places where there is a diverse crowd of users with different physical traits (Barshan & Yurtman, 2016; Kristiansen, Madeleine, Hansen, & Samani, 2015). A way to reduce the inter-subject data variability is to address each user separately by collecting data from the user of the HAR application to train a subject-specific HAR model. However, the cost of training a subject-specific HAR model is often prohibitive and requires labeled data from the user (Kobsar & Ferber, 2018; Lubetzky-Vilnai, Ciol, & McCoy, 2014; Mourão-Miranda et al., 2011). Therefore, there is an inherent trade-off between cross-subject models (cheaper but less accurate) versus subject-specific models (more expensive and more accurate). Furthermore, such a trade-off often exacerbates in specialized cases in HAR (e.g., muscle fatigue detection), where manual or semi-supervised labeling is usually required (Fredriksson, Mattos, Bosch, & Olsson, 2020; Nweke et al., 2018). As a result, this increases the data cost in the case of the subject-specific models; or, if we want to spare that data cost, we will choose the less accurate option, the cross-subject models.

In this chapter, we attempt to bridge the trade-off between cross-subject and subject-specific models. We propose a personalization approach to improve the performance of cross-subject fatigue detection models. Recently, other researchers suggest breaking down the crowd into smaller groups and using a cross-subject model to serve each group based on the common features extracted per group known as the personalized model (Ferrari et al., 2020; Khan, Roy, & Misra, 2018; Palmius et al., 2018). Therefore, in this chapter, we hypothesize that utilizing the personalization approach in bicep muscle fatigue detection is beneficial to the cross-subject models' performance and it can reduce the hindering effect of the inter-subject data variability. Also, to strengthen our hypothesis, we study the similarity traits between the test subject and individuals in the crowd to improve the quality of selected data for training our models. We believe that our hypothesis is achievable for two reasons:

(1) Previous studies show that the personalization of the cross-subject models can improve their performance in classifying the activities of daily living with promising results (Fallahzadeh & Ghasemzadeh, 2017; Ferrari et al., 2020; Sztyler & Stuckenschmidt, 2017).

(2) Other studies show that prioritizing collected data from individuals in the crowd who share similarities with the test subject can reduce inter-subject variability in the training dataset (Y. Chen, Wang, Huang, & Yu, 2019; Khan et al., 2018; Lane et al., 2011).

6.2 Data Processing

Similar to section 3.2, we extract and label the data collected from each volunteer manually, according to the RPE values reported for the set. In this section, we use the three-dimensional Cartesian coordinates (x, y, z) from gyroscope and accelerometer; in addition to the two computed signals: total acceleration and exerted force from section 5.2. Also,

we use Acc–gyro data fusion to compute a 3rd additional signal combined from the accelerometer and the gyroscope data: Kalman filter.

- Total Acceleration: this is the vector sum of the tangential and centripetal accelerations, which makes it a place-independent signal, which means it does not rely on the exact attachment of the accelerometer because it combines x, y, and z acceleration signals at time t_i to compute a total acceleration, defined as: $\sqrt{a_{x_i}^2 + a_{y_i}^2 + a_{z_i}^2}$.

- Exerted Force: $F_{exerted} = m \times a$ is the exerted force by the a volunteer to lift the dumble. $F_{exerted}$ is calculated by multiplying the mass m of the lifted dumble by acceleration a.

- Acc–gyro data fusion (Kalman filter): A complementary filter is often used to detect human body movement patterns by combining the gyroscope and the accelerometer (Alarfaj, Qian, & Liu, 2021; Webber & Rojas, 2021). Gyroscope's data are used for precision because it is not vulnerable to external forces, while the accelerometer's data are used for long-term tracking as it does not drift. We use the Kalman filter algorithm to estimate roll, pitch, and yaw angles (Q. Li, Li, Ji, & Dai, 2015). However, we use the yaw angle because it indicates the sideways vibration for the volunteer's hand during the extension and flexion of the bicep. Previous studies show fatigue may cause a temporary movement disorder, such as skeletal muscles vibration, which indicates fatigue backlogs and increases the vibration angle (Nweke, Teh, Mujtaba, & Al-Garadi, 2019; Palumbo, Gallicchio, Pucci, & Micheli, 2016; Y. Wang, Cang, & Yu, 2018; Wichit & Choksuriwong, 2015). In the filter's simplest form, the equation is defined as: $angle = 0.98 \times (angle + gyro \times dt) + 0.02 \times acc$.

6.3 Feature Extraction

Feature extraction is a crucial component of HAR systems because it establishes the most significant parameters to identify or predict human body movements. In addition, feature extraction reduces the data dimensionality while preserving the relevant characteristics of the signal. In this section, we compute a total of eleven hand-crafted features, as shown in Table 6.1. Eight of the selected features are proven accurate in previous works, especially in the general classification of human activities (Bianco, Napoletano, & Schettini, 2019; Ferrari, Micucci, Mobilio, & Napoletano, 2019; Z. Huang, Niu, You, & Pau, 2021; Janidarmian et al., 2017; Op De Beéck et al., 2018; Vanrell, Milone, & Rufiner, 2017). These features include min, max, mean, median, SD, variance, kurtosis, and RMS. Besides the eight features mentioned before, we also select three other features often associated with fatigue for better performance: skewness, IoP, and MSP (Aghamohammadi-Sereshki, Bayazi, Ghomsheh, & Amirabdollahian, 2019; Q. Ji, Lan, & Looney, 2006; Mallis, Mejdal, Nguyen, & Dinges, 2004; Sant'Ana, Li, & Zhang, 2019). A previous study suggests considering the skewness of the data when detecting fatigue in repetitive muscle movements such as bicep curls (Elshafei et al., 2021). We select skewness as a fatigue feature because, during the repetitions' extraction and labeling process, we observed the following: (1) Non-fatigue repetitions are relatively symmetrical during the repetitions' extraction and labeling process. (2) In contrast, fatigue repetitions are often positively skewed. Another work shows that fatigue often occurs in later sets, which increases the time to complete repetitions of bicep curls while decreasing the force exerted by the muscles (Elshafei & Shihab, 2021). This is observable through the increments of intervals between peaks, e.g., IoP, and decrements of peaks' amplitudes, e.g., MSP. For each volunteer, we extract the eleven features on all repetitions, across all nine signals, including the two 3D signals (x, y, z) from the accelerometer and the gyroscope, total acceleration, exerted Force, and acc–gyro signal fusion.

Table 6.1: Eleven hand-crafted features: eight HAR-related features and three fatigue-related features.

	Feature	Formula
Centralized	Minimum	$min = min_{i=1,...N}(x_i)$
	Maximum	$max = max_{i=1,...N}(x_i)$
	Mean	$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$
	Median	$M = \begin{cases} x_{\frac{N+1}{2}}, \text{ N odd} \\ \frac{1}{2}(x_{\frac{N}{2}} + x_{\frac{N}{2}+1}), \text{N even} \end{cases}$
	Standard Deviation (SD)	$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2}$
	Variance	$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$
	Kurtosis	$\text{K} = \frac{1}{N}\sum_{i=1}^{N}\frac{(x_i-\overline{x})^4}{\sigma^4}$
	Root Mean Square (RMS)	$\text{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x\_\overline{x})^2}$
Fatigue	Skewness	$\text{Sk} = \frac{1}{N}\sum_{i=1}^{N}\frac{(x_i-\overline{x})^3}{\sigma^3}$
	Interval of Peaks (IoP)	$IoP = T_p - T_{p-1} : p = 2, ...N$
	Mean Slope between Peaks (MSP)	$MSP = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{p_j-p_i}{T_{p_j}-T_{p_i}}$

6.4 Measuring Similarities

To visualize the concept of our work, let us assume that a test subject is selected from a diverse population P of size n, as shown in Figure 6.1. Each member of the population reports their physical traits along with bicep concentration curl data signals. Meanwhile, the test subject provides only partial data, often one set of repetitions, of their bicep concentration curl data signals along with their physical traits. We measure the similarities between the test subject and members of the population so that the data from whom the test subject is similar gain more weight while training the model. We are keen to utilize two types of similarities, physical similarity and signal similarity, because previous studies have reported gains in performance when harnessing those similarities to weight data from

the crowd (Ferrari et al., 2020; Lane et al., 2011).

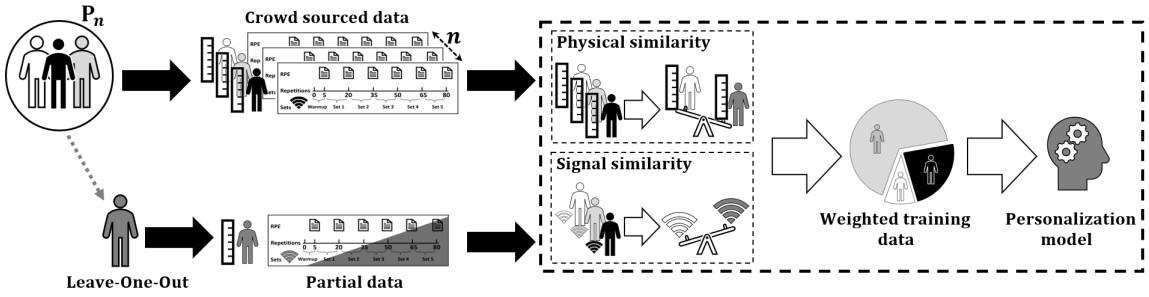


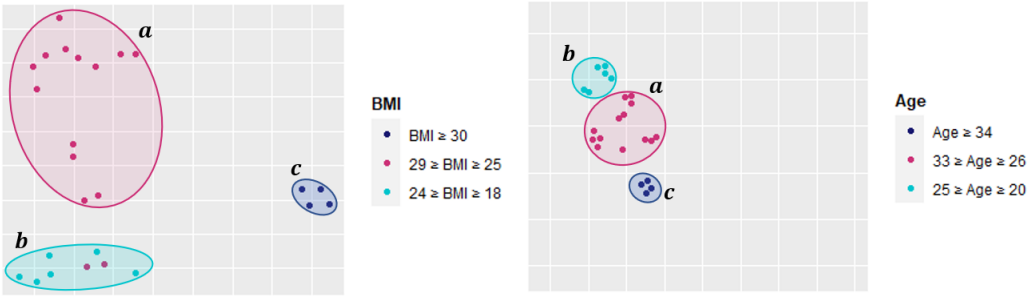
Figure 6.1: Visualization of the concept of personalizing general model using crowd-sourced wearables' data.

We hypothesize that combining both physical and signal similarities may further improve the personalized models' performance. Therefore, we measure and compare the performance between the personalized models trained using the weighted data and the cross-subject models. Furthermore, it comes to our minds that if we already possess and use a part of the test subject's data to measure the similarities between the test subject and the crowd, then we may let the personalized models consume it in training to improve their learning. Therefore, we also decided to let subject-specific models consume the same part of the test subject's data; then, we compare the performances of personalized and subject-specific models. Moreover, we allow the subject-specific models to consume more of the test subject's data if needed until it can reach the same performance of personalized models so that we quantify the amount of spared data by using personalized models.

Measuring Physical Similarity

Physical characteristics of people (e.g., age, weight, height, or BMI) vary from one person to another within a large population. Such differences can affect the way people move and perform physical activities. We believe that a user with different physical traits, e.g., age and BMI, may show signs of fatigue differently. At the same time, we expect groups

of people who share similar physical traits to show similar signs of fatigue (Morgan, Smeuninx, & Breen, 2020; Tomlinson et al., 2021). For example, let us capture the signs of fatigue using the three fatigue-related features, skewness, IoP, and MSP, and plot the principal component analysis (PCA). In Figure 6.2a, we often observe that individuals within specific limits of BMI values tend to share similar signs of fatigue. Moreover, Figure 6.2b shows similar observations where we use age instead of BMI. We can observe that individuals of certain ages tend to share similar signs of fatigue. This strengthens our hypothesis that if we construct the training data from individuals within the population who are more similar, it may reduce the inter-subject data variability and hence improve the performance of the fatigue detection model.



(a) (b)

Figure 6.2: PCA plots showing signs of fatigue captured by the three fatigue-related features and BMI/age. (**a**) BMI perspective. (**b**) Age perspective.

To compute the physical similarity value between a pair of users, we employ four types of physical traits: age, height, weight, and BMI. To limit the widespread of the values, due to subjects' variations, we apply min–max normalization to each physical trait, on training data, to normalize each trait between 0 and 1. We combine these four traits per user to form a dedicated physical vector V^{Phy} = {age, height, weight, BMI} representing their physical traits separately. We measure the distance d^{Phy} between the physical traits V^{Phy} of two users (q, p) based on the Manhattan distance, as shown in Equation (9). Previous works show that Manhattan distance is preferable to Euclidean for high dimensional data and if

the dimensions are not comparable (Aggarwal, Hinneburg, & Keim, 2001; Malkauthekar, 2013; Shirkhorshidi, Aghabozorgi, & Wah, 2015). The physical similarity between users (q, p) is based on the universal law of generalization proposed in previous works (Ferrari et al., 2020; Lane et al., 2011; Shepard, 1987; Tenenbaum & Griffiths, 2001), where distance and perceived similarity are related via an exponential function, as shown in Equation (10):

$$d^{Phy}(q,p) = \sum_{k=1}^{4} |V_{q_k}^{Phy} - V_{p_k}^{Phy}| \tag{9}$$

$$sim^{Phy}(q,p) = \frac{1}{e^{\gamma d^{Phy}(q,p)}} \tag{10}$$

where γ is an empirically determined scaling parameter that affects the shape of the exponential function. For example, $\lim_{\gamma \to \infty} sim^{Phy}(q,p) = 0$, which indicates that as γ approaches infinity, the physical similarity approaches zero, causing more segregation between users. This can be a double-edged sword because as we segregate dissimilar users from each other, we may increase the segregation between similar users unintentionally. On the other hand, $\lim_{\gamma \to 0} sim^{Phy}(q,p) = 1$, which indicates that as gamma approaches zero, the physical similarity approaches one, implying that all subjects show similar signs of fatigue; in other words, the changes in their data patterns are similar. Again, this is a double-edged scenario where we may unintentionally pull dissimilar users near to the similar users. Therefore, further investigation is required to estimate the optimal value of γ.

Measuring Signal Similarity

In the context of signal similarities, we use one set of repetitions, approximately 20% of the subject's data needed for the subject-specific models. We believe that users within the same population may show similar signs of fatigue, leading to similar changes in data patterns while performing the exercise. To compute the signal similarity value between a

pair of users, we employ the 11 extracted features in Table 6.1 to form a dedicated signal vector V^{Sig} = {min, max, ... , MSP} for each user. We measure the distance d^{Sig} between the signal traits V^{Sig} of two users (q, p) based on the Manhattan distance for all repetitions l = {1, 2, ... , L}, as shown in Equation (11).

$$d^{Sig}(q,p) = \sum_{k=1}^{11} \sum_{l=1}^{L} |V^{Sig}_{q_{(k,l)}} - V^{Sig}_{p_{(k,l)}}| \tag{11}$$

$$sim^{Sig}(q,p) = \frac{1}{e^{\gamma d^{Sig}(q,p)}} \tag{12}$$

The signal similarity between users (q, p) is based on the distance between their vectors, as shown in Equation (12).

Measuring Total Similarity

We measure the total similarity sim^{Total} between two users (q, p) by summing their weighted physical $sim^{Phy}(q, p)$ and signal $sim^{Sig}(q, p)$ similarities, as shown in Equation (13).

$$sim^{Total}(q,p) = \alpha \times sim^{Phy}(q,p) + \beta \times sim^{Sig}(q,p) \tag{13}$$

where $\alpha + \beta = 1$. If α is greater than β, the physical similarity will contribute more than signal similarity in determining the total similarity value. On the other hand, if β is greater than α, the signal similarity will be the one that dominates the total similarity value. Therefore, we further investigate the impact of (α, β) values on the performance of the personalized models. Moreover, we examine (γ) values to achieve the highest performance possible.

Next, we evaluate the personalization approach in boosting the performance of cross-subject models and answer the following research questions:

- RQ1: What is the impact of the physical and signal parameters on the performance of the personalized biceps fatigue detection models?
- RQ2: Can the personalization approach improve the performance of cross-subject models in detecting biceps muscle fatigue?
- RQ3: Can the personalization approach reduce the consumption of the test subject's data in comparison to subject-specific models?

6.5 Examining the Parameters in the Personalized Biceps Fatigue Detection Model

This section searches through different values of personalization parameters and observes their effects on the models to answer RQ1: What is the impact of the physical and signal parameters on the performance of the personalized biceps fatigue detection models? **Motivation:** We think that the performance of users similarity-based models, such as those driven from the personalization approach, may degrade if the inadequate parameters are selected. In other words, valuable data from the crowd, e.g., similar users, may be discarded due to an unintended preference for the physical similarity over signal similarity and vice versa. Previous work shows that finding a balance between the extracted similarities is important to improve the accuracy of the models constantly (Zhu, Wang, Zhang, & Xu, 2014).

Approach: To fulfill our motivation, we examine the possible values for the parameters (α, β, γ) in Equations (10), (12), and (13) so that we observe the impact of these parameters on the models' performance. We start with Equations (10) and (12) to examine γ, which has an arbitrary value between (0, ∞). To observe the impact of the different γ values on the models' performance, we employ each γ value to run two pairs of models:

AdaBoost-DT and AdaBoost-ANN. The first pair is physical similarity-based models of AdaBoost-DT and AdaBoost-ANN, while the second pair is signal similarity-based models of AdaBoost-DT and AdaBoost-ANN. Then, we compute the changes in these models' performance as the value of γ increases and select the gamma value that corresponds to the best performance. In Equation (13), we examine the physical α and signal β parameters while satisfying the condition $\alpha + \beta = 1$. We experiment with different α and β values and observe the effect by running two similarity-based models (AdaBoost-DT, AdaBoost-ANN). Then, we compute the models' performance as the values of α and β change; then, we select the (α, β) values that correspond to the best performance.

Findings: Figure 6.3 shows the average changes in both models' performance as the value of γ increases. The performance in this context is measured using accuracy. We use the accuracy at $\gamma = 0$ as the reference point to measure the changes in accuracy ($\Delta Accuracy$) as the γ value increases. We observe that $\Delta Accuracy$ increases as the γ value increases until both reach maximum values of 3.83 and 14 at the dashed line, respectively. Then, the changes in accuracy start to decline as the γ value continues to increase. Therefore, we select the $\gamma = 14$ to let the models perform at maximum accuracy.

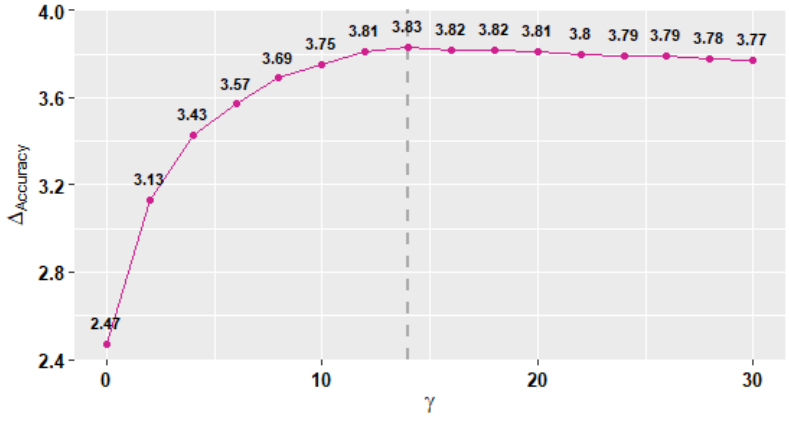


Figure 6.3: The average changes in both models' accuracy as the value of γ increases.

Figure 6.4 shows the impact of the physical α and signal β parameters on models' performance and there are three important findings in the figure. In the first finding, at $\alpha = 0$

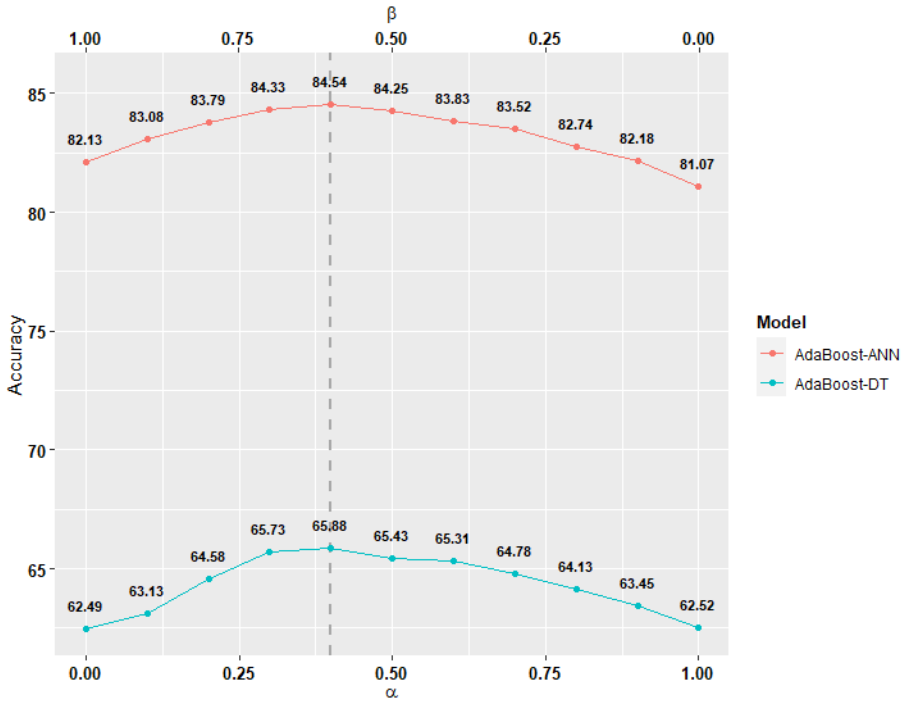


Figure 6.4: The average models' accuracy as the values of α and β change.

and $\beta = 1$, the models (AdaBoost-ANN, AdaBoost-DT) solely depend on the signal similarity, leading the training dataset, for these models, to be selected from the crowd with whom the test subject's signal is similar while discarding the physical similarity. In this case, the two models (AdaBoost-ANN, AdaBoost-DT) achieve accuracy of 82.13% and 62.49%, respectively. In the second finding, at $\alpha = 1$ and $\beta = 0$, the models (AdaBoost-ANN, AdaBoost-DT) solely depend on the physical similarity, leading the training dataset, for these models, to be selected from the crowd with whom the test subject's physical traits is similar while discarding the signal similarity. In this case, the accuracy for the two models (AdaBoost-ANN, AdaBoost-DT) drops to 81.07% and 62.52%, respectively. In the third finding, at $\alpha = [0.25, 0.50]$ and $\beta = (1-\alpha)$, the models (AdaBoost-ANN, AdaBoost-DT) depend on both physical and signal similarities; however, the training dataset, for these models, is selected from the crowd with whom the test subject's is similar while prioritizing those with the highest signal similarity. In this case, we observe that the accuracy

for the two models (AdaBoost-ANN, AdaBoost-DT) rises to reach 84.54% and 65.88%, respectively.

Our findings show that both physical and signal similarities are important. Moreover, the models' performance reaches its peak when $\gamma = 14$. The best selected values for $\alpha = 0.4$ and $\beta = 0.6$ which we use for the rest of the evaluations.

6.6 Evaluating the Performance of Personalized Models

This section uses personalized fatigue models to estimate RPE values bicep curls repetition with a leave-one-out cross-validation (LOOCV) to answer RQ2: Can the personalization approach improve the performance of cross-subject models in detecting biceps muscle fatigue? **Motivation:** The common trade-off for having cross-subject models to server large crowds is accuracy loss, especially for users with particular activity patterns. In other words, users who do not share enough similarities with the crowd may look as outliers where the cross-subject models are less accurate to detect their biceps muscle fatigue. We think that adding weight to user's data from whom the test subject is similar can improve model accuracy, including marginal users. Results of a previous study show that the personalization of cross-subject models constantly improves their accuracy compared with the standard cross-subject models (Sztyler & Stuckenschmidt, 2017).

Approach: To fulfill our motivation, we use the 11 hand-crafted features along with the 2 similarity-based models (AdaBoost-DT, AdaBoost-ANN) to predict the Borg rating for each repetition and detect whether a repetition contains fatigue or not; therefore, we run two experiments.

In the first experiment, we set $(\alpha = 0, \beta = 0)$ in both models to mimic the standard cross-subject models. This means that the training dataset for these models is collected without considering any type of similarity between the crowd and the test subject. For this experiment, we use LOOCV, which is a K-fold cross-validation with K equal to the

number of volunteers ($K = 25$). In the second experiment, we set (α, β) to optimal values as identified in section 6.5, leading the training dataset, for these models, to be selected based on physical and signal similarities, in addition to prioritizing data coming from users of highest signal similarity in the crowd. For each experiment, we calculate the accuracy using the confusion matrix shown in Table 6.2, where non-fatigue repetition represents a Borg score from 6 to 16, and fatigue status represents a Borg score from 17 to 20. We calculate the accuracy using Equation (14), precision using Equation (15), recall using Equation (16), and F1 using Equation (17).

Table 6.2: Fatigue detection confusion matrix

		Actual	
		Fatigue ∈ [17,20]	**Non-Fatigue** ∈ **[6,16]**
Predict	**Fatigue** ∈ **[17,20]**	TRUE Fatigue	FALSE Fatigue
	Non-Fatigue ∈ **[6,16]**	FALSE Non-Fatigue	TRUE Non-Fatigue

$$Accuracy = \frac{True(Repeat + NonRepeat)}{True(Repeat + NonRepeat) + False(Repeat + NonRepeat)} \tag{14}$$

$$Precision = \frac{True(Repeat)}{True(Repeat) + False(Repeat)} \tag{15}$$

$$Recall = \frac{True(Repeat)}{True(Repeat) + False(NonRepeat)} \tag{16}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{17}$$

Findings: It is important to mention that we use one set of 15 repetitions approximately from the test subject's data during the personalization of DT and ANN to measure the signal similarity between the test subject and individuals in the crowd. However, we do not include these 15 repetitions nor any data from the test subject in the training set. Our findings show that the personalization approach improves the accuracies for the models

by 5.89% (DT) and by 3.38% (ANN), as shown in Table 6.3. The accuracy improved after prioritizing training data from the crowd based on the total similarity score, where individuals with high scores contribute more to the models' training dataset. Moreover, we observe other improvements in terms of precision, recall, and F1-measure across the models. The results show that the personalization improves the DT model in terms of precision by (1.20%), recall by (4.51%), and F1-measure to (2.81%). On the other hand, the personalization improves the ANN model in terms of precision by (6.96%), recall by (4.55%), and F1-measure to (5.82%). Moreover, we observe that the standard cross-subject ANN model outperforms both DT models, which is expected in fatigue detection wherein ANN models often perform better than other cross-subject models (Ghazal et al., 2018).

Table 6.3: Average precision, recall, and accuracy, with a CI of 95%, for detecting fatigue in biceps repetitions before and after the personalization of cross-subject models.

	Models					
	DT			**ANN**		
	Cross-Subject	**Personalized**	Δ	**Cross-Subject**	**Personalized**	Δ
Precision	60.57% $\pm$ 0.66	61.77% $\pm$ 0.58	1.20%	73.29% $\pm$ 0.43	80.25% $\pm$ 0.47	6.96%
Recall	61.32% $\pm$ 0.53	65.83% $\pm$ 0.49	4.51%	78.53% $\pm$ 0.39	83.08% $\pm$ 0.43	4.55%
Accuracy	60.08% $\pm$ 0.49	65.97% $\pm$ 0.67	5.89%	82.41% $\pm$ 0.58	85.79% $\pm$ 0.48	3.38%
F1	60.94% $\pm$ 0.59	63.75% $\pm$ 0.53	2.81%	75.82% $\pm$ 0.41	81.64% $\pm$ 0.45	5.82%

Overall, our findings indicate that the personalization approach improves both models in terms of performance. For the DT model, the personalization improves its F1-measure from (60.94%) to (63.75%), while for the ANN model, the personalization improves its F1-measure from (75.82%) to (81.64%).

6.7 Examining the Consumption of the Test Subject's Data in the Personalization Approach

This section measures the test subject's required data for subject-specific models to detect fatigue in bicep curls accurately compared to personalized models to answer RQ3: Can the personalization approach reduce the consumption of the test subject's data in comparison to subject-specific models? **Motivation:** The subject-specific models are known for their high performance and demand of the test subject data. In contrast, the cross-subject models have relatively lower performance and no test subject data demand. We hypothesize that the personalized approach can combine the best aspects of these two models. In other words, the personalization approach can improve the performance of the cross-subject models while consuming less test subject data than the subject-specific models. A previous work shows that adding a small amount of the test subject's data to the training dataset for personalized models helps to improve the performance further closer to the subject-specific models (Weiss & Lockhart, 2012).

Approach: To fulfill our motivation, we utilize the 11 hand-crafted features and 4 models: subject-specific (DT, ANN) models and personalization (AdaBoost-DT, AdaBoost-ANN) models. We use these models to predict the Borg rating for each repetition to determine whether it is fatigue repetition or not. Similar in section 6.6, we set (α, β) to optimal values so that the training dataset for these models is selected based on physical and signal similarities while prioritizing the signal similarity selection. Our experiment consists of seven runs where we incrementally add 10% of the test subject's data to the training set after each run. This means, there is 0% of the test subject's data added to the training dataset at the 1st run, while at the 7th run, there is 60% of the test subject's data added to the training dataset. We stop at 60% of the test subject's data to prevent overfitting the personalized model; otherwise, there will not be much of a difference between the subject-specific and

the personalization approaches. Moreover, this allows us to measure the amount of the test subject's data needed to improve the personalization models' performance closer to the subject-specific models. Or, in other words, how little data we need from the test subject if we use the personalized models instead of subject-specific ones while keeping the accuracy relativity high. For each run, we calculate the accuracy using Equation (1) and accuracy gain ratio (AGR) using Equation (18).

$$AGR = \frac{\Delta accuracy}{\text{Test subject's data (repetitions)}} \tag{18}$$

Table 6.4: The accuracy averages for the subject-specific and personalized models after adding 10% of the test subject's data to the training set in each run incrementally. We include a version of this table with the confidence intervals in the appendix .1.

			Number of biceps repetitions collected from the test subject (% of used test's data)						
			0 (0%)	**8 (10%)**	**15 (20%)**	**23 (30%)**	**30 (40%)**	**38 (50%)**	**45 (60%)**
Models	**DT**	**Subject-specific**	Accuracy	15.34%	41.30%	58.40%	68.60%	78.90%	82.20%	84.03%
			$\Delta accuracy$	-	25.96%	17.10%	10.20%	10.30%	3.30%	1.83%
			AGR	-	3.25%	1.14%	0.44%	0.34%	0.09%	0.04%
		Personalized	Accuracy	65.88%	70.64%	76.08%	77.77%	79.15%	79.55%	79.91%
			$\Delta accuracy$	-	4.76%	5.44%	1.69%	1.38%	0.40%	0.36%
			AGR	-	0.60%	0.36%	0.07%	0.05%	0.01%	0.01%
	ANN	**Subject-specific**	Accuracy	55.23%	62.48%	74.68%	82.95%	87.37%	90.45%	92.99%
			$\Delta accuracy$	-	7.25%	12.20%	8.27%	4.42%	3.08%	2.54%
			AGR	-	0.91%	0.81%	0.36%	0.15%	0.08%	0.06%
		Personalized	Accuracy	84.54%	87.37%	92.74%	93.56%	93.88%	94.25%	94.78%
			$\Delta accuracy$	-	2.83%	5.37%	0.82%	0.32%	0.37%	0.53%
			AGR	-	0.35%	0.36%	0.04%	0.01%	0.01%	0.01%

Findings: Our findings show that the more the test subject's data are added to the training set, the higher the accuracy of the subject-specific and personalized models. Table 6.4 shows that the subject-specific DT model achieves an accuracy of 78.90% after consuming 40% of the test subject's data. On the other hand, the personalization of the DT model

achieves an accuracy of 76.08% after consuming 20% of the test subject's data while compensating the rest of the training data from the similar users in the crowd. In other words, the subject-specific DT model requires twice the amount of test subject data, at 40%, to achieve similar accuracy to the personalized DT at 20% of the test subject's data consumption with taking into consideration that the personalized DT model compensates the rest of the training data from the crowd. Moreover, our findings show that with 20% of the test subject's data, the personalized DT model reaches the lowest accuracy gain ratio of 0.36% per test subject's repetition while maintaining the highest accuracy gain of 5.44%.

Furthermore, the subject-specific ANN model achieves an accuracy of 92.99% after consuming 60% of the test subject's data. On the other hand, the personalization of the ANN model achieves an accuracy of 92.74% after consuming 20% of the test subject's data while compensating the rest of the training data from the similar users in the crowd. In other words, the subject-specific ANN model requires triple the amount of test subject's data, at 60%, to achieve similar accuracy to the personalized ANN at 20% of the test subject's data consumption with taking into consideration that the personalized ANN model compensates the rest of the training data from the crowd. Moreover, our findings show that with 20% of the test subject's data, the personalized ANN model reaches the lowest accuracy gain ratio of $\approx$0.36% per test subject's repetition while maintaining the highest accuracy gain of 5.37%.

Our findings show that the personalization approach may reduce the test subject's data consumption by 33.3% up to 50.0% while reducing the accuracy gap compared to the subject-specific models.

6.8 Discussion

In RQ1, we observe that $\Delta Accuracy$ increases following a γ increase until it reaches a maximum of 3.83 at $\gamma = 14$. Then, the accuracy starts to drop slightly with higher

values of γ. Previous studies report similar behavior for the gamma parameter in their results sections (Ferrari et al., 2020; Lane et al., 2011). They observe their models' accuracy increases with an increase in γ value until γ reaches an optimal point, where their models then start losing accuracy. Although gamma's behavior seems similar, the γ values are different and depend on the dataset. The reason behind gamma's behavior resides in Equations (10) and (12) where we find that the physical and signal similarities approach zero as $\gamma \rightarrow \infty$. This means, if we keep increasing the gamma values, we will push the test subject further away from the crowd. In other words, we, unintentionally, decrease the possibility of finding similar users in the crowds, resulting in fewer similar data points, and hence smaller training data. On the other hand, when $\gamma \rightarrow 0$, the physical and signal similarities approach 1. This means, if we keep decreasing the gamma values, we will push the test subject closer toward the crowd, increasing the possibility of finding similar users. However, this can increase the risk of including low-quality data points from similar users with low ranks, which is usually the case in a cross-subject model; therefore, the accuracy often drops.

In RQ2, our findings show improvements in accuracy for both personalized models, AdaBoost-DT and AdaBoost-ANN, compared to the standard cross-subject ones, by 5.89% and 3.38%, respectively. Such improvements occur because the training datasets for the personalized models are selected from users who's physical and signal traits are similar to the test subject. A previous study reports similar findings to ours, indicating that personalized models often perform +3% better than standard cross-subject models (Sztyler & Stuckenschmidt, 2017). However, their proposed personalized model averages 0.78 for F1-score, while our AdaBoost-ANN model performs 3.64% better with an average of 81.64% $\pm$ 0.45 for F1-score. Overall, while this has been shown in previous works, our results help consolidate the benefits on relying on similarity as a method for boosting the performance of cross-subject models.

In RQ3, our finding indicates that the personalized models, AdaBoost-DT and AdaBoost-ANN, achieve comparable performance to subject-specific models while consuming 50.0% and 66.77% less test subject data. This is an important finding to motivate approaches that rely less on the data of the subject, particularly in cases where the test subject's data are difficult to obtain or very limited. A previous study utilized a personalization approach to cut down the cost of data labeling by up to 90% for new users (Hong, Ramos, & Dey, 2016). The study reports model accuracy between 77.7% and 83.4%. In contrast, we can observe that our personalized models, AdaBoost-DT and AdaBoost-ANN, achieve closer or higher accuracies at 76.08% $\pm$ 0.71 and 92.74% $\pm$ 0.49 at similar rates of 20% test data consumption, respectively, as shown in Table 6.5. This table shows the accuracy achieved by fatigue detection models including the cross-subject, subject-specific, and personalized models. As an implication, our findings suggest that personalized models are an effective approach to reduce data dependency—when data on the target subject is scarce—without severely compromising the model's performance.

Table 6.5: Percent accuracy achieved on, with a CI of 95%, the cross-subject, subject-specific, and personalization models.

Models	(% of Used Test's Data)	Accuracy	ΔAccuracy		
			Cross-Subject	Personalized	Subject-Specific
DT	Cross-Subject (0%)	60.08% $\pm$ 0.49	–	-16.00%	-28.67%
	Personalized (20%)	76.08% $\pm$ 0.71	16.00%	–	-12.67%
	Subject-Specific (100%)	88.75% $\pm$ 0.59	28.67%	12.67%	–
ANN	Cross-Subject (0%)	82.41% $\pm$ 0.58	–	-10.33%	-16.89%
	Personalized (20%)	92.74% $\pm$ 0.49	10.33%	–	-6.56%
	Subject-Specific (100%)	99.30% $\pm$ 0.37	16.89%	6.56%	–

Moreover, we can observe that both of the personalization models achieve higher accuracies compared to the cross-subject models. However, we find that the personalization ANN models achieve lesser accuracy improvement than the personalization DT models.

This agrees with previous work that shows that AdaBoost usually achieves higher improvement results on weak classifiers such as DT than stronger ones such as ANN (Subasi et al., 2018).

6.9 Summary

This work aims to mitigate the hindering effect of subject data variability in fatigue detection. We propose the personalization approach to utilize data from the crowd based on the total similarity score between the test subject and the crowd. Our dataset consists of 3,750 concentration curl repetitions from 25 volunteers with ages and BMI ranging between 20–46 and 24–46, respectively. We compute the total similarity score between each individual in the crowd and the test subject based on the physical and signal similarity scores. Then, we extract a weighted dataset to train our models. Our findings show that the AdaBoost-DT model outperforms the cross-subject-DT model by 5.89%, while the AdaBoost-ANN model outperforms the cross-subject-ANN model by 3.38%. On the other hand, at 50.0% less of the test subject's data consumption, our AdaBoost-DT model outperforms the subject-specific-DT model by 16%, while the AdaBoost-ANN model outperforms the subject-specific-ANN model by 10.33%. Our findings indicate that crowd data are usable to build personalized bicep fatigue detection models to prevent athletes from fatigue-induced injuries. Moreover, our personalization approach benefits real-life applications when the data from the test subject is unavailable or insufficient. We believe that our work is useful and represents a solid start for moving into real-world applications for detecting the fatigue level in bicep muscles using wearables' data from the crowd.

Chapter 7

Conclusions and Future Work

This chapter concludes our research thesis where we summarize research findings, list work limitations, and propose future works.

7.1 Conclusion and Findings

Chapter 4: On the Impact of Biceps Muscle Fatigue in Human Activity Recognition

In this chapter, we used the biceps concentration curls exercise as an example of a gym activity to observe the impact of fatigue in wearable-based HAR system. Our findings indicate that fatigue often occurs in later sets of an exercise and extends the completion time of later sets by up to 31% and decreases muscular endurance by 4.1%. Another finding shows that changes in data patterns are often occurring during fatigue presence, causing seven features to become statistically insignificant. Further findings indicate that fatigue can cause a substantial decrease in performance in both subject-specific and cross-subject models. Finally, we observed that a FNN showed the best performance in both cross-subject and subject-specific models in all our evaluations.

Chapter 5: Towards Detecting Biceps Muscle Fatigue in Gym Activity Using Wearables

In this chapter, we adopted a wearable approach to detect biceps muscle fatigue during a bicep concentration curl exercise as an example of a gym activity. We observed from our data that fatigue reduces the biceps' angular velocity; therefore, it increases the completion time for later sets. We extracted a total of 33 features from our dataset, which have been reduced to 16 features. These features are the most overall representative and correlated with bicep curl movement, yet they are fatigue-specific features. We utilized these features in five detection models; however, we found that using a two-layer FNN achieves an accuracy of 98% and 88% for subject-specific and cross-subject models, respectively. The results presented in this chapter are useful and represent a solid start for moving into a real-world application for detecting the fatigue level in bicep muscles using wearable sensors as we advise athletes to take fatigue into consideration to avoid fatigue-induced injuries.

Chapter 6: The Personalization of Biceps Fatigue Detection Model For Gym Activity: An Approach To Utilize Wearables' Data From The Crowd

In this chapter, we presented a personalized model that achieves higher performance than the cross-subject model while maintaining a lower data cost than the subject-specific model. Our personalization approach sources data from the crowd based on similarity scores computed between the test subject and the individuals in the crowd. We compute 11 hand-crafted features and train two personalized models: AdaBoost-DT and AdaBoost-ANN, using data from whom the test subject shares similar physical and single traits. Our findings show that the AdaBoost-DT model outperforms the cross-subject-DT model by 5.89%, while the AdaBoost-ANN model outperforms the cross-subject-ANN model by

3.38%. On the other hand, at 50.0% less of test subject's data consumption, our AdaBoost-DT model outperforms the subject-specific-DT model by 16% while the AdaBoost-ANN model outperforms the subject-specific-ANN model by 10.33%. Yet, the subject-specific models achieve the best performances at 100% of test subjects' data consumption.

7.2 Limitations

7.2.1 Limited Data

The first limitation of our work is the data size, which may affect the external validity of our study. While some HAR studies have opted to use public datasets, datasets with fatigue data are not common nor often available to the public (Ferrari et al., 2020; Lin & Marculescu, 2020). Since we have to collect our fatigue data during the COVID-19 pandemic, it has been a daunting task due to social distancing and restrictive measures. Although our dataset may look small in size, we believe it is suitable for our research under such circumstances as other studies also collected their dataset with similar sizes to ours (Jebelli & Lee, 2019; Wan, Qi, Xu, Tong, & Gu, 2020). We agree that a bigger dataset is beneficial to our work, but we believe our experiments/ approach can generate similar performance approximately.

7.2.2 Technology and Equipment

The second limitation of our work is the reliance on the Apple Watch Series which uses the photoplethysmography (PPG) sensor to measure participants' heart rate during the exercise. Although Apple Watch can provide the most accurate readings amongst the optical wrist wearables (Gillinov et al., 2017), previous works show that PPG often suffers from inaccuracies. This means our results may be indirectly impacted (Gil et al., 2010; Schäfer & Vagedes, 2013); however, we believe such technology does not compromise our findings,

especially in real-life applications. Previous works show that PPG achieves clinically acceptable accuracy and might be considered safe for rehabilitation training programs (Falter, Budts, Goetschalckx, Cornelissen, & Buys, 2019; Shcherbina et al., 2017).

On the other hand, dumbbell weight can directly vary the data points collected during the exercise because of each participant's physical capacity or strength. A previous study shows lightweight dumbbells lead to a long recording session with many similar data points until participants reach fatigue (Reis et al., 2017). In contrast, heavyweight dumbbells lead to shorter recording sessions with fewer data entries, which do not capture kinetic changes clearly throughout the exercise because participants reach fatigue quickly. Although we use a 4.5 kg weight dumbbell as recommended by previous studies, we believe having dumbbell weights will provide us with more information and different patterns of biceps muscle fatigue (Bergquist et al., 2018; Hwang et al., 2016; Liao et al., 2021).

7.2.3 Reliability of Borg Scale

The third limitation of our work is the use of the Borg scale and the dumbbell weight. Although the Borg scale is often used in sports science, some studies are often cautious about its implications (Arney et al., 2019; Ciolac et al., 2015; Sala et al., 2021). Using subjective measures such as the Borg scale to report RPE may introduce a dependency between the correctness of selected Borg rating and participants' awareness. Therefore, we introduce the concept of the Borg scale to the participants in advance to avoid misevaluating their perceived exertion rate. Also, a previous study on an Asthma Quality of Life Questionnaire (AQLQ) study shows that the Borg scale can provide highly correlated fatigue assessment scores with other accurate fatigue approaches (e.g., correlation coefficients with %HRmax, VO2max, and total AQLQ score were 0.86, 0.89 and 0.61, respectively). Thus, overall, the Borg scale is more convenient yet considered valid and reliable scale to assess fatigue (Grammatopoulou, Skordilis, Koutsouki, & Baltopoulos, 2008).

7.3 Future Work

Although this thesis has taken several steps towards addressing the muscle fatigue challenge in HAR, many different avenues for future work remain unexplored. This section lists some of the potential and interesting future work.

7.3.1 Replication: Using a Different Data Source in Fatigue Detection

Our work uses inertia data collected from a 50 Hz Neblina IMU alongside an Apple Watch Series 4. Chapter 3 of this thesis describes how we collected our data set from 25 volunteers and processed the data for labeling. However, an important question is left unanswered: **What about different data sources such as Electromyography (EMG) in fatigue detection?**

Previous studies show that the EMG sensors can provide detailed information about muscle conditions during incremental exercises (De la Peña, Polo, & Robles-Algarín, 2019; Jebelli & Lee, 2019). Moreover, recent studies that utilize the EMG sensors show that some EMG-based fatigue detection models can achieve average accuracies ranging between 97.2% – 98.5% during incremental exercises (L. K. Huang et al., 2020; M. Li, Li, & Shu, 2020; S. Wang, Tang, Wang, & Mo, 2021; G. Zhang, Morin, Zhang, & Etemad, 2018). We believe utilizing such a different data source like EMG can improve the performance of our fatigue detection models. However, collecting and processing EMG sensors data from bicep muscles will require effort, coding, and time; therefore, we opt for EMG sensors in our future work.

7.3.2 Investigation: Is Fatigue Just Noise?

Chapter 4 evaluates the impact of fatigue in wearable-based HAR, including the collected data, extracted features, and models performance. Our findings showed that the

performance of HAR models tends to decrease in the presence of fatigue. However, an important question is left unanswered: **What differentiates fatigue from data noise in the HAR's dataset?**

Previous studies mention fatigue as one of the main challenges of HAR in healthcare applications because it often, especially in elderly cases, accounts for a considerable portion of the data (Dinarević, Husić, & Baraković, 2019; Nguyen, Coelho, Bastos, & Krishnan, 2021; Schrader et al., 2020). However, previous studies present fatigue as a natural symptom that manifests due to decreases in muscles' ability to perform exercise over time. Therefore, omitting fatigue from the HAR dataset would weaken the models in real-life applications. Also, the data analysis of the previous datasets shows fatigue presence is too frequent for it to be data noise in the data collected from elderly volunteers (Dinarević et al., 2019; Triwiyanto et al., 2018; Yu et al., 2019). We believe that a comparison study between fatigue and data noise may shed light on the differences between these two. Also, this will answer whether noise reduction approaches apply to fatigue or not. This requires a dedicated effort, coding, and time to select the proper exercise for the elderly; then, collecting and processing their data; therefore, we propose such a comparison study in our future work.

7.3.3 Extension: Exploring Deep Learning in Fatigue Detection

Chapter 5 combines the wearable approach and learning models to detect biceps muscle fatigue during bicep curls. Although we used hand-crafted features and conventional learning models, our findings show that a two-layer FNN can achieve an accuracy of 98% and 88% for subject-specific and cross-subject models, respectively. In chapter 6, we propose a personalization approach to improve the performance of cross-subject fatigue detection models; we select DT and ANN models to examine our personalization approach (Elshafei et al., 2021; Elshafei & Shihab, 2021). However, an important question is left unanswered:

What can deep learning achieve in fatigue detection?

After applying deep learning, a recent study shows a significant improvement in HAR systems performance, from 89.83% to 96.62% (Gil-Martín, San-Segundo, Fernandez-Martinez, & Ferreiros-López, 2020). However, none of the experiments presented in the study included fatigue detection or exhaustion rates prediction; therefore, future work should expand to include deep learning in fatigue detection, believing that it may provide better accuracy results. Furthermore, future work should examine the features extracted by Convolutional Neural Networks (CNNs) and compare their performance with the hand-crafted ones.

7.3.4 Application: Personalized Fatigue Detection in Parkinson's Patients

This section lists a potential application for the fatigue personalization approach. This application is based on the multi-agent system (MAS) and usability of crowdsource data. In chapter 6, we mitigated the hindering effect of subject data variability in fatigue detection from previous works to improve the cross-subject models' performance (Elshafei et al., 2021; Elshafei & Shihab, 2021). Now, an important question is left unanswered: **What are the applications of the fatigue personalization approach in real-life?**

Data collection is often a common challenge in applying HAR to elderlies, especially Parkinson's patients (Antar, Ahmed, & Ahad, 2019; Dinarević et al., 2019; Y. Wang et al., 2018; Y. Wang, Cang, & Yu, 2019). The challenge lies in two problems: 1) Parkinson's fatigue data is usually small because Parkinson's patients often describe fatigue as quick and extensive exhaustion that prevents them from moving. 2) Parkinson's disease leads to muscle shaking and stiffness which builds up fatigue continuously. As a result, fatigue may suddenly reach extreme levels and become life-threatening with the slightest patient activities (e.g., making breakfast, answering a call, taking a shower). Therefore, making use of

each bit of collected fatigue data from Parkinson's patients is crucial. With the fatigue personalization approach, we can utilize fatigue data from similar Parkinson's patients to train and improve the performance of cross-subject Parkinson's fatigue models while demanding fewer data points from the target Parkinson's patient. Therefore, we suggest applying our proposed fatigue personalization approach in fatigue detection for Parkinson's patients in future work.

Table .1: Appendix-A The accuracy averages for the subject-specific and personalized models after adding 10% of the test subject's data to the training set in each run incrementally.

			Number of biceps repetitions collected from the test subject (% of used test's data)						
			0 (0%)	**8 (10%)**	**15 (20%)**	**23 (30%)**	**30 (40%)**	**38 (50%)**	**45 (60%)**
Models	**DT**								
	Subject-specific	Accuracy	15.34%±0.83	41.30%±0.78	58.40%±0.71	68.60%±0.63	78.90%±0.57	82.20%±0.48	84.03%±0.41
		$\Delta accuracy$	-	25.96%	17.10%	10.20%	10.30%	3.30%	1.83%
		AGR	-	3.25%	1.14%	0.44%	0.34%	0.09%	0.04%
	Personalized	Accuracy	65.88%±0.65	70.64%±0.68	76.08%±0.71	77.77%±0.66	79.15%±0.59	79.55%±0.51	79.91%±0.47
		$\Delta accuracy$	-	4.76%	5.44%	1.69%	1.38%	0.40%	0.36%
		AGR	-	0.60%	0.36%	0.07%	0.05%	0.01%	0.01%
	ANN								
	Subject-specific	Accuracy	55.23%±0.79	62.48%±0.72	74.68%±0.64	82.95%±0.58	87.37%±0.45	90.45%±0.47	92.99%
		$\Delta accuracy$	-	7.25%	12.20%	8.27%	4.42%	3.08%	2.54%
		AGR	-	0.91%	0.81%	0.36%	0.15%	0.08%	0.06%±0.33
	Personalized	Accuracy	84.54%±0.63	87.37%±0.55	92.74%±0.49	93.56%±0.42	93.88%±0.48	94.25%±0.45	94.78%±0.39
		$\Delta accuracy$	-	2.83%	5.37%	0.82%	0.32%	0.37%	0.53%
		AGR	-	0.35%	0.36%	0.04%	0.01%	0.01%	0.01%

Table .2: Appendix-B Abbreviations list

Abbreviations	**Meaning**	**Page**
HAR	Human Activity Recognition	1
IMU	Inertial Measurement Unit	4
RPE	Rate of Perceived Exertion	4
VO$_2$ max	Maximum Volume of Oxygen Consumption	13
O$_2$	Oxygen	13
VO$_2$	Volume of Oxygen Consumption	13
GSR	Galvanic Skin Response	14
BMI	Body Mass Index	17
GLM	Generalized Linear Model	27
LR	Logistic Regression	27
RF	Random Forest	27
DT	Decision Trees	27
FNN	Feedforward Neural Network	27
MAD	Mean Absolute Deviation	33
SD	Standard Deviation	33
LOOCV	Leave-One-Out Cross-Validation	39
EMG	Electromyography	44
RMSE	Root Mean Square Error	44
IoP	Interval of Peaks	67
MSP	Mean Slope between Peaks	67
PCA	Principal Component Analysis	69
ANN	Artificial Neural Network	73
AdaBoost-ANN	AdaBoost-Artificial Neural Network	73
AdaBoost-DT	AdaBoost-Decision Trees	73
AGR	Accuracy Gain Ratio	79

References

Abbood, H., Al-Nuaimy, W., Al-Ataby, A., Salem, S. A., & AlZubi, H. S. (2014). Prediction of driver fatigue: Approaches and open challenges. In *2014 14th uk workshop on computational intelligence (ukci)* (pp. 1–6).

Adirim, T. A., & Cheng, T. L. (2003). Overview of injuries in the young athlete. *Sports medicine*, *33*(1), 75–81.

Aggarwal, C. C., Hinneburg, A., & Keim, D. A. (2001). On the surprising behavior of distance metrics in high dimensional space. In *International conference on database theory* (pp. 420–434).

Aghamohammadi-Sereshki, A., Bayazi, M.-J. D., Ghomsheh, F. T., & Amirabdollahian, F. (2019). Investigation of fatigue using different emg features. In *2019 ieee 16th international conference on rehabilitation robotics (icorr)* (pp. 115–120).

Alarfaj, M., Qian, Y., & Liu, H. (2021). Detection of human body movement patterns using imu and barometer. In *2020 international conference on communications, signal processing, and their applications (iccspa)* (pp. 1–6).

Al-Mulla, M. R., Sepulveda, F., & Colley, M. (2011). An autonomous wearable system for predicting and detecting localised muscle fatigue. *Sensors*, *11*(2), 1542–1557.

Alsheikh, M. A., Selim, A., Niyato, D., Doyle, L., Lin, S., & Tan, H.-P. (2016). Deep activity recognition models with triaxial accelerometers. In *Workshops at the thirtieth aaai conference on artificial intelligence.*

Antar, A. D., Ahmed, M., & Ahad, M. A. R. (2019). Challenges in sensor-based human activity recognition and a comparative analysis of benchmark datasets: a review. In *2019 joint 8th international conference on informatics, electronics & vision (iciev) and 2019 3rd international conference on imaging, vision & pattern recognition (icivpr)* (pp. 134–139).

Apriantono, T., Nunome, H., Ikegami, Y., & Sano, S. (2006). The effect of muscle fatigue on instep kicking kinetics and kinematics in association football. *Journal of sports sciences*, *24*(9), 951–960.

Arney, B. E., Glover, R., Fusco, A., Cortis, C., de Koning, J. J., van Erp, T., ... Foster, C. (2019). Comparison of rpe (rating of perceived exertion) scales for session rpe. *International journal of sports physiology and performance*, *14*(7), 994–996.

Aviles-Cruz, C., Rodriguez-Martinez, E., Villegas-Cortez, J., & Ferreyra-Ramirez, A. (2019). Granger-causality: An efficient single user movement recognition using a smartphone accelerometer sensor. *Pattern Recognition Letters*, *125*, 576-583. doi: https://doi.org/10.1016/j.patrec.2019.06.029

Barshan, B., & Yüksek, M. C. (2014). Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units. *The Computer Journal*, *57*(11), 1649–1667.

Barshan, B., & Yurtman, A. (2016). Investigating inter-subject and inter-activity variations in activity recognition using wearable motion sensors. *The Computer Journal*, *59*(9), 1345–1362.

Batchuluun, G., Kim, J. H., Hong, H. G., Kang, J. K., & Park, K. R. (2017). Fuzzy system based human behavior recognition by combining behavior prediction and recognition. *Expert Systems with Applications*, *81*, 108-133. doi: https://doi.org/10.1016/j.eswa.2017.03.052

Bergquist, R., Iversen, V. M., Mork, P. J., & Fimland, M. S. (2018). Muscle activity

in upper-body single-joint resistance exercises with elastic resistance bands vs. free weights. *Journal of human kinetics*, *61*, 5.

Biagetti, G., Crippa, P., Falaschetti, L., Orcioni, S., & Turchetti, C. (2017). Human activity recognition using accelerometer and photoplethysmographic signals. In *International conference on intelligent decision technologies* (pp. 53–62).

Bianco, S., Napoletano, P., & Schettini, R. (2019). Multimodal car driver stress recognition. In *Proceedings of the 13th eai international conference on pervasive computing technologies for healthcare* (pp. 302–307).

Billat, L. V., & Koralsztein, J. P. (1996). Significance of the velocity at vo 2max and time to exhaustion at this velocity. *Sports Medicine*, *22*(2), 90–108.

Bogart, B. I., & Bogart, B. I. (2007). *Elsevier's Integrated Anatomy and Embryology*. Elsevier. (Book, Whole)

Borg, G. (1998). *Borg's perceived exertion and pain scales.* Human kinetics.

Borg, G. A. (1982). Psychophysical bases of perceived exertion. *Med sci sports exerc*, *14*(5), 377–381.

Borga, M., West, J., Bell, J. D., Harvey, N. C., Romu, T., Heymsfield, S. B., & Leinhard, O. D. (2018). Advanced body composition assessment: from body mass index to body composition profiling. *Journal of Investigative Medicine*, *66*(5), 1–9.

Bosquet, L., Léger, L., & Legros, P. (2001). Blood lactate response to overtraining in male endurance athletes. *European journal of applied physiology*, *84*(1-2), 107–114.

Burkhauser, R. V., & Cawley, J. (2008). Beyond bmi: the value of more accurate measures of fatness and obesity in social science research. *Journal of health economics*, *27*(2), 519–529.

Burt, C. W., & Overpeck, M. D. (2001). Emergency visits for sports-related injuries. *Annals of emergency medicine*, *37*(3), 301–308.

Cannon, D. T., White, A. C., Andriano, M. F., Kolkhorst, F. W., & Rossiter, H. B. (2011).

Skeletal muscle fatigue precedes the slow component of oxygen uptake kinetics during exercise in humans. *The Journal of physiology*, *589*(3), 727–739.

Chen, K., Zhang, D., Yao, L., Guo, B., Yu, Z., & Liu, Y. (2021). Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities. *ACM Computing Surveys (CSUR)*, *54*(4), 1–40.

Chen, Y., Wang, J., Huang, M., & Yu, H. (2019). Cross-position activity recognition with stratified transfer learning. *Pervasive and Mobile Computing*, *57*, 1–13.

Ciolac, E. G., Mantuani, S. S., Neiva, C. M., Verardi, C. E. L., Pêssoa-Filho, D. M., & Pimenta, L. (2015). Rating of perceived exertion as a tool for prescribing and self regulating interval training: a pilot study. *Biology of sport*, *32*(2), 103.

Coelho, A. C., Cannon, D. T., Cao, R., Porszasz, J., Casaburi, R., Knorst, M. M., & Rossiter, H. B. (2015). Instantaneous quantification of skeletal muscle activation, power production, and fatigue during cycle ergometry. *Journal of Applied Physiology*, *118*(5), 646–654.

Crewe, H., Tucker, R., & Noakes, T. D. (2008). The rate of increase in rating of perceived exertion predicts the duration of exercise to fatigue at a fixed power output in different environmental conditions. *European journal of applied physiology*, *103*(5), 569.

Dang, L. M., Hassan, S. I., Im, S., & Moon, H. (2019). Face image manipulation detection based on a convolutional neural network. *Expert Systems with Applications*, *129*, 156–168.

Dang, L. M., Min, K., Wang, H., Piran, M. J., Lee, C. H., & Moon, H. (2020). Sensor-based and vision-based human activity recognition: A comprehensive survey. *Pattern Recognition*, *108*, 107561.

De, C. L. (1984). Myoelectrical manifestations of localized muscular fatigue in humans. *Critical reviews in biomedical engineering*, *11*(4), 251–279.

Debold, E. (2015). Potential molecular mechanisms underlying muscle fatigue mediated

by reactive oxygen and nitrogen species. *Frontiers in physiology*, *6*, 239.

Debold, E. P., Walcott, S., Woodward, M., & Turner, M. A. (2013). Direct observation of phosphate inhibiting the force-generating capacity of a miniensemble of myosin molecules. *Biophysical journal*, *105*(10), 2374–2384.

De la Peña, S., Polo, A., & Robles-Algarín, C. (2019). Implementation of a portable electromyographic prototype for the detection of muscle fatigue. *Electronics*, *8*(6), 619.

Demrozi, F., Pravadelli, G., Bihorac, A., & Rashidi, P. (2020). Human activity recognition using inertial, physiological and environmental sensors: a comprehensive survey. *IEEE Access*.

Dinarević, E. C., Husić, J. B., & Baraković, S. (2019). Issues of human activity recognition in healthcare. In *2019 18th international symposium infoteh-jahorina (infoteh)* (p. 1-6). doi: 10.1109/INFOTEH.2019.8717749

Edwards, T., Spiteri, T., Piggott, B., Bonhotal, J., Haff, G. G., & Joyce, C. (2018). Monitoring and managing fatigue in basketball. *Sports*, *6*(1). Retrieved from https://www.mdpi.com/2075-4663/6/1/19

Elshafei, M., Costa, D. E., & Shihab, E. (2021). On the impact of biceps muscle fatigue in human activity recognition. *Sensors*, *21*(4), 1070.

Elshafei, M., & Shihab, E. (2021). Towards detecting biceps muscle fatigue in gym activity using wearables. *Sensors*, *21*(3), 759.

Enoka, R. M., & Duchateau, J. (2008). Muscle fatigue: what, why and how it influences muscle function. *The Journal of physiology*, *586*(1), 11–23.

Enoka, R. M., & Duchateau, J. (2016). Translating fatigue to human performance. *Medicine and science in sports and exercise*, *48*(11), 2228.

Fallahzadeh, R., & Ghasemzadeh, H. (2017). Personalization without user interruption: Boosting activity recognition in new subjects using unlabeled data. In *Proceedings*

of the 8th international conference on cyber-physical systems (pp. 293–302).

Falter, M., Budts, W., Goetschalckx, K., Cornelissen, V., & Buys, R. (2019). Accuracy of apple watch measurements for heart rate and energy expenditure in patients with cardiovascular disease: Cross-sectional study. *JMIR mHealth and uHealth*, *7*(3), e11889.

Ferrari, A., Micucci, D., Mobilio, M., & Napoletano, P. (2019). Hand-crafted features vs residual networks for human activities recognition using accelerometer. In *2019 ieee 23rd international symposium on consumer technologies (isct)* (pp. 153–156).

Ferrari, A., Micucci, D., Mobilio, M., & Napoletano, P. (2020). On the personalization of classification models for human activity recognition. *IEEE Access*, *8*, 32066–32079.

Fredriksson, T., Mattos, D. I., Bosch, J., & Olsson, H. H. (2020). Data labeling: an empirical investigation into industrial challenges and mitigation strategies. In *International conference on product-focused software process improvement* (pp. 202–216).

Fu, B., Damer, N., Kirchbuchner, F., & Kuijper, A. (2020). Sensing technology for human activity recognition: A comprehensive survey. *IEEE Access*, *8*, 83791–83820.

Garrett Jr, W. E. (1996). Muscle strain injuries. *The American journal of sports medicine*, *24*(6_suppl), S2–S8.

Ghazal, M., Alhalabi, M., Fraiwan, L., Yaghi, M., & Alkhatib, L. (2019). Assessment of motion quality using an iot-based wearable and mobile joint flexion sensors. In *2019 7th international conference on future internet of things and cloud workshops (ficloudw)* (pp. 44–48).

Ghazal, M., Haeyeh, Y. A., Abed, A., & Ghazal, S. (2018). Embedded fatigue detection using convolutional neural networks with mobile integration. In *2018 6th international conference on future internet of things and cloud workshops (ficloudw)* (pp. 129–133).

Gil, E., Orini, M., Bailon, R., Vergara, J. M., Mainardi, L., & Laguna, P. (2010). Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions. *Physiological measurement*, *31*(9), 1271.

Gillinov, S., Etiwy, M., Wang, R., Blackburn, G., Phelan, D., Gillinov, A. M., ... Desai, M. Y. (2017). Variable accuracy of wearable heart rate monitors during aerobic exercise. *Med Sci Sports Exerc*, *49*(8), 1697–1703.

Gil-Martín, M., San-Segundo, R., Fernandez-Martinez, F., & Ferreiros-López, J. (2020). Improving physical activity recognition using a new deep learning architecture and post-processing techniques. *Engineering Applications of Artificial Intelligence*, *92*, 103679.

Golestani, N., & Moghaddam, M. (2020). Human activity recognition using magnetic induction-based motion signals and deep recurrent neural networks. *Nature communications*, *11*(1), 1–11.

González-Izal, M., Malanda, A., Gorostiaga, E., & Izquierdo, M. (2012). Electromyographic models to assess muscle fatigue. *Journal of Electromyography and Kinesiology*, *22*(4), 501–512.

Grammatopoulou, E., Skordilis, E., Koutsouki, D., & Baltopoulos, G. (2008). An 18-item standardized asthma quality of life questionnaire-aqlq (s). *Quality of Life Research*, *17*(2), 323–332.

Green, B., & Pizzari, T. (2017). Calf muscle strain injuries in sport: a systematic review of risk factors for injury. *British journal of sports medicine*, *51*(16), 1189–1194.

Gruet, M., Temesi, J., Rupp, T., Levy, P., Millet, G., & Verges, S. (2013). Stimulation of the motor cortex and corticospinal tract to assess human muscle fatigue. *Neuroscience*, *231*, 384–399.

Halson, S. L. (2014). Monitoring training load to understand fatigue in athletes. *Sports medicine*, *44*(2), 139–147.

Hauke, J., & Kossowski, T. (2011). Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data. *Quaestiones geographicae*, *30*(2), 87–93.

Hong, J.-H., Ramos, J., & Dey, A. K. (2016). Toward personalized activity recognition systems with a semipopulation approach. *IEEE Transactions on Human-Machine Systems*, *46*(1), 101-112.

Hopkins, W. G., Marshall, S. W., Quarrie, K. L., & Hume, P. A. (2007). Risk factors and risk statistics for sports injuries. *Clinical Journal of Sport Medicine*, *17*(3), 208–210.

Hsu, Y.-L., Yang, S.-C., Chang, H.-C., & Lai, H.-C. (2018). Human daily and sport activity recognition using a wearable inertial sensor network. *IEEE Access*, *6*, 31715–31728.

Huang, L. K., Huang, L. N., Gao, Y., Vasić, Ž. L., Cifrek, M., & Du, M. (2020). Electrical impedance myography applied to monitoring of muscle fatigue during dynamic contractions. *IEEE Access*, *8*, 13056–13065.

Huang, Z., Niu, Q., You, I., & Pau, G. (2021). Acceleration feature extraction of human body based on wearable devices. *Energies*, *14*(4), 924.

Hwang, H.-J., Chung, W.-H., Song, J.-H., Lim, J.-K., & Kim, H.-S. (2016). Prediction of biceps muscle fatigue and force using electromyography signal analysis for repeated isokinetic dumbbell curl exercise. *Journal of Mechanical Science and Technology*, *30*(11), 5329–5336.

Ignatov, A. (2018). Real-time human activity recognition from accelerometer data using convolutional neural networks. *Applied Soft Computing*, *62*, 915–922.

Jalal, A., Kim, Y.-H., Kim, Y.-J., Kamal, S., & Kim, D. (2017). Robust human activity recognition from depth video using spatiotemporal multi-fused features. *Pattern recognition*, *61*, 295–308.

Janidarmian, M., Roshan Fekr, A., Radecka, K., & Zilic, Z. (2017). A comprehensive analysis on wearable acceleration sensors in human activity recognition. *Sensors*,

17(3), 529.

Janssen, I., Katzmarzyk, P. T., & Ross, R. (2002). Body mass index, waist circumference, and health risk: evidence in support of current national institutes of health guidelines. *Archives of internal medicine*, *162*(18), 2074–2079.

Javaid, H. A., Rashid, N., Tiwana, M. I., & Anwar, M. W. (2018). Comparative analysis of emg signal features in time-domain and frequency-domain using myo gesture control. In *Proceedings of the 2018 4th international conference on mechatronics and robotics engineering* (pp. 157–162).

Jebelli, H., & Lee, S. (2019). Feasibility of wearable electromyography (emg) to assess construction workers' muscle fatigue. In *Advances in informatics and computing in civil and construction engineering* (pp. 181–187). Springer.

Jeong, G.-M., Truong, P. H., & Choi, S.-I. (2017). Classification of three types of walking activities regarding stairs using plantar pressure sensors. *IEEE Sensors Journal*, *17*(9), 2638–2639.

Ji, Q., Lan, P., & Looney, C. (2006). A probabilistic framework for modeling and real-time monitoring human fatigue. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and humans*, *36*(5), 862–875.

Ji, X., Cheng, J., Feng, W., & Tao, D. (2018). Skeleton embedded motion body partition for human action recognition using depth sequences. *Signal Processing*, *143*, 56–68.

Jones, B. H., Hauret, K. G., Dye, S. K., Hauschild, V. D., Rossi, S. P., Richardson, M. D., & Friedl, K. E. (2017). Impact of physical fitness and body composition on injury risk among active young adults: a study of army trainees. *Journal of science and medicine in sport*, *20*, S17–S22.

Karatzaferi, C., Franks-Skiba, K., & Cooke, R. (2008). Inhibition of shortening velocity of skinned skeletal muscle fibers in conditions that mimic fatigue. *American Journal*

of Physiology-Regulatory, Integrative and Comparative Physiology, *294*(3), R948–R955.

Kellmann, M. (2010). Preventing overtraining in athletes in high-intensity sports and stress/recovery monitoring. *Scandinavian journal of medicine & science in sports*, *20*, 95–102.

Khan, M. A. A. H., Roy, N., & Misra, A. (2018). Scaling human activity recognition via deep learning-based domain adaptation. In *2018 ieee international conference on pervasive computing and communications (percom)* (pp. 1–9).

Kobayashi, Y., Takeuchi, T., Hosoi, T., Yoshizaki, H., & Loeppky, J. A. (2005). Effect of a marathon run on serum lipoproteins, creatine kinase, and lactate dehydrogenase in recreational runners. *Research quarterly for exercise and sport*, *76*(4), 450–455.

Kobsar, D., & Ferber, R. (2018). Wearable sensor data to track subject-specific movement patterns related to clinical outcomes using a machine learning approach. *Sensors*, *18*(9), 2828.

Koutsos, E., Cretu, V., & Georgiou, P. (2016). A muscle fibre conduction velocity tracking asic for local fatigue monitoring. *IEEE Transactions on Biomedical Circuits and Systems*, *10*(6), 1119-1128. doi: 10.1109/TBCAS.2016.2520563

Kristiansen, M., Madeleine, P., Hansen, E. A., & Samani, A. (2015). Inter-subject variability of muscle synergies during bench press in power lifters and untrained individuals. *Scandinavian journal of medicine & science in sports*, *25*(1), 89–97.

Kuhn, M., et al. (2008). Building predictive models in r using the caret package. *Journal of statistical software*, *28*(5), 1–26.

Lan, N., Feng, H.-Q., & Crago, P. E. (1994). Neural network generation of muscle stimulation patterns for control of arm movements. *IEEE Transactions on Rehabilitation Engineering*, *2*(4), 213–224.

Lane, N. D., Xu, Y., Lu, H., Hu, S., Choudhury, T., Campbell, A. T., & Zhao, F. (2011).

Enabling large-scale human activity inference on smartphones using community similarity networks (csn). In *Proceedings of the 13th international conference on ubiquitous computing* (pp. 355–364).

Lee, C.-B., Eun, D., Kim, K.-H., Park, J.-W., & Jee, Y.-S. (2017). Relationship between cardiopulmonary responses and isokinetic moments: the optimal angular velocity for muscular endurance. *Journal of exercise rehabilitation*, *13*(2), 185.

LEIGH, J. P. (2011). Economic burden of occupational injury and illness in the united states. *The Milbank Quarterly*, *89*(4), 728-772.

Li, M., Li, J., & Shu, M. (2020). Detection of muscle fatigue by fusion of agonist and synergistic muscle semg signals. In *2020 ieee 33rd international symposium on computer-based medical systems (cbms)* (pp. 95–98).

Li, Q., Li, R., Ji, K., & Dai, W. (2015). Kalman filter and its application. In *2015 8th international conference on intelligent networks and intelligent systems (icinis)* (pp. 74–77).

Liao, F., Zhang, X., Cao, C., Hung, I. Y.-J., Chen, Y., & Jan, Y.-K. (2021). Effects of muscle fatigue and recovery on complexity of surface electromyography of biceps brachii. *Entropy*, *23*(8), 1036.

Lin, C.-Y., & Marculescu, R. (2020). Model personalization for human activity recognition. In *2020 ieee international conference on pervasive computing and communications workshops (percom workshops)* (pp. 1–7).

Liu, Y., Nie, L., Liu, L., & Rosenblum, D. S. (2016). From action to activity: sensor-based activity recognition. *Neurocomputing*, *181*, 108–115.

Lockhart, J. W., & Weiss, G. M. (2014). Limitations with activity recognition methodology & data sets. In *Proceedings of the 2014 acm international joint conference on pervasive and ubiquitous computing: Adjunct publication* (pp. 747–756).

Lubetzky-Vilnai, A., Ciol, M., & McCoy, S. W. (2014). Statistical analysis of clinical prediction rules for rehabilitation interventions: current state of the literature. *Archives of physical medicine and rehabilitation*, *95*(1), 188–196.

Ma, C., Li, W., Cao, J., Wang, S., & Wu, L. (2014). A fatigue detect system based on activity recognition. In *International conference on internet and distributed computing systems* (pp. 303–311).

Mair, S. D., Seaber, A. V., Glisson, R. R., & Garrett JR, W. E. (1996). The role of fatigue in susceptibility to acute muscle strain injury. *The American Journal of Sports Medicine*, *24*(2), 137–143.

Malkauthekar, M. (2013). Analysis of euclidean distance and manhattan distance measure in face recognition. In *Third international conference on computational intelligence and information technology (ciit 2013)* (pp. 503–507).

Mallis, M. M., Mejdal, S., Nguyen, T. T., & Dinges, D. F. (2004). Summary of the key features of seven biomathematical models of human fatigue and performance. *Aviation, space, and environmental medicine*, *75*(3), A4–A14.

Maughan, R. J., Maughan, R. J., & Gleeson, M. (2010). *The biochemical basis of sports performance*. Oxford University Press.

Min, Y., Htay, Y. Y., & Oo, K. K. (2020). Comparing the performance of machine learning algorithms for human activities recognition using wisdm dataset. *International Journal of Computer (IJC)*, *38*(1), 61–72.

Mokaya, F., Lucas, R., Noh, H. Y., & Zhang, P. (2016). Burnout: a wearable system for unobtrusive skeletal muscle fatigue estimation. In *2016 15th acm/ieee international conference on information processing in sensor networks (ipsn)* (pp. 1–12).

Moradi, B., Aghapour, M., & Shirbandi, A. (2019). Compare of machine learning and deep learning approaches for human activity recognition. *EasyChair, Tech. Rep.*.

Morgan, P. T., Smeuninx, B., & Breen, L. (2020). Exploring the impact of obesity on

skeletal muscle function in older age. *Frontiers in Nutrition*, *7*, 286.

Mourão-Miranda, J., Hardoon, D. R., Hahn, T., Marquand, A. F., Williams, S. C., Shawe-Taylor, J., & Brammer, M. (2011). Patient classification as an outlier detection problem: an application of the one-class support vector machine. *Neuroimage*, *58*(3), 793–804.

Mueller-Wohlfahrt, H.-W., Haensel, L., Mithoefer, K., Ekstrand, J., English, B., McNally, S., ... Ueblacker, P. (2013). Terminology and classification of muscle injuries in sport: The munich consensus statement. *British Journal of Sports Medicine*, *47*(6), 342–350. doi: 10.1136/bjsports-2012-091448

Nesterenko, S., Domire, Z. J., Morrey, B. F., & Sanchez-Sotelo, J. (2010). Elbow strength and endurance in patients with a ruptured distal biceps tendon. *Journal of shoulder and elbow surgery*, *19*(2), 184–189.

Nguyen, B., Coelho, Y., Bastos, T., & Krishnan, S. (2021). Trends in human activity recognition with focus on machine learning and power requirements. *Machine Learning with Applications*, *5*, 100072. doi: https://doi.org/10.1016/j.mlwa.2021.100072

Nweke, H. F., Teh, Y. W., Al-Garadi, M. A., & Alo, U. R. (2018). Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert Systems with Applications*, *105*, 233–261.

Nweke, H. F., Teh, Y. W., Mujtaba, G., & Al-Garadi, M. A. (2019). Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions. *Information Fusion*, *46*, 147–170.

of Labor Statistics, U. B. (2016). Nonfatal occupational injuries and illnesses requiring days away from work. *Technical Report USDL-16-2130*(4), 1–4.

Opar, D., Williams, M., & Shield, A. (2012, 03). Hamstring strain injuries factors that lead to injury and re-injury. *Sports medicine (Auckland, N.Z.)*, *42*, 209-26. doi: 10.2165/11594800-000000000-00000

Op De Beéck, T., Meert, W., Schütte, K., Vanwanseele, B., & Davis, J. (2018). Fatigue prediction in outdoor runners via machine learning and sensor fusion. In *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining* (pp. 606–615).

Orizio, C., Gobbo, M., Diemont, B., Esposito, F., & Veicsteinas, A. (2003). The surface mechanomyogram as a tool to describe the influence of fatigue on biceps brachii motor unit activation strategy. historical basis and novel evidence. *European journal of applied physiology*, *90*(3-4), 326–336.

Palmius, N., Saunders, K. E., Carr, O., Geddes, J. R., Goodwin, G. M., & De Vos, M. (2018). Group-personalized regression models for predicting mental health scores from objective mobile phone data streams: observational study. *Journal of medical Internet research*, *20*(10), e10194.

Palumbo, F., Gallicchio, C., Pucci, R., & Micheli, A. (2016). Human activity recognition using multisensor data fusion based on reservoir computing. *Journal of Ambient Intelligence and Smart Environments*, *8*(2), 87–107.

Pigou, L., Van Den Oord, A., Dieleman, S., Van Herreweghe, M., & Dambre, J. (2018). Beyond temporal pooling: Recurrence and temporal convolutions for gesture recognition in video. *International Journal of Computer Vision*, *126*(2), 430–439.

Prentice, A. M., & Jebb, S. A. (2001). Beyond body mass index. *Obesity reviews*, *2*(3), 141–147.

Qi, J., Yang, P., Hanneghan, M., Tang, S., & Zhou, B. (2018). A hybrid hierarchical framework for gym physical activity recognition and measurement using wearable sensors. *IEEE Internet of Things Journal*, *6*(2), 1384–1393.

Ramanujam, E., Perumal, T., & Padmavathi, S. (2021). Human activity recognition with smartphone and wearable sensors using deep learning techniques: A review. *IEEE Sensors Journal*.

Ramasamy Ramamurthy, S., & Roy, N. (2018). Recent trends in machine learning for human activity recognition—a survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), e1254.

Reis, V. M., Garrido, N. D., Vianna, J., Sousa, A. C., Alves, J. V., & Marques, M. C. (2017). Energy cost of isolated resistance exercises across low-to high-intensities. *PloS one*, *12*(7), e0181311.

Robergs, R. A., Ghiasvand, F., & Parker, D. (2004). Biochemistry of exercise-induced metabolic acidosis. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *287*(3), R502–R516.

Robson-Ansley, P. J., Gleeson, M., & Ansley, L. (2009). Fatigue management in the preparation of olympic athletes. *Journal of sports sciences*, *27*(13), 1409–1420.

Sadoyama, T., & Miyano, H. (1981). Frequency analysis of surface emg to evaluation of muscle fatigue. *European journal of applied physiology and occupational physiology*, *47*(3), 239–246.

Sala, E., Lopomo, N. F., Tomasi, C., Romagnoli, F., Morotti, A., Apostoli, P., & De Palma, G. (2021). Importance of work-related psychosocial factors in exertion perception using the borg scale among workers subjected to heavy physical work. *Frontiers in Public Health*, *9*, 461.

Sanchez-Medina, L., & González-Badillo, J. J. (2011). Velocity loss as an indicator of neuromuscular fatigue during resistance training. *Medicine and science in sports and exercise*, *43*(9), 1725–1734.

Sant'Ana, M., Li, G., & Zhang, H. (2019). A decentralized sensor fusion approach to human fatigue monitoring in maritime operations. In *2019 ieee 15th international conference on control and automation (icca)* (pp. 1569–1574).

Schäfer, A., & Vagedes, J. (2013). How accurate is pulse rate variability as an estimate of heart rate variability?: A review on studies comparing photoplethysmographic

technology with an electrocardiogram. *International journal of cardiology*, *166*(1), 15–29.

Schrader, L., Vargas Toro, A., Konietzny, S., Rüping, S., Schäpers, B., Steinböck, M., ... Bock, T. (2020). Advanced sensing and human activity recognition in early intervention and rehabilitation of elderly people. *Journal of Population Ageing*, *13*(2), 139–165.

Seshadri, D. R., Li, R. T., Voos, J. E., Rowbottom, J. R., Alfes, C. M., Zorman, C. A., & Drummond, C. K. (2019). Wearable sensors for monitoring the physiological and biochemical profile of the athlete. *NPJ digital medicine*, *2*(1), 1–16.

Shcherbina, A., Mattsson, C. M., Waggott, D., Salisbury, H., Christle, J. W., Hastie, T., ... Ashley, E. A. (2017). Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *Journal of personalized medicine*, *7*(2), 3.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323.

Shirkhorshidi, A. S., Aghabozorgi, S., & Wah, T. Y. (2015). A comparison study on similarity and dissimilarity measures in clustering continuous data. *PloS one*, *10*(12), e0144059.

Shoaib, M., Bosch, S., Incel, O. D., Scholten, H., & Havinga, P. J. (2014). Fusion of smartphone motion sensors for physical activity recognition. *Sensors*, *14*(6), 10146–10176.

Smith, I. C. H., & Newham, D. J. (2007). Fatigue and functional performance of human biceps muscle following concentric or eccentric contractions. *Journal of applied physiology*, *102*(1), 207–213.

Soro, A., Brunner, G., Tanner, S., & Wattenhofer, R. (2019). Recognition and repetition counting for complex physical exercises with deep learning. *Sensors*, *19*(3), 714.

Steffen, L. M., Arnett, D. K., Blackburn, H., Shah, G., Armstrong, C., Luepker, R. V., & Jacobs, J. D. (2006). Population trends in leisure-time physical activity: Minnesota heart survey, 1980-2000. *Medicine and Science in Sports and Exercise*, *38*(10), 1716–1723.

Stoudemire, N. M., Wideman, L., Pass, K. A., Mcginnes, C. L., Gaesser, G. A., & Weltman, A. (1996). The validity of regulating blood lactate concentration during running by ratings of perceived exertion. *Medicine and Science in Sports and Exercise*, *28*(4), 490–495.

Subasi, A., Dammas, D. H., Alghamdi, R. D., Makawi, R. A., Albiety, E. A., Brahimi, T., & Sarirete, A. (2018). Sensor based human activity recognition using adaboost ensemble classifier. *procedia computer science*, *140*, 104–111.

Subasi, A., & Kiymik, M. K. (2010). Muscle fatigue detection in emg using time–frequency methods, ica and neural networks. *Journal of medical systems*, *34*(4), 777–785.

Suto, J., Oniga, S., & Sitar, P. P. (2017). Feature analysis to human activity recognition. *International Journal of Computers Communications & Control*, *12*(1), 116–130.

Sztyler, T., & Stuckenschmidt, H. (2017). Online personalization of cross-subjects based activity recognition models on wearable devices. In *2017 ieee international conference on pervasive computing and communications (percom)* (pp. 180–189).

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and bayesian inference. *Behavioral and brain sciences*, *24*(4), 629–640.

Thalman, C. M., Lam, Q. P., Nguyen, P. H., Sridar, S., & Polygerinos, P. (2018, Oct). A novel soft elbow exosuit to supplement bicep lifting capacity. In *2018 ieee/rsj international conference on intelligent robots and systems (iros)* (p. 6965-6971).

Theofilidis, G., Bogdanis, G. C., Koutedakis, Y., & Karatzaferi, C. (2018). Monitoring exercise-induced muscle fatigue and adaptations: making sense of popular or emerging indices and biomarkers. *Sports*, *6*(4), 153.

Tomlinson, D. J., Erskine, R. M., Morse, C. I., Pappachan, J. M., Sanderson-Gillard, E., & Onambélé-Pearson, G. L. (2021). The combined effects of obesity and ageing on skeletal muscle function and tendon properties in vivo in men. *Endocrine*, *72*(2), 411–422.

Triwiyanto, T., Wahyunggoro, O., Nugroho, H. A., & Herianto, H. (2018). Muscle fatigue compensation of the electromyography signal for elbow joint angle estimation using adaptive feature. *Computers & Electrical Engineering*, *71*, 284–293.

Troiano, R. P., Berrigan, D., Dodd, K. W., Masse, L. C., Tilert, T., & McDowell, M. (2008). Physical activity in the united states measured by accelerometer. *Medicine & Science in Sports & Exercise*, *40*(1), 181–188.

Vanrell, S. R., Milone, D. H., & Rufiner, H. L. (2017). Assessment of homomorphic analysis for human activity recognition from acceleration signals. *IEEE journal of biomedical and health informatics*, *22*(4), 1001–1010.

Wan, S., Qi, L., Xu, X., Tong, C., & Gu, Z. (2020). Deep learning models for real-time human activity recognition with smartphones. *Mobile Networks and Applications*, *25*(2), 743–755.

Wang, S., Tang, H., Wang, B., & Mo, J. (2021). A novel approach to detecting muscle fatigue based on semg by using neural architecture search framework. *IEEE Transactions on Neural Networks and Learning Systems*.

Wang, Y., Cang, S., & Yu, H. (2018). A data fusion-based hybrid sensory system for older people's daily activity and daily routine recognition. *IEEE Sensors Journal*, *18*(16), 6874–6888.

Wang, Y., Cang, S., & Yu, H. (2019). A survey on wearable sensor modality centred human activity recognition in health care. *Expert Systems with Applications*, *137*, 167–190.

Webber, M., & Rojas, R. F. (2021). Human activity recognition with accelerometer and gyroscope: a data fusion approach. *IEEE Sensors Journal*.

Weiss, G. M., & Lockhart, J. (2012). The impact of personalization on smartphone-based activity recognition. In *Workshops at the twenty-sixth aaai conference on artificial intelligence.*

Whittaker, R. L., Sonne, M. W., & Potvin, J. R. (2019). Ratings of perceived fatigue predict fatigue induced declines in muscle strength during tasks with different distributions of effort and recovery. *Journal of Electromyography and Kinesiology*, *47*, 88–95.

Wichit, N., & Choksuriwong, A. (2015). Multi-sensor data fusion model based kalman filter using fuzzy logic for human activity detection. *International Journal of Information and Electronics Engineering*, *5*(6), 450.

Xu, L., Yang, W., Cao, Y., & Li, Q. (2017). Human activity recognition based on random forests. In *2017 13th international conference on natural computation, fuzzy systems and knowledge discovery (icnc-fskd)* (pp. 548–553).

Yoo, S., Ackad, C., Heywood, T., & Kay, J. (2017). Evaluating the actual and perceived exertion provided by virtual reality games. In *Proceedings of the 2017 chi conference extended abstracts on human factors in computing systems* (p. 3050–3057). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3027063.3053203 doi: 10.1145/3027063.3053203

Yu, Y., Li, H., Yang, X., Kong, L., Luo, X., & Wong, A. Y. (2019). An automatic and non-invasive physical fatigue assessment method for construction workers. *Automation in construction*, *103*, 1–12.

Zhang, G., Morin, E., Zhang, Y., & Etemad, S. A. (2018). Non-invasive detection of low-level muscle fatigue using surface emg with wavelet decomposition. In *2018 40th annual international conference of the ieee engineering in medicine and biology society (embc)* (pp. 5648–5651).

Zhang, M., & Sawchuk, A. A. (2011). A feature selection-based framework for human activity recognition using wearable multimodal sensors. In *Bodynets* (pp. 92–98).

Zhou, S., Ogihara, A., Nishimura, S., & Jin, Q. (2017). Analysis of health and physiological index based on sleep and walking steps by wearable devices for the elderly. In *2017 ieee 10th conference on service-oriented computing and applications (soca)* (pp. 245–250).

Zhu, Y., Wang, C., Zhang, J., & Xu, J. (2014). Human activity recognition based on similarity. In *2014 ieee 17th international conference on computational science and engineering* (pp. 1382–1387).