Prediction of Fatigue in Lower Extremity using EMG Sensor and Machine Learning

by

Mahdokht Golmohammadishouraki

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Applied Science

in

Mechanical Engineering

Carleton University
Ottawa, Ontario

Abstract

Predicting the fatigue behavior in muscles significantly improves a stroke patient's road to recovery as it allows minimizing danger while encouraging the patients to participate in the rehabilitation. This essential feature in the recovery process assists physiotherapists with fatigue onset recognition or fatigue level quantification while patients undergo therapy with a rehabilitation robot. In this research, muscle fatigue detection was studied on the squat motion of healthy subjects while using an exergame as the same muscles are utilized for the sit-to-stand motion in stroke patients. First, a procedure was developed, and different analysis methods were investigated to recognize fatigue using processed muscle activity data (MNF and MDF) collected by sEMG as various researches used to recognize fatigue, designed exergame data, and perceived fatigue data collected from participants during the experiment. The sEMG data were collected from four muscles of ten subjects while playing an exergame designed based on squat motion. Among three different methods for analyzing the data (discrete window analysis, moving window analysis, and phasing analysis), the most accurate analysis method was developed by analyzing each phase of the squats separately.

Furthermore, there was a need to develop a method to predict both fatigue onset and fatigue level by using all the muscle data and perceived data. The appropriate muscle activity values were extracted from the data and combined for further classification and regression analysis. Six machine learning classifiers with two different implementation strategies (5-fold cross-validation and leave-one-participant-out cross-validation) were utilized to find the best model for detecting the onset of fatigue, whereas five machine learning regressors

were used for finding the best regression analyzer in predicting the fatigue level. The classifiers and regressors were evaluated by seven and five different algorithm validation respectively to not only evaluate the accuracy of the method, but also consider the reliability of the performance to make sure that patients are not facing any danger during the recovery process. In both scenarios, Random Forest performed best.

Acknowledgements

I would like to extend my sincere gratitude to my supervisor, Prof. Mojtaba Ahmadi, who have given me this opportunity and always supported me during this research. I am grateful to my colleagues at the Advanced Biomechatronics Locomotion Lab (ABL) who have always been supportive and friendly. I would like to especially thank Nick Berezny, who introduced me to the research and was always there to help.

I would like to take this opportunity to express my gratitude to my family and friends for their unconditional love and support. Shahryar, thanks a lot for the encouragements and kind words when things were tough.

I do not know how to say thank you to my loving parents who helped me to become a better person and create a better life. Finally, my dearest brother, Mahdi, I love you and I can not express how thankful I am for having you in my life.

Table of Contents

List of Tables

List of Figure

This is the list of figures.

Nomenclature

Name	Acronym
5-Fold Cross-Validation	5F-CV
Bicep Femoris	BF
Central Nervous System	CNS
Classical Machine learning	C-ML
Decision Tree	DT
Degree of Freedom	DoF
Electromyogram	EMG
False Negative	FN
False Positive	FP
Frequency Domain	FD
Gastrocnemius Lateralis	GL
Gastrocnemius Medialis	GM
Gaussian Naive Bayes	GNB
Histogram	HIST
Human Robot Interaction	HRI
K-Nearest Neighbors	KNN
Leave-One-Participant-Out Cross-Validation	LOPO-CV
Linear Discriminant Analysis	LDA
Machine Learning	ML
Mean Absolute Error	MAE
Mean Absolute Value	MAV
Mean Squared Error	MSE
Motor Units	MU
Neural Network	NN
Quadratic Discriminant Analysis	QDA
Random Forest	RF
Receiver Operating Characteristic	ROC
Root Mean Square	RMS
Root Mean Square Error	RMSE
Support Vector Machine	SVM
Surface Electromyogram	sEMG
Tibialis Anterior	TA
Time Domain	TD
True Negative	TN
True Positive	TP
True Positive Rate	TPR
False Positive Rate	FPR
Vastus Lateralis	VL
Vastus Medialis	VM
Waveform Length	WL

Chapter 1

Introduction

1.1 Motivation

Each year approximately 62,000 Canadians and 795,000 Americans are annually affected by strokes which is a cerebrovascular disease [1], [2]. It can lead to adult disability, body weakness, muscle spasticity, limited mobility, and muscle control, affecting patients' daily life activity [3], [4]. It requires medical intervention and rehabilitation to reduce performance impairment. Rehabilitation utilizes new neural connections in the brain, which allow the patient to re-learn tasks through repetitive exercise [5].

Occupational therapists and physiotherapists use a broad range of therapeutic exercises in the stroke rehabilitation process, mostly focused on motor dexterity and motor movement involving the repetition of the target movement. Over the past decades, various rehabilitation robotics with different target exercises were developed to be used by occupational therapists and physiotherapists to improve and increase the speed of stroke patients' rehabilitation process. Using these robots has many advantages, such as increasing the intensity and duration of the therapy session since they are not relying on the therapist's assistance and can be used until the patient is tired. Also, unlike traditional therapy sessions, these rehabilitation robots can provide some feedback information like trajectory error, force, and effort trends. Tracking these parameters can

help the therapist during the recovery period. Mental tiredness of patients is another factor which decreases the therapy session duration. Virtual environment is another feature added to some of these rehabilitation robots to increase the encouragement in using the robot by patients.

1.2 Project Background

Advanced Biomechatronics and Locomotion Laboratory (ABL) at Carleton University had developed a 4-DOF end plate-based rehabilitation robot named ViGRR (Virtual Gait Rehabilitation Robot), which assists the patient to move through sagittal plane trajectories. The robot has an assistive feature which is controlled by an admittance controller that helps the patient in case of trajectory error. It consists of a footplate that interfaces with the user's foot which is magnetically attached to the robot, as shown in Figure 1. The magnet can be released in case of high force, high velocity, or activation of the emergency button. The robot also has a virtual environment to increase user engagement, including different games with desired trajectories and feedback [6], [7].

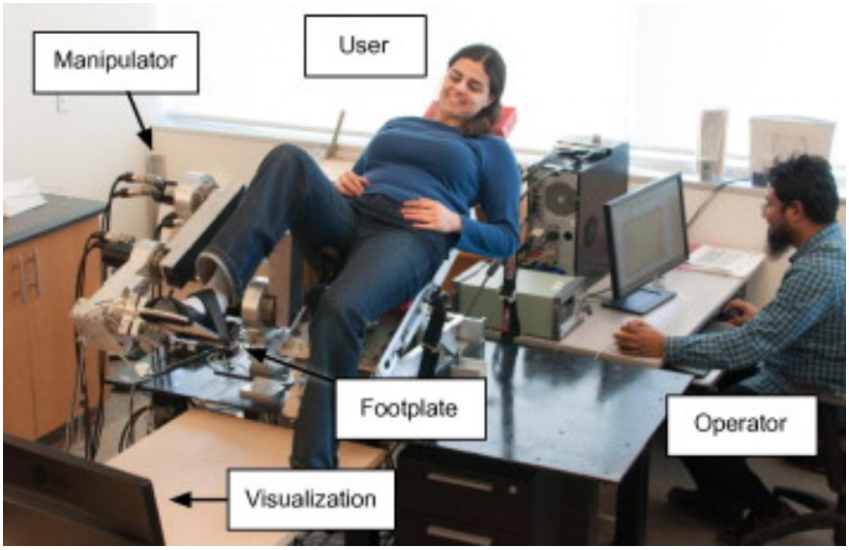


Figure 1: The original ViGRR that was developed at Carleton University [6]

ViGRR was designed with the goal of lower-limb rehabilitation and compatibility with a patient who is reclined in a hospital bed. However, ViGRR has some aspects which prevent it from being used easily as it is a heavy robot and cannot easily be transported to different locations. Therefore, another robot has been developed at ABL named ViGRR-Lite with the goal of increasing the functional patient's recovery outcomes for therapists in hospitals. It utilized the primary components of ViGRR and repackaged it to a lighter bed-bound rehabilitation robot that is easier to install. ViGRR-Lite simulates sit-to-stand motion by a leg extension exercise, and similar to ViGRR, it engages the patient using games with desired trajectories [7]. Figure 2 represents the ViGRR-Lite robot.

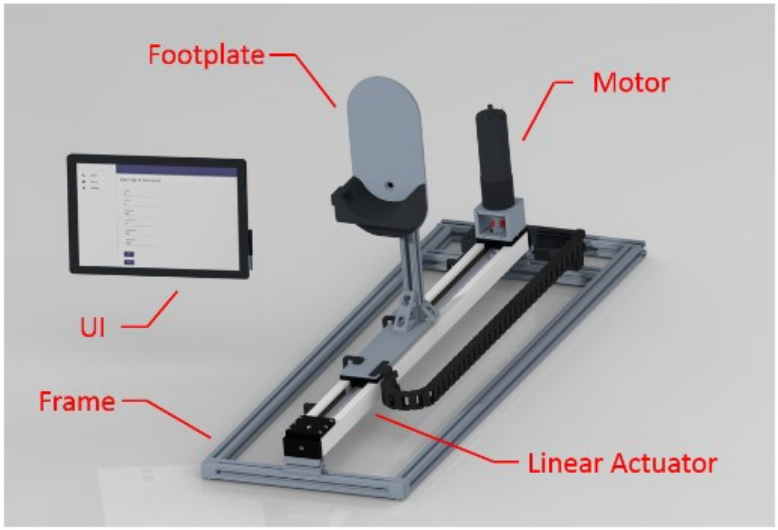


Figure 2: The schematic of the ViGRR-Lite robot developed at Carleton University [7]

1.3 Thesis Objective

Both robots discussed in 1.1 assist the patient to move their foot in the desired trajectory to utilize their muscles and try to re-learn the motions. Games have been used for engaging the patient and increasing the exercise duration. There are some feedbacks provided by the robots which can help

the therapist to measure the improvement of the muscles during the recovery journey. During the period of the exercise, physical fatigue or potential boredom emerges as the patients continue to play the game.

Fatigue prediction can be very beneficial for rehabilitation process as per the following reasons:

- Increasing the duration of exercise is one of the most important goals of rehabilitation robotics. Earlier prediction of fatigue onset, and even quantifying the fatigue level could help the therapists to prevent fatigue in patients. This can be done by increasing the resting period, decreasing the intensity of the exercise, increasing the assistance level of the controller, etc.
- Disregarding the psychological fatigue may be harmful; the patient would not correctly perform the exercise rendering the therapy ineffective.
- Physical tardiness due to excessive exercise could potentially be harmful to an already damaged muscle.

Therefore, there is a need to recognize fatigue while the patient is performing a targeted exercise with a rehabilitation robot.

The original goal of this thesis was to utilize Electromyography sensors to track muscle activity and recognize associated fatigue, in addition to analyzing perceived fatigue data to recognize mental tiredness while the patient is using the ViGRR-Lite.

During the COVID-19 pandemic and restricted physical distancing period, it would be impossible to undertake any experiments involving human-to-human physical contact, especially with vulnerable stroke patients. Therefore, a modified experiment using an exergame was conducted remotely without compromising safety and aimed to replicate the original motion of the robot. The robotic system was replaced with a reward-based game operated by healthy subjects performing

basic squats to gain the most points. For this, my colleague, Nick Berezny, had developed a vision-based body tracking system which only requires setting up a webcam and affixing printed markers to the participant's joints. The main objective of this thesis is to prepare the means of detecting and quantifying muscle fatigue (based on perceived fatigue) during an exergame focused on squat motion.

1.3.1 Thesis Contributions

The following contributions assisted in reaching the objectives:

1. Developing the best scenario for a new experiment on healthy subjects to simulate the patient's real-life circumstances while adhering to public health guidelines.
2. Collecting three different sets of data to consider participants' fatigue: EMG sensors (muscle activity data), game data, and perceived data.
3. Comparing and identifying a minimal adequate set of muscle groups for EMG collection and finding a suitable analysis method to identify the most appropriate muscle activity analysis by focusing on the timing or phase of each squat in order to synchronize the game data and the perceived fatigue data.
4. Identifying muscle fatigue from EMG data, comparing fatigue behavior on different muscles, and correlating it with other variables (Gender, age, BMI, game score, etc.) were among the tasks performed for understanding the data and in preparation of the next step.
5. Utilizing machine learning classifiers and regressors to find the best model to detect the fatigue onset and quantify the fatigue level, respectively, based on the perceived data considering different features such as physical tiredness, BMI, age, etc.

6. The outcome will provide the necessary preliminary information needed to set up the ViGRR-Lite to conduct robotic haptic gaming with healthy and ultimately target population participants.

1.4 Thesis Outline

- **Chapter 1: Introduction**

  The current chapter aims to introduce the research goals and provide the objectives of the thesis.

- **Chapter 2: Literature Review**

  This chapter represents a review of fatigue recognition research in muscles, Electromyography, and provides an overview of different machine learning definitions and training algorithms implemented in this research.

- **Chapter 3: Fatigue Recognition Through Collected Game, EMG, and Perceived Data**

  This chapter explains the experimental trial, the process of choosing the desired muscles, the development of different analysis methods, and how their results led to the primary trial. Muscle selection and analysis method, showing the results of EMG collected data, game data, and perceived data.

- **Chapter 4: Muscle Fatigue Predictions using Machine Learning**

  This chapter covers the implementation of different machine learning algorithms to recognize the onset fatigue and the fatigue level. It includes the process of preparing various input features for the analysis and provides the best parameters for the classification and regression based on the results.

- **Chapter 5: Conclusion and Future Work**

  This chapter includes the concluding remarks along with the recommendations for future work.

Chapter 2

Literature Review

2.1 Muscles and Associated Causes of Fatigue

2.1.1 Muscle Fatigue

Many people experience **Fatigue** as a common symptom in a wide range of daily situations. It can be felt as a significant tiredness, lack of energy, or feeling exhausted, and relates to having difficulty in performing voluntary tasks [8], [9]. Fatigue can be classified based on various measures as mental fatigue or physical fatigue, acute fatigue or chronic fatigue, etc. [10]–[12]. One of the most important types of fatigue is muscle fatigue which in [13] is defined as a decrease in power production in case of responding to contractile activity, or as in [14] is defined as a decline in the performance of the intensive activity of the muscle.

Muscle fatigue is a broad phenomenon which can be referred to as a motor deficit, decline in mental function, or a perception and it can describe the sustained activity end point or the decrease in muscle force capacity [15]. This broadness is problematic in identifying the causes of muscle fatigue; therefore, a more focused definition used by most of the investigators can be utilized. It is defined as an exercise-induced reduction in muscle ability to produce force or power for sustained or intermittent tasks [16], [17]. Moreover, muscle fatigue is not the exhaustion of a muscle or

failing to do a task; it is the decrease of maximal force or power which can be produced by the muscle [15].

Fatigue in muscle occurs since the process of generating a force by contractile proteins is enabled by several physiological impaired processes, or in other words, it is proceeded from central or peripheral motor pathway levels [18], [19].

2.1.2 The Mechanism of fatigue in muscles

As previously discussed in 2.1.1, fatigue is a common feeling in daily life, and it has been under evaluation over decades in different applications and clinical projects [20]. Different studies reveal various views of muscle fatigue mechanism as it is dependent on features like exercise models and the methods used. Therefore, studying muscle fatigue is challenging as the data from different studies cannot be transferred, there is no certain result to rely on, or the results are based on voluntary exercises that cannot be carried out on daily life [20].

For any exercise, a voluntary force is generated from a sequence of events. First, motor-neurons are activated with motivational factors and sensory information integration. The signal transfer between motor-neurons and neuro-muscular junction occurs at this point. This transfer of signal in a healthy body should be adequate [21], [22]. Different fatigue protocols with different metabolic, ionic, and electrical changes have different mechanisms, but mostly tiredness in muscles is caused by changing in balance of different ions like Na^+ and K^+ over the sarcolemma [20].

It should be noted that unlike studies on animal muscles, studies on human muscles have various limitations, and need lots of considerations which render the study of muscle fatigue mechanism more difficult.

2.2 Electromyography

Electromyography or EMG system is used to track the muscle health and the motor neurons which control the muscle's activity by measuring the response of the electrical activity of the muscle nerves to stimulation [23], [24]. An electromyograph is used to detect the electrical potential generated from muscles when they are neurologically active [25]. EMG is used in different areas to find the muscle disorders like muscle weakness, muscular dystrophy, radiculopathies, muscle fatigue, etc. [23].

2.2.1 EMG Types

There are two types of EMG: surface EMG and Needle EMG. Each of these are used for different purposes.

Surface EMG is used in different situations as used by physiotherapists, kinesiologists, biomedical engineers, and experimental researchers. It uses surface electrodes to collect muscle activity data from the surface of skin. As shown on Figure 3, it consists of an amplifier, sensors, and input module [26].

Figure 3: Delsys Surface EMG set-up [26]

The collected data is the voltage difference between two separate electrodes. The potentials are measured with respect to a reference voltage generated from a sensor on a bone.

Figure 4: Surface EMG Sensor and the operating schematic [26]

As it can be seen in Figure 4, the voltage is measured by subtracting V_2 from V_1. Surface EMG is used in this research.

Needle electromyography is another popular EMG system which is used for detecting illnesses such as mimicking diseases and localized abnormality of muscles. Needle EMG contains fine wires which are inserted in the target muscle and collect muscle activity signals [27]. There is a need for a reference electrode which can be a surface electrode or an inserted fine wire. The fine wire has an exposed tip which serve as an active electrode. The electrical activity of the muscle can be seen on an oscilloscope [24].

These two electromyography types are vastly used in various medical and engineering fields, and they have been certainly helpful in detecting diseases and helping researchers to have a better understanding of muscle activities and abnormalities to make improvement in areas such as rehabilitation robotics. Each of these systems have their risks which limit their adoption for different applications. As the Needle EMG utilizes needles to insert the electrodes in the muscle, it causes pain and discomfort for the participants [27], but the surface EMG does not cause any pain during the attachment of the electrodes to the subjects' body. There is a minor risk of discomfort or skin irritation when adhesive tapes are taken off the skin at the end of the experiment with surface EMG.

Also, electrodes used in surface EMG system are not active and only use ± 5 v DC power supply [26] resulting in experiments with no electric shocks. However, Needle EMG systems are using active electrodes which require professional supervision for ensuring that the electrodes are placed exactly on the right position in muscles in order to collect the correct muscle activity signals while minimizing the skin discomfort and bleeding for the participants [28]. On the other hand, using surface EMG requires a lot of attention to find the right location on skin to attach electrodes and collect the desired data; electrodes are easier to attach and does not cause a lot of annoyance for

subjects. Surface EMG is more preferrable these days to collect intensity of muscle activation [29]. Therefore, the surface EMG system has been used in this research.

2.3 EMG Data and Data Analysis

For understanding the different behaviors in human body and detecting abnormalities, data processing is the primary step after recording EMG data that is prone to collecting lots of noises [30]. The raw EMG data is affected by noises induced by incorrect electrode location, incorrect selection of signal channel, inherent electrode noises [31], movement of connection cables, etc.[32]. Therefore, there is a need for preprocessing the raw data before using them. The first step is filtering on the frequency spectrum of the signal. The most suitable filters for EMG signals are low-pass, high-pass, and band-pass filters [33] which have been used in this work. Furthermore, there is a need to extract some reliable features from the signal to have a better observation and analyzation from the informative data features. Features can be extracted from both the time domain (such as Mean Absolute Value (MAV) and Waveform length (WL)) and the frequency domain (such as Mean power frequency (MNF) and Median power frequency (MDF)) [34].

- **Mean Absolute Value (MAV):** For each window, MAV represents the mean of a fully rectified signal.

$$MAV(x_i) = \frac{1}{L}\sum_{k=1}^{L}|x_{i,k}|$$

where $x_{i,k}$ represents k^{th} sample data in i^{th} segment and L denotes the length of the EMG signal.

- **Waveform length (WL):** It shows a simple signal waveform characteristic which can be calculated as below:

$$WL(x_i) = \sum_{k=1}^{L} \left| x_{i,k} - x_{i,k-1} \right|$$

where $x_{i,k}$ represents k^{th} sample data in i^{th} segment, and L denotes to the length of the EMG signal.

- **Mean Power Frequency (MNF):** It represents the average frequency calculated as below:

  $MNF = \sum_{j=1}^{M} f_j P_j \,/\, \sum_{j=1}^{M} P_j$

  Where f_j is the EMG power spectrum frequency value at the frequency bin j, P_j is the EMG power spectrum at the frequency bin j, and M is the frequency bin length [35].

- **Median Power Frequency (MDF):** the power spectrum will be divided into two regions with equal amplitude in this frequency as below:

$$\sum_{j=1}^{MDF} P_j = \sum_{j=MDF}^{M} P_j$$

  Where P_j is the power spectrum of EMG at the frequency bin j, and M is the length of the frequency bin [35].

2.4 EMG and Fatigue

In most of the muscle fatigue research, the observation of neuromuscular responses and evaluation of the motor unit activations have utilized electromyography [36], [37]. The EMG signals can provide helpful information about motor unit activation and muscular responses [38]. However,

there is a need for feature extraction to make sense of the data and draw meaningful conclusions. The amplitude of the EMG signals represents an increase, and there is a decrease in the spectral frequency of the signals when it comes to muscle fatigue studies [39].

There are various studies that aim to recognize muscle fatigue. They mostly used EMG signals to observe muscle activity and consider the slope of features such as MNF and MDF of signals as a mean of recognizing fatigue in muscles. These studies used different exercises to collect data from one muscle. In [40], muscle activity of the Rectus Femoris is collected using EMG while doing a knee extension exercise. Integrated EMG (IEMG), Average Motor Unit Potential (AMUP) and Mean Power Frequency (MNF) were calculated. IEMG results indicated a slight and continuous rise during the fatigue period, whereas AMUP showed an increase in the signal's amplitude. The MNF represented a slight decrease during fatigue time. Rahnama et al. [41] used a soccer stimulation fatiguing protocol to investigate the lower limb muscle activity. The results indicate a decrease in EMG activity when the participant is fatigued while using the stimulation. Bicep concentration curl exercise are used in [42] to observe the effect of muscle fatigue on Human Activity Recognition systems. Based on their collected data from 20 participants, fatigue has occurred in later sets of the exercise, and muscle endurance decreased. They also find that some changes happen in data during the fatigue period. Chang et al. [43] used a wireless surface EMG recording system to monitor Vastus Lateralis fatigue while the participant used a pedaled-multifunctional elliptical trainer. A decrease in MDF is seen in signal processing as the muscle exercise time increases. In some studies, an EMG-based exergame is developed to model the exercise in addition to engaging the participant during the experiment [44], [45]. In [44], an EMG-based system is developed in a customized isometric exergame. The Flexor digitorum superficialis muscle has been examined to monitor the performance of participants. A swimming exergame is

developed in [45], to investigate the upper limb muscle activities. There are a limited numbers of studies which used machine learning methods to detect muscle fatigue, as in [46], the Support Vector Machine (SVM) was used to recognize the fatigue and non-fatigue of lower limb muscle during isokinetic exertions during the target exercise. Finally, Papakostas et al. [47] investigated upper limb muscle fatigue in exercises like shoulder flexion and shoulder abduction. They used a machine learning approach which considered the muscle activity in addition to user subjective report. To the best of the author's knowledge, this research was the only one which considered the perceived fatigue of participants as an input feature of their machine learning models.

Despite different studies in monitoring muscle fatigue, there is still a lack of research in some areas. First, most of the studies in this area monitor the upper limb muscles during exercises, and there is not enough research on lower limb muscles. Second, there are a few studies which are using an exergame to detect muscle fatigue. Another important feature that has not been investigated enough is considering the subjective user report during the experiment which can be a reflection of potential boredom or mental fatigue. Fourth, most studies in this area are about detecting the onset of muscle fatigue, but there is a need to investigate fatigue level recognition in muscles. Finally, in most of the studies, only one muscle was observed while the participant was performing an exercise, but it is known that in every exercise, there is more than one muscle involved. To better understand the participant fatigue behavior, it is better to study at least the main muscles active during the desired exercise. This research aims to investigate these gaps.

2.5 Machine Learning

Machine learning is a type of AI (Artificial Intelligence) that utilizes different algorithms to let the software learn from a dataset and train itself to predict the relationship between the data and the

desired target [48]. Tom Mitchell's definition of Machine Learning is “a computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E” [49]. A daily example of machine learning utilization is in the email spam detection system. At first, some of the emails are flagged as spam emails by the user, and they have been used by machine learning as the training examples, which are called **Training instances or Samples**. In this scenario, flagging spam for new emails are called task or T, the training experience or E is the database of the emails flagging as spam and the performance measure or P is the proportion of correctly selected emails as spam by the software [33],[50]. As each of these three processes has an important effect on the performance of the machine learning system, they should be defined carefully. Using Machine Learning is great for different scenarios such as getting an insight into complex problems with big data sets, finding the solutions for complex problems which cannot be done by the traditional approaches, etc. [48], [50], [51]. There are many different types of Machine Learning systems that can be categorized based on different criteria. The procedure to train the data is a defining parameter for the type of machine learning, Supervised and Unsupervised learning are the main categories that will be discussed next.

2.5.1 Supervised Learning

Supervised learning is an approach which trains the algorithm based on a group of labeled data that is called training sets. These training sets include inputs and desired outputs and help to teach the model to evaluate desired results [48], [52]–[54]. Classification and Regression are the two types of supervised learning algorithms.

2.5.1.1 Classifications and Regression

The classification method is training the algorithm task to classify the data into k categories where $\{k \in \mathbb{N} \mid k > 1\}$. In most cases, k=2 which is called binary classifications. The algorithm function is: $f: \mathbb{R}^n \rightarrow \{1, \ldots, k\}$. In other words, the algorithm is able to predict that the given vector $\boldsymbol{x} \in \mathbb{R}^n$ should be grouped with which categories of k [55]. In the second part of this research, the goal is to define if the subject is fatigued or not. In this situation, the binary classification (k=2) is used as two categories of being fatigued and not fatigued [48], [54], [56].

The regression method enables an algorithm to predict a target numeric value. To train such a system many input features with the numerical value are required [48]. Note that some classification algorithms can be used as regressors and vice versa. Here is the list of classification and regression tools used in this research.

Logistic Regression

Logistic regression is a model which is mainly used in solving binary problems and categorizing the output in two different groups such as “True” and “False”, or “Yes” and “No” [57]–[59]. The logistic regression model is defined as:

$$f_{w,b}(x) = \frac{1}{1 + e^{-(wx+b)}}$$

Here, x is the feature vector representing an instance, w represents a D-dimensional vector of parameters (weights), and b is a real number. This equation is called logistic function, which has an output range of 0 to 1. The model predicts an output based on the proximity of the output to 1 [48], [57].

Gaussian Classifier

Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) are two different types of the Gaussian Classifier. The LDA concept is finding a linear combination of predictors that separates classes in the best way. It uses a score function to estimate the linear coefficients that maximize the score [60]. QDA is a general discriminant function with a nonlinear separation boundary used for classifying the data in two or more classes [60].

For the instance *x,* the probability can be written as a conditional probability as follows:

$$P(y = k \mid x) = \frac{P(x|\, y = k)\, P(y = k)}{P(x)} = \frac{P(x|\, y = k)\, P(y = k)}{P(x|y = 1)\, P(y = 1) + P(x|y = 0)\, P(y = 0)}$$

(The instance *x* has label $y = k$ where $k \in \{0,1\}$).

By replacing $P(x|\, y = k)$ with $f_k(x)$, and $P(y = k)$ with π_k, the shortened form of the equation is achieved. Maximizing $P(y = k \mid x)$ by any value of *k*, is considered the label for *x,* which yields:

$$h(x) = argmax_k \frac{f_k(x)\, \pi_k}{\sum_y f_y(x)\, \pi_y}$$

The *argmax* operation refers to maximizing $h(x)$ based on the value of *k*. Please note that the denominator is independent of *k,* and constant. By simplifying the expression independent of *k* and assuming $f_k(x)$ to multivariate Gaussian model and $h(x)$ can be created as a classifier [61]:

$$h(x) = argmax_k \delta_k(x)$$

where,

$$\delta_k(x) = -\frac{1}{2}\log|\Sigma_k| - \frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1}(x - \mu_k) + \log \pi_k$$

In the equation above, Σ_k is the training data covariance matrix in class k, and $|\Sigma_k|$ indicates its determinant. This classifier presents the Quadratic Discriminant Analysis (QDA), which has a quadratic decision function. μ_k is the mean of training examples vector in class k, Σ_k is the training example covariance of class k, and π_k is the training data proportion in class k in comparison with the total number of data.

If the covariance matrices are assumed to be equal between two classes, the $\delta_k(x)$ can be simplified as:

$$\delta_k(x) = x^T \Sigma^{-1} \mu_k - \frac{1}{2} {\mu_k}^T \Sigma^{-1} + \log \pi_k$$

where,

$$\Sigma_0 = \Sigma_1 = \Sigma$$

It has a linear decision boundary and is called Linear Discriminant Analysis (LDA).

Decision Tree Classifier (DT)

This classifier utilizes recursive partitioning to classify data into a tree structure form and is built from a single root node. For each step, instances with similar classes will be partitioned into the same subsets. The main advantages of DTs are high accuracy, easy interpretation, capability of determining the best and worst attributions, along with estimating the expected value for various attributes.

By dividing the feature space in each dataset into m non-overlapping subsets $R_1, R_2, \ldots\ R_m$, a function can be defined to predict instance x label as per the following:

$$f(x) = \sum_{i=1}^{m} c_m \, I\{x \in R_m\}$$

I is an identified function which returns 1 and 0 when $x \in R_m$ and $x \notin R_m$, subsequently. c_m is a constant determined from the majority of the labels in subset m. Therefore, the label for each new instance is determined from the subset it falls into. Although there are different strategies to divide the feature space into m portions, they all try to minimize the impurity measure (p_{mk}) in them. If there are N_m observations in subset R_m, the observation proportion of class k in a R_m subset can be written as[62]:

$$p_{mk} = \frac{1}{N_m} \sum_{x \in R_m} I\,(y = k)$$

The Gini index and Cross-entropy are two ways of calculating minimum impurity in partitioning the feature space into m subsets. They are calculated as follows:

$$Gini\ index = \sum_{k=1}^{K} p_{mk}(1 - p_{mk})$$

$$Cross - entropy = -\sum_{k=1}^{K} p_{mk} \log p_{mk}$$

Useful partitions can be created for classification by minimizing these two equations [48], [63]–[65].

Random Forest (RF)

Decision Trees have high variance and are noisy. Therefore, building a collection of uncorrelated decision tree classifiers and then averaging them improves the classifiers and reduces the variance; a RF classifier is an average of some decision tree classifiers [48], [62], [63], [65]–[67].

Support Vector Machine Classifier (SVM)

The Support Vector Machine (SVM) classifier aims to find a decision function in the hyperplane or line form that separates two classes and stays far away from the closest training examples. SVM's objective is to maximize the margin (the distance between the decision boundaries and the closest instance is called the margin). As the real-world data is noisy and has some overlap instances from each class, finding such a line or hyperplane is not always possible. Therefore, finding a good balance between the margin width and the number of instances violating the margin constraint is the operating tool for SVM.

The equation $wx + b = 0$ can be defined as a decision function: *x* is the multidimensional training data, *w* is the real-value vector of weights with the same dimension as *x*, and *b* is a value referred to as bias. $||w|| = \sqrt{\sum_{j=1}^{D}(w^{(j)})^2}$ is defined as the decision tree slope, which is inversely proportional to the margin. Therefore, the objective of SVM is to minimize $||w||$ to get larger margins. SVM objective can also be formulated as follows given that $t^{(i)}$ is -1 for negative instances and +1 for positive instances:

$$\text{Minimize } \big|\big|w\big|\big| + C \sum_{i=1}^{m} \zeta^{(i)}$$

$$\text{Subject to } t^{(i)}(wx + b) \geq 1 - \zeta^{(i)} \ \ and \ \zeta^{(i)} \geq 0 \ \text{ for } \ i = 1,2,3, \dots, m$$

$\zeta^{(i)}$ is a slack variable which measures how much each instance is violating the margin [48], [63], [65], [67].

Neural Network

"Artificial Neural Network or ANN is a machine learning algorithm inspired by the networks of biological neurons found in our brains." Written in [48]. ANN is powerful, scalable, and capable

of using for highly complex Machine Learning tasks. It is composed of an input layer, hidden layers, and an output layer, as shown in Figure 5:

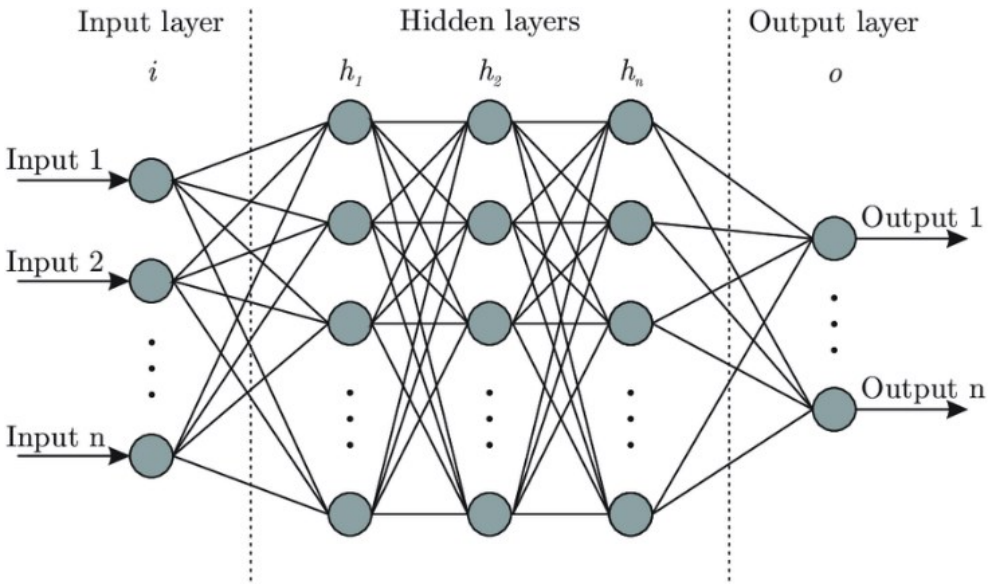


Figure 5: Neural Network Layers [68]

And here is the equation for a fully connected layer:

$$h_{w,b}(x) = \emptyset(xw + b)$$

Here, x is representing the input features, w shows the connection weights, b contains the bias vector. Last, the $\emptyset$ is called activation function, which depends on the application and the context of the study [48]. The input information is passed from the input layer to the hidden layers. In each hidden layer, the output is calculated by the algorithm for each of the neurons in this layer, and then the results will pass to the next layer and so on. When the last hidden layer calculates its result, it will go through the output layer. This is called the **forward pass**. Then, by using different functions, the algorithm measures the output error and calculates how much each connection output contributes to this error by applying the chain rule. For each hidden layer, the algorithm again

calculates how much of this error came from the previous layer working backward until it gets to the input layer. This backward movement helps measure the error across all connection weights backward through the entire network; this is called the **reverse pass.** The algorithm for each training instance uses a forward pass for making the first predictions and uses the reverse pass to measure the error associated with each layer, and finally changes the connection weights to reduce the error [48], [69]–[71].

Linear Regression

The linear regression model is the most common regressor which is used in Machine Learning applications. This model predicts based on computing input features weighted sum and the bias term:

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

Which y is the prediction value, n is the features number, x_i is the i-th feature value and θ_j is the j-th model parameter. The linear regression model will first fit the training data and evaluate the performance. Then, based on the selected performance measure, tries to come up with the θ value, which minimizes the error [48], [72].

K-Nearest Neighbors (KNN)

K-Nearest Neighbors is a nonparametric supervised learning method which classifies the training set data based on the nearest neighbor's majority votes. In other words, by using distance function measurement, a case is assigned to the class of K nearest neighbors for an instance in time which

is their identifications generated about them [73]. For instance, in Figure 6, K is chosen to be equal to 15 which means the predicted class is chosen by the majority votes of the 15 nearest neighbors.

Figure 6: 15-Nearest-Neighbors

Gaussian Naive Bayes (GNB)

Gaussian Naïve Bayes is a specific type of Naïve Bayes which is a classification method based on Bayes theorem [74]. By assigning the probability that its label is y=C using the Bayes Rule:

$$P(y = C\ |x) = \frac{P(x|\ y = C)\ P(y = C)}{P(x)}$$

Which can be rearranged for $y = 1$ and $y = 0$ as follows:

$$f_b(x) = \frac{P(y = 1 \mid x)}{P(y = 0 \mid x)} = \frac{P(x \mid y = 1)\ P(y = 1)}{P(x \mid y = 0)\ P(y = 0)}$$

The relative probability of x belonging to each class is computed by $f_b(x)$ which is called the Bayesian classifier.

2.5.2 Unsupervised Learning

There are many applications that use Supervised Learning these days, but there are many applications without any available labeled data. This kind of application requires to use of unsupervised learning to train their algorithm, which uses the data and labels them innately. The most common unsupervised learning tasks are Clustering, Anomaly Detection, and Density Estimation. As unsupervised learning is not used in this research, it will not be discussed further.

2.5.3 Main challenges of Machine Learning

Although there are so many advantages of using machine learning in technology and sciences, it has its own challenges like any other application. First, a large amount of data is needed for most machine learning algorithms to work appropriately [48], [75]. In [76], a notion was explained that the data is more important than the algorithm; however, data collection is not feasible in all cases, and the algorithms should still be utilized.

Also, there is a need for generalized training dataset in order to have a classifier which expands over the entire range of inputs. The model would not be predicting accurately if it was trained by a nonrepresentative training set [48].

Accurate prediction by classifiers requires organized and consistent training data sets. Errors, outliers, and noises mislead the system toward detecting useful patterns, and the classifier is unlikely to perform as accurate as possible. Therefore, there is a need to spend a reasonable amount of time to clean up the training data set [48], [77], [78].

Another critical step in using machine learning in a project and expecting a successful result is to ensure that the training data contains relevant features as opposed to irrelevant features. This can help the classifier to come up with more powerful correlations between the input features. This is called feature engineering, which has three steps [48], [77], [79]:

1. Feature Selection: The first step is to come up with the most valuable features for training among all the features.
2. Feature Extraction: Many useful and crucial features can be produced by combining existing features. This can expand the number of input features and will help to have a more powerful classifier.
3. Creating new features by gathering new data: Although, as explained before, in some cases, collecting new data is time-consuming and expensive, but if it is possible, it can help to create more input features for the classifier.

On the other hand, there are many traps that the classifier may fall into, such as overfitting the training data or underfitting it. Overfitting in machine learning refers to the lack of further applicability of the model even though it is performing well on the training data. When the training set is noisy or too small, some complex models are likely to detect patterns in the noise in addition to the patterns in the actual data. This can be solved by selecting a fewer number of parameters, collecting more training data set, or spending some time to reduce the noise in the training dataset [48]. Underfitting is the opposite of overfitting, which happens when the model is too simple to learn the data structure and determine the pattern between the input features. This problem can be solved by empowering the model by having more parameters, selecting better features for the algorithm, or reducing the constraints on the model [48].

As explained, preprocessing the data is a critical step to filter the training data from errors and outliners, but this process may result in missing some of the data [73], which is an important challenge for machine learning algorithms [80]. There are different solutions for this problem but by replacing the missing values, the original dataset is affected. In [81]–[84], some practical solutions are explained for this challenge. Even though there are further issues with machine learning, they exceed the scope of this research.

2.5.4 Feature Scaling

There is an essential need for scaling the input features as most of the algorithms cannot perform well when the input features have a different scale. There are different methods for dealing with this issue; however, the two most common methods are Min-Max Scaling and Standardization [48].

Min-Max Scaling is the most common feature scaling way that rescales values to 0 to 1. It is the ratio of the difference between the parameter (X) and the smallest feature (X_{min}), over the range of the data set ($X_{max} - X_{min}$) as shown in the following equation [48], [85]:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Standardization is also used in different applications. It subtracts the mean value from the data point and then divides it by the standard deviation. Therefore, the resulting distribution has a unit variance. The standardization scaling does not scale the data to a specific range, unlike the Min-Max Scaling. The equation for standardization is provided below [48], [86]:

$$X_{standardized} = \frac{X - \bar{X}}{S}$$

Where, $\bar{X}$ is the mean of all the data points, S is the standard deviation of all the datapoints, X is the original datapoint that $X_{standardized}$ will represent in the new dataset.

2.5.5 Model Tunning

Each machine learning algorithm has parameters that cannot be calculated from the data, and they are external to the model. These parameters are called hyperparameters [87]. There is a need to choose the best hyperparameters values for optimizing the prediction result of the algorithms. One way is to iterate them manually until the best combination is found, but it is time-consuming [48], [88]. **Grid search** and **Randomized search** are two procedures undertaken to choose the best values for the hyperparameters. **Grid search** uses cross-validation to determine the best values of selected hyperparameters to reach the optimized result of the model, but when the hyperparameter search area is large, it is better to use a **Randomized search.** Randomized search can be used in the same applications as Grid search, but it will evaluate a given number of random combinations instead of all the possible combinations of hyperparameters [48], [88], [89]. Based on the application, using each of these procedures will help to improve the prediction of the algorithm.

2.5.6 Test and Train set selection

Testing is one of the most important steps in developing a machine learning model since it provides the information on how well a model is performing and being generalized to new cases [90], [91]. For this, there should be a train set on which the model will be trained and a test set on which the model will be tested. Choosing a test set can be done in different ways [48]. There are various ways

of dividing the collected data into a test set and train set; however, **cross-validation** provides an appropriate means of selecting the best test/train approach and understanding the consistency in the performance of the model [48], [65], [92]. This will either randomly or methodically divide the collected data into a specific number of groups and perform the analysis on all except one group (set aside for testing). Using this evaluation based on these subsampling makes the algorithm be trained on several training sets. The final performance of the model can be studied by the average and standard deviation of the performances [91], [93].

2.5.7 Algorithm Validation

Some metrics should be chosen based on the defined problem to evaluate the algorithm's performance. Here, for both classification and regression, the metrics will be discussed which are used for validation in this work.

2.5.7.1 Classification Validation

Imaging a binary classification task where each example $\boldsymbol{x_i} \in \mathbb{R}^n$ and its corresponding true value $\boldsymbol{y_i} \in \{0,1\}$, where 0 and 1 denote negative and positive examples, respectively. A classifier f predicts the class of example $\boldsymbol{x_i}$ such that $f(x_i) \in \{0,1\}$. The resulting prediction for each example can be categorized into four groups as shown in Figure 7: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) [48].

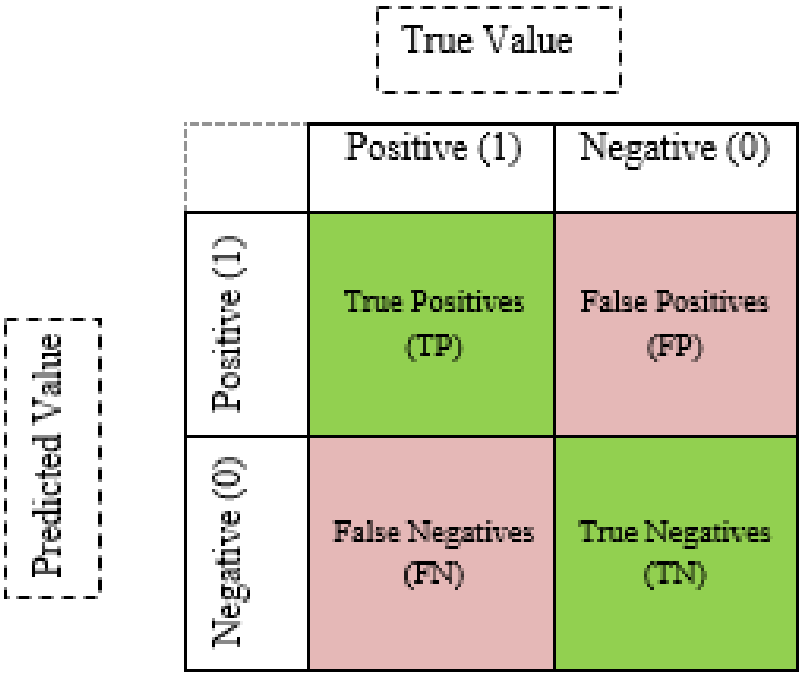


Figure 7: Confusion matrix with four categories of prediction. The green squares indicate the correct predictions, and the orange squares show the wrong predictions.

The first and most popular validating metric is called Accuracy, which is defined as the overall proportion of true predictions with respect to the entire dataset. Accuracy is the most common metric in classification, which is calculated as per the following equation [48], [94], [95]:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP}$$

Recall (sensitivity or true-positive-rate (TPR)) is the proportion of the positive examples that are correctly predicted to be positive [48], [94], [95]:

$$Recall = \frac{TP}{TP + FN}$$

Specificity is the proportion of the negative examples that are correctly predicted to be negative [48], [94], [95]:

$$Specificity = \frac{TN}{TN + FP}$$

Precision is the proportion of the true positives with respect to all the examples that the classifiers predicted as positive [48], [94], [95]:

$$Precision = \frac{TP}{TP + FP}$$

F1 measure is the harmonic mean of precision and recall [48], [94], [95]:

$$F1\ measure = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

These are the metrics that are used to measure the performance of the algorithm for classification.

2.5.7.2 Regression Validation

Regression has its own evaluation metrics, which will be discussed further. The first popular evaluation metric used in this research is the Max Error function which computes the worst error between the predicted and true value. Here is the Max Error function:

$$Max\ Error(y, \hat{y}) = \max(|y_i - \hat{y}_i|)$$

Where, y_i is the true value of the i_{th} sample and the $\hat{y}_i$ is the predicted one.

Mean Absolute Error (MAE) is another evaluation metric which computes the mean absolute error between the predicted and true value [96] as described in the function below:

$$MAE(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} |y_i - \hat{y}_i|$$

Similar to the previous function, y_i is the true value of the i_{th} sample and the $\hat{y}_i$ is the predicted one, and $n_{samples}$ is the number of samples which the MAE is estimated from.

The Mean Squared Error (MSE) function is a risk metric of the expected value of squared error [96]. The function below illustrates its computation:

$$MSE(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} (y_i - \hat{y}_i)^2$$

Lastly, one of the most important regression evaluation metrics is R^2 which computes the coefficient of determination. It is an indication of the fitting performance. The best possible score for R^2 is 1 but it cannot be compared between different data sets [96]. Here is the R^2 function:

$$R^2\ (y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \in_i^2$.

2.5.8 Error Analysis of Machine Learning Model

Another way of improving a model is to analyze different types of errors made by the model. Using a confusion matrix can be helpful where the rows are representing the actual values and columns are representing the predicted values. The color of the matrix cells will show the error rates. These will help to understand where the errors occur most and then try to decrease them [48], [97].

2.5.9 ROC Curve

One of the ways of comparing different models is using the **Receiver Operating Characteristic (ROC) Curve.** The ROC Curve plots the true positive rate (or as discussed previously, recall) versus the false positive rate, which is the ratio of negative values incorrectly classified as positive. Plotting ROC Curves for all the models allows comparing the models and choosing the most efficient one.

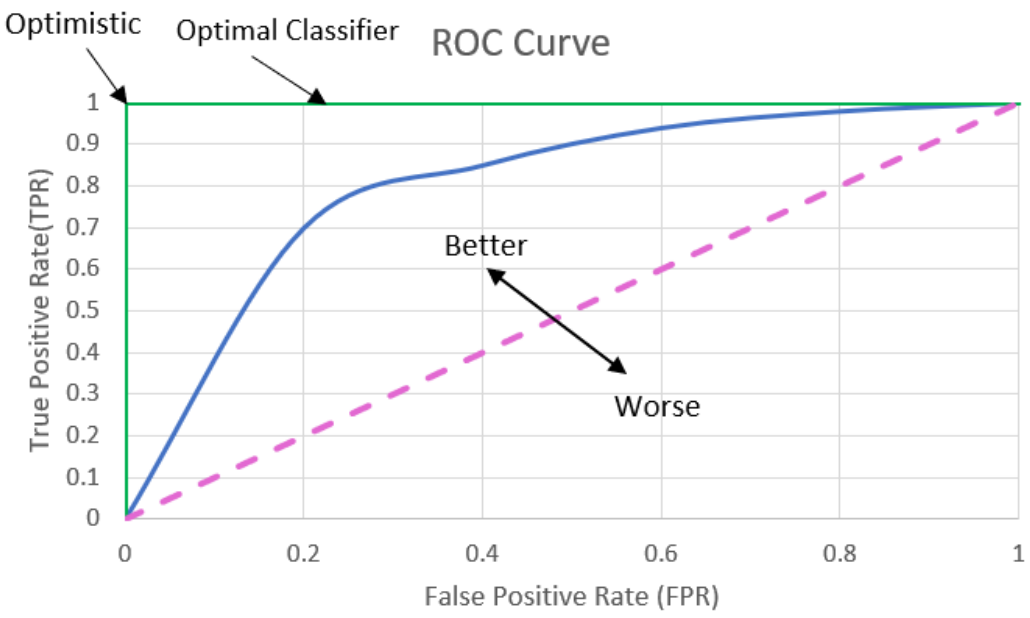


Figure 8: A sample ROC Curve along with the associated optimal classifier

As shown in Figure 8, the optimal classifier is the one whose True Positive Rate is equal to one, and its False Positive Rate is equal to zero. Any of the models' curve which is closer to that point (Point B) is predicting better than the others and have fewer errors [48], [98], [99]. The purple line is representing the curve for a baseline (or dummy) classifier which randomly assigns an output to the inputs. A poor classifier is closer to the purple line since it has a larger False Positive Rate and a less True Positive Rate [99], [100].

2.5.10 PR Curve

Knowing the trade-off between Precision and Recall can be helpful for most of the applications which require choosing a threshold. Plotting Precision over Recall can help to determine the threshold. By looking at the plot, the best threshold can be chosen by observing the rate of the change of the slope [48], [101].

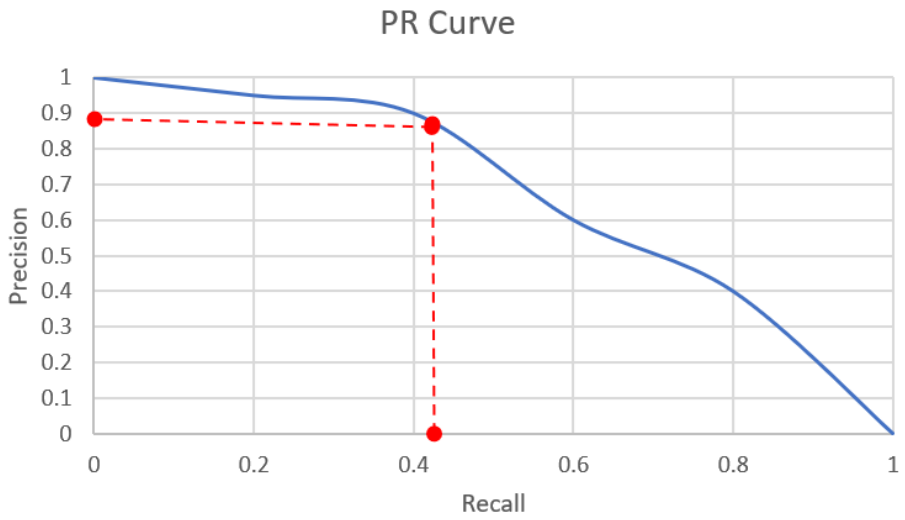


Figure 9: A sample PR-Curve along with a point that was chosen as the threshold based on the application

The example from Figure 9 indicates that the line trend is decreasing sharply from Precision=0.8. Therefore, the best choice for threshold is the one before the point.

2.5.11 Steps to follow for Machine Learning

For every Machine Learning project, there are some steps which should be taken. The first step is to understand the problem and look at it from a broader scope. The objectives of the problem should be defined and how the performance should be measured. The second step is listing the required data. Then the data should be collected and explored to find out the characteristics of the data, remove the noises, and develop appropriate filters. Then, the data should be visualized, and the correlation between attributions should be determined. The next step is preparing the data, which involves filtering the data, feature selection, feature engineering, and feature scaling. Now that the data is neat, it is time to run different classifiers and see which one is more efficient and effective for the project objectives. In this step, there is a need to use cross-validation and analyze the most important variables for each algorithm while trying to reduce the errors. The model's

hyperparameters should be tuned to improve the model and reduce errors. Once all these steps are taken, and the model is improved, the performance of the final model should be measured and utilized as the final solution for the project [49]. Figure 10 shows a block diagram of the steps which are needed to be taken for machine learning projects:

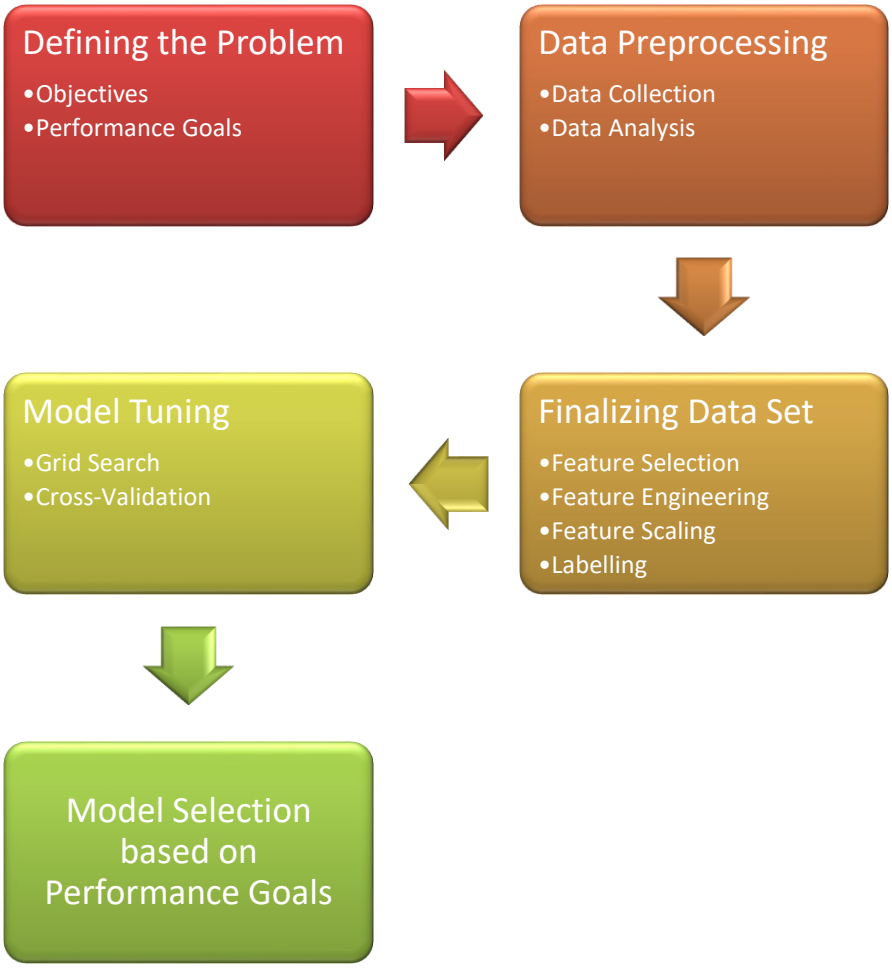


Figure 10: Steps in developing a Machine Learning project from the beginning to the end.

Chapter 3

Experimental and Primary Trial

3.1 Introduction

Stroke rehabilitation seeks to restore motor control to patients in order to increase their independence when they leave the hospital. This is typically carried out by physio- and occupational therapists manually manipulating the patient's limbs through a series of exercises without knowing the muscle's activity. Duration of exercise is an important parameter in the recovery process, which is based on the feeling of the patient. However, it is not clear if the patient muscles are tired, or it is the feeling of being bored.

Rehabilitation robots can be used as tools by therapists to provide more exercise at a higher intensity, which has been linked to more significant clinical outcomes, and by using an EMG sensor, there would be useful data for researchers and therapists to monitor the patient's progress while recording their muscle activity. The original research focused on utilizing an EMG system to monitor the lower limb muscle activity to recognize the muscle fatigue while the patient was using ViGRR-Lite. In addition to using the EMG system, the virtual game was going to be used to reduce the probability of boredom in the patient. Consequently, the level of tiredness of the participants was going to be considered by frequently asking them during the experiment.

During the COVID-19 quarantine and physical distancing period, experiments with human-to-human physical contact had been avoided. However, we performed a modified set of experiments that were done remotely without compromising safety. This enabled the collection of basic data needed to continue the research until restrictions are lifted. The robotic system was replaced with a gaming activity that tracks the participant as they perform squats. This motion was selected since squats are an appropriate sample of sit-to-stand motion, and it is a suitable exercise to induce fatigue in the muscles of healthy subjects to simulate the endurance of stroke patients while using the ViGRR-Lite robot. For this, we had developed a vision-based body motion system that is very lightweight and easy to set up, which is explained further in Section 3.2.

Two different experiments were conducted in this research:

1. Experimental Trial: The goals of this phase of the experiment were (i) to develop a set of analysis methods for determining muscle fatigue, (ii) to select the appropriate muscles for the primary trials, and any potential clinical trial in the future.
2. Primary Trial: The goals of the second phase of the experiment were, (i) to validate a procedure for future clinical studies by examining the procedure on ten participants, (ii)to recognize the muscle fatigue using the developed analysis method in the experimental trial phase, (iii) to prepare the data for further use in machine learning.

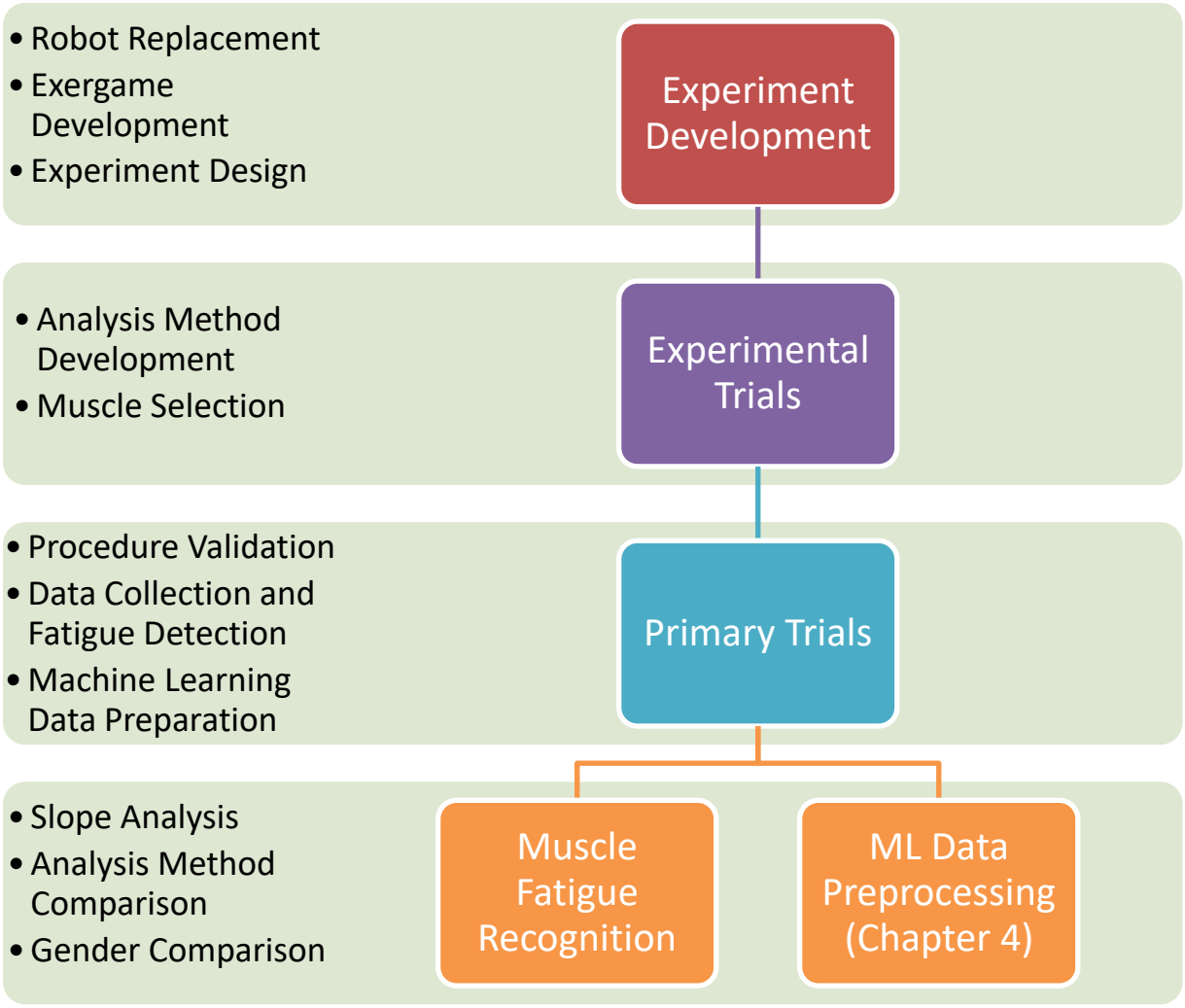


Figure 11: Summary of different phases of research

3.2 Overall Setup

Figure 12 represents the overall setup of the experiment and each part will be explain individually.

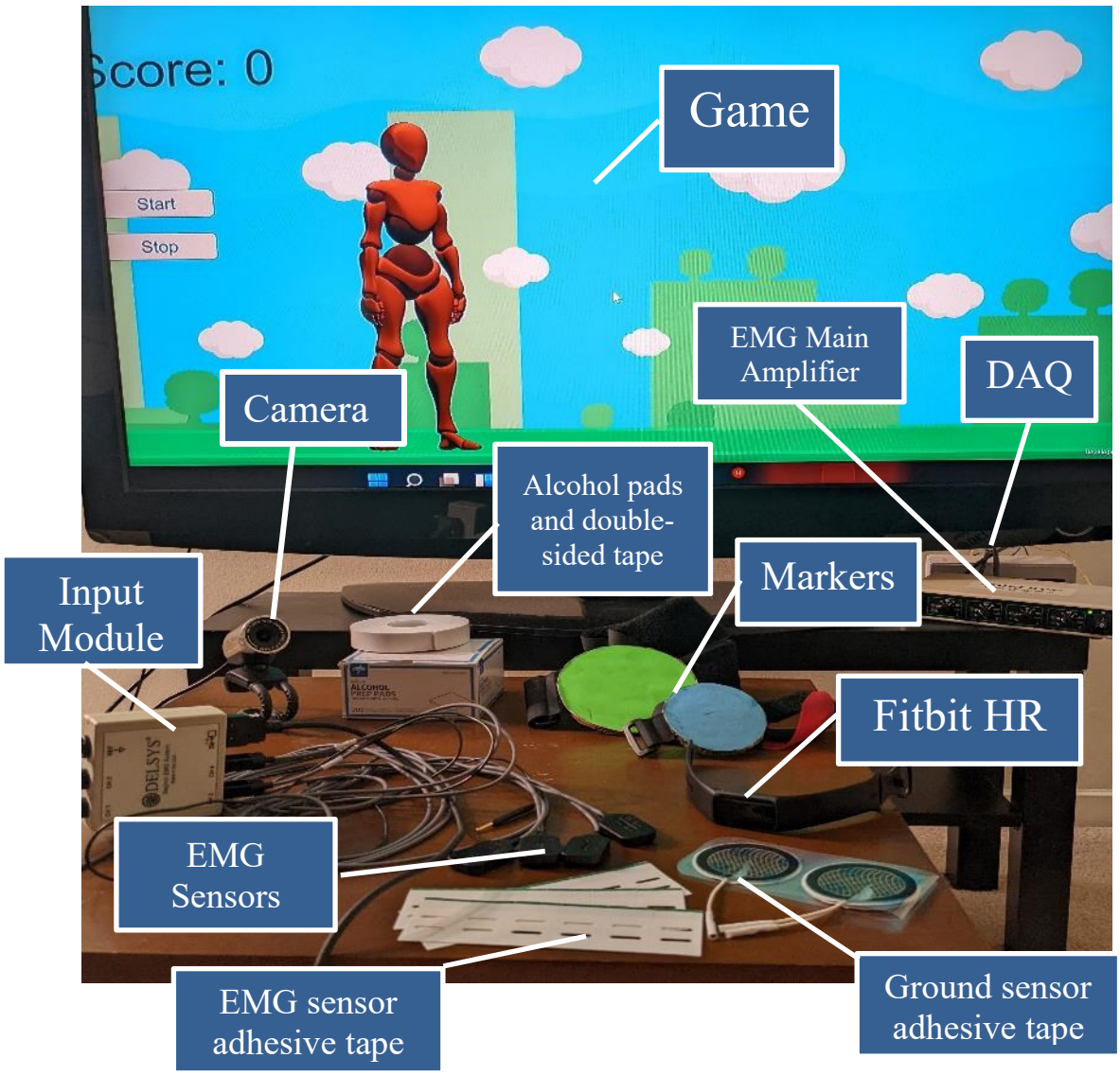
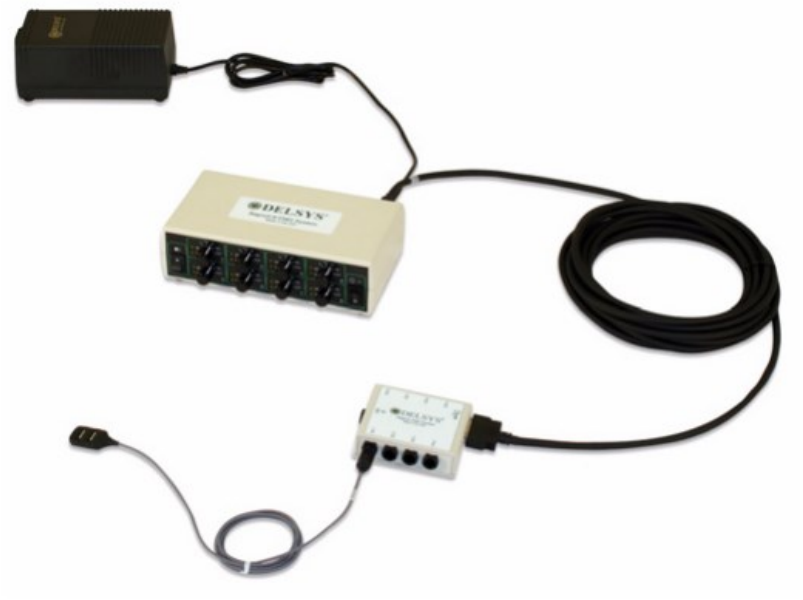


Figure 12: Overall experiment setup used in this research

3.2.1 Surface EMG Sensing Equipment

A Delsys Bangoli Surface EMG acquisition system was used for this experiment. Delsys is a reputable and well-known brand in the industry for producing equipment with exceptional quality and safety. Additionally, they are easy to use, considering no conductive gel is required, as they have dry sensing electrodes. In Figure 13, two different set-ups used in this experiment are represented: an 8-channel EMG sensor for the experimental trial and a 4-channel EMG sensor for the primary trial. The system uses skin-safe double-sided tape to adhere the electrodes to the skin surface. An image of the system can be seen below:



(A)

(B)

Figure 13: Delsys EMG Sensors, A: 8-Channel and B: 4-Channel on the right [102]

In Figure 14, different parts of the system are numbered: 1) Sensor 2) Main Amplifier, 3) Input Modules, 4) Input Module Cable, 5) Inter-Module Cable, 6) International Power Supply.

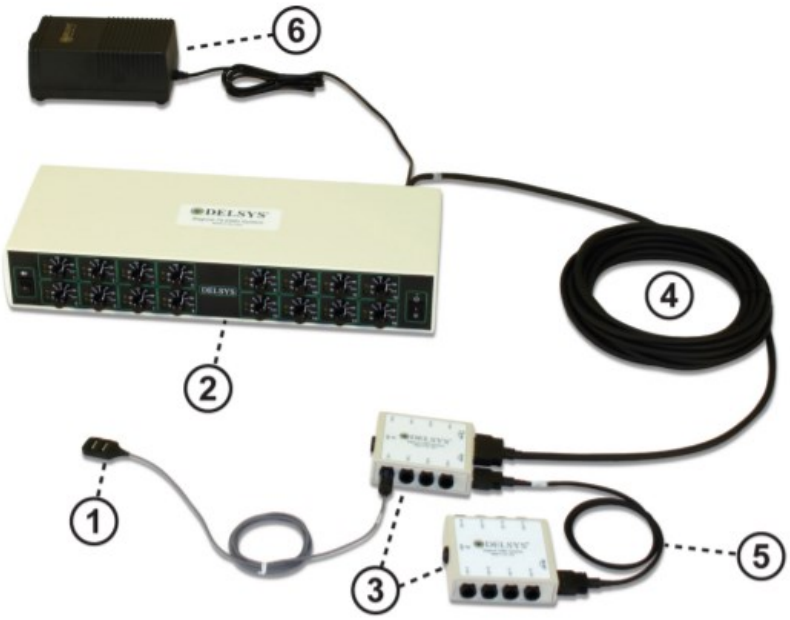


Figure 14-EMG Sensor System[102]

3.2.2 Fitbit Inspire HR

For added safety of the participants, the Fitbit Inspire HR is utilized to monitor their heart rate and ensure the maximum advisable heart rate is not reached. The maximum heart rate is calculated for each participant by subtracting the participant age from 220 as per [103]. Furthermore, the desired heart rate zone is also calculated- the level at which heart is being exercised and conditioned but not overworked. The American Heart Association recommends a target heart rate of 50% to about 70% of the maximum heart rate for moderate exercise. The participants will be asked to pause the experiment to get back to the resting rate if they exceed the target heart rate range. This device has been designed with a three axes accelerometer to track the heart rate during different activities and uses a real-time heart rate tracker with 1 Hz sampling rate that is connected to an application on mobile devices [104].

Figure 15: The Fitbit Inspire HR that was utilized for monitoring participant heartrates

3.2.3 National Instruments USB-6216 Data Acquisition Box

A National Instruments USB-6216 Data Acquisition Box is used in the experiment. They are USB-based measurement and automation devices which provide analog inputs/outputs, digital inputs/outputs, and more [105]. All computer-based systems that acquire readings from various sensors/transducers, and control multiple outputs/actuators, require a data acquisition to go between the computer and the various inputs and outputs. USB-6216 Data Acquisition Box is a 16-bit analog-to-digital converter which has been used to sample signal at 1000 Hz in this research. Channels 1 to 8 was utilized for the experimental trial, whereas channels 1 to 4 were utilized for primary trial. A custom-coded interface is used to store the data on computer during the experiment.

Figure 16: National Instruments USB-6216 Data Acquisition

3.2.4 AUSDOM AW615 Webcam

The webcam has been used to track the ankle and hip positions of the participants throughout the exercise. It uses a simple USB connection with the computer. Video has not been saved or used for further analysis; only the tracking data were stored.

Figure 17: AUSDOM Webcam

3.2.5 Markers

Two markers were used for tracking the movement of participants. As shown on Figure 18, the green marker is placed on the participant's upper thigh and the blue marker was placed on the ankle. These two markers and the camera tracked the squats during the experiment. The lighting was adjusted in a way that the motion of the participant induces any shadow over the markers.

Figure 18: Markers used in the experiment for recording the squat motion which is also required for interacting with the exergame

3.3 Game Set-up

For this experiment, an exergame was designed to prompt the squat motion at a predetermined pace. The virtual character in the game replicates the participant's motion based on the data acquired from the markers and the camera. Figure 19 represents the game environment where the participant must perform the squat motion based on a trajectory that collects the most coins and prevent the obstacles:

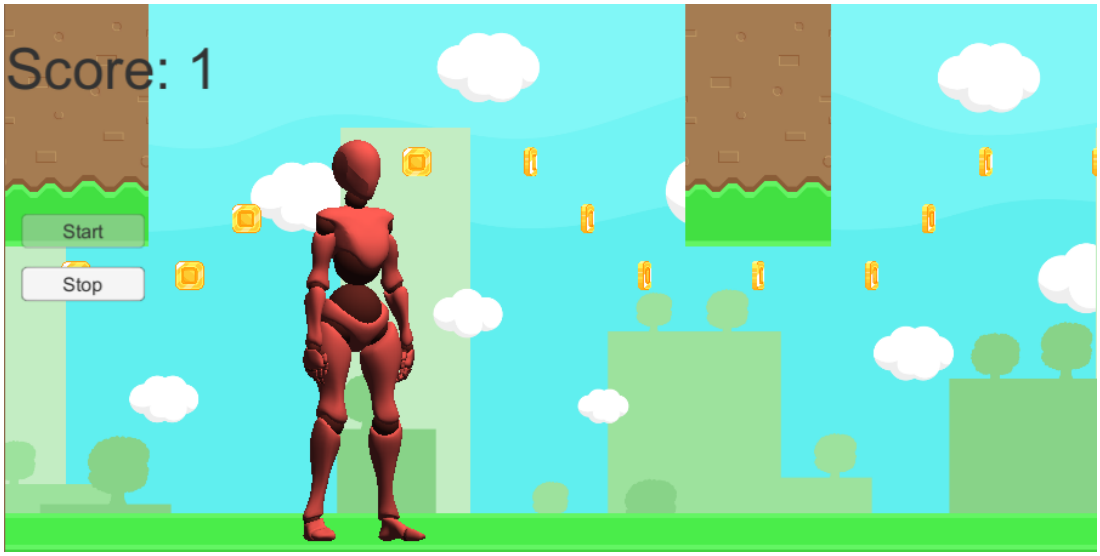


Figure 19: The exergame environment where the virtual character replicates the motion of the participant with the goal of collecting the most coins without hitting the obstacles.

The coins are representing the desired trajectory for squatting and by going through the coins, the player will collect points (each coin will have one point). Obstacles move towards the character in the game. They should be avoided by doing squats and passing them from the underneath; failure to do so deducts five points. During the game, the score is shown on the screen along with the start and stop buttons. Whenever the participant does not feel good and wants to stop the game, or the heart rate is over the safe region, the researchers would stop the game.

Figure 20: Game environment with a notification for the collection of perceived fatigue level

Every 30 seconds, "RATE FATIGUE 1-10" appears on the screen as it is shown in Figure 20. It is the time that the researcher will ask the subjective muscle fatigue level which in this work is called Perceived Fatigue. The ratings are added to the form designated for collecting the perceived fatigue from each participant. The form is attached to the Appendix Section A.

3.4 Ethics Approval

Any research involving human interactions requires an ethical approval from Carleton university. The Carleton University's Research Ethics Boards (CUREB-B) for engineering research experiments approved this research titled "Design and implementation of a Haptic Gait Rehabilitation robotic platform" with an approval number: 10090 11-1631.

3.5 Email Invitation

After acquiring the ethical approval from Carleton University, an invitation email was sent which is attached to Appendix B.

3.6 Questionnaires

The experiment consists of two different questionnaires which has been filled up by the participants before and after the experiment.

3.6.1 Pre-Questionnaire

A pre-experiment questionnaire was designed to ensure the participants did not have any health issue or did any moderate exercise in the month prior to the experiment to be sure that they would not have any issues or potential injuries during the experiment. It is attached to Appendix C.1.

3.6.2 Post-Questionnaire

When the experiment is done, each participant was asked to fill up a post-experiment questionnaire which consists of some questions about the experiment, the game, how fatigued they feel in their thigh and shank area, and which one got tired first so that make them to stop the game. The questionnaire is attached to Appendix C.2. This information can be helpful for further discussion and future experiments.

3.7 Pre-Experiment Procedure

After ensuring all the safety protocols regarding the prequestionnaire are followed, the participants are asked to sign a consent form (attached to Appendix D). The participants were advised against strenuous leg workouts on the day of the experiment. They are also advised to wear a comfortable short for squats which is neither green nor blue as it will interfere with the tracking system. After explaining the experiment, a video has been shown to the participant to make sure that they know the right way of doing squats [106]. Some leg measurements from hip to ankle and knee to ankle was performed on each participant.

3.8 Test Procedure

The procedure of doing the test with participants have different steps which will be discussed below:

3.8.1 Attaching EMG Sensors

First step of doing the experiment is attaching the EMG sensors to measure the muscle activities during the experiment. For both the experimental and primary trails, a document was prepared containing appropriate electrode placements on the muscles and methods to find these locations. This information was collected from SENIAM (Surface Electromyography for the Non-Invasive Assessment of Muscles) project [107]. All the electrodes should be attached to the muscles of the same leg as shown in Figure 21. Skin preparation for EMG electrode attachment involved shaving any hair and cleaning with an alcohol wipe. The ground sensor was placed on the knee bone. EMG electrodes are not active, and they operate with a ±5V DC power supply, hence there is no risk of

shocks during these experiments. There is a minor risk of discomfort or skin irritation when sticky tapes are taken off from the skin at the end of the experiment.

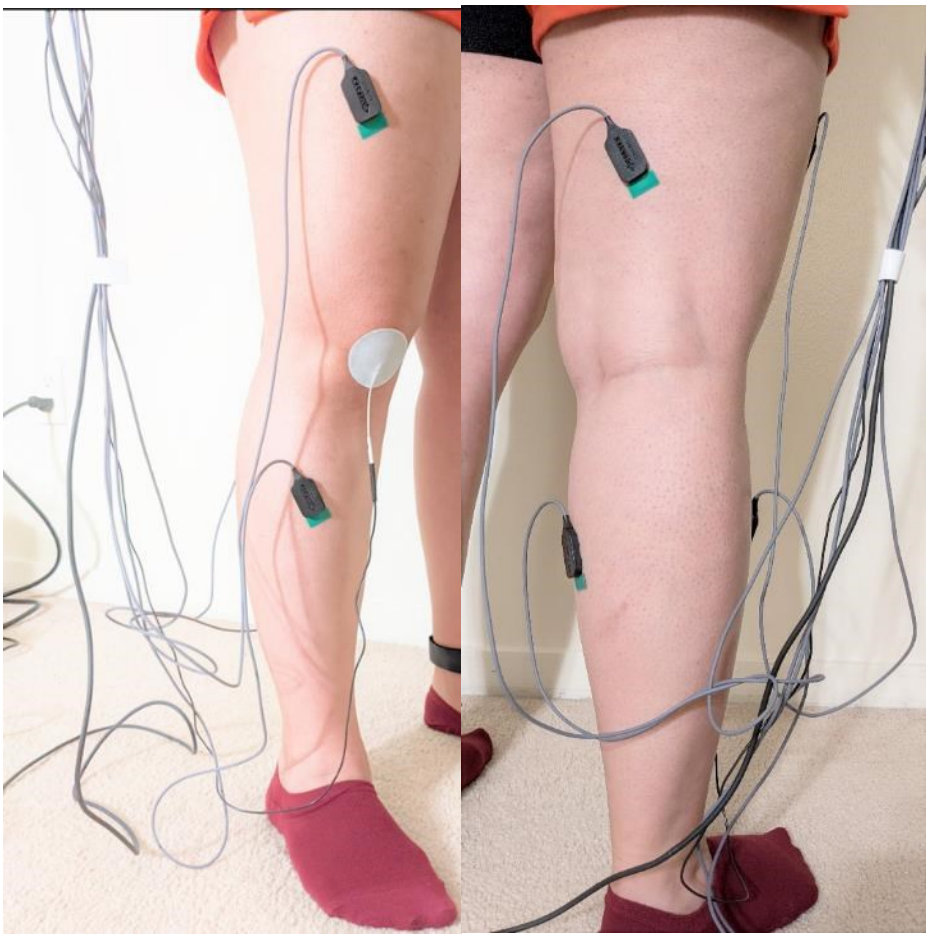


Figure 21: 4-channel surface EMG sensors attachment in the experimental trial for front and back of the leg

3.8.2 Tracker Set Up

As discussed in 3.3, there is a need of tracking the participant movement to play the exergame. The second step in the procedure is setting up the tracker. The markers should be attached to the leg that does not have the EMG sensors, and the webcam should be placed in front of them approximately at knee level as it is shown in Figure 22.

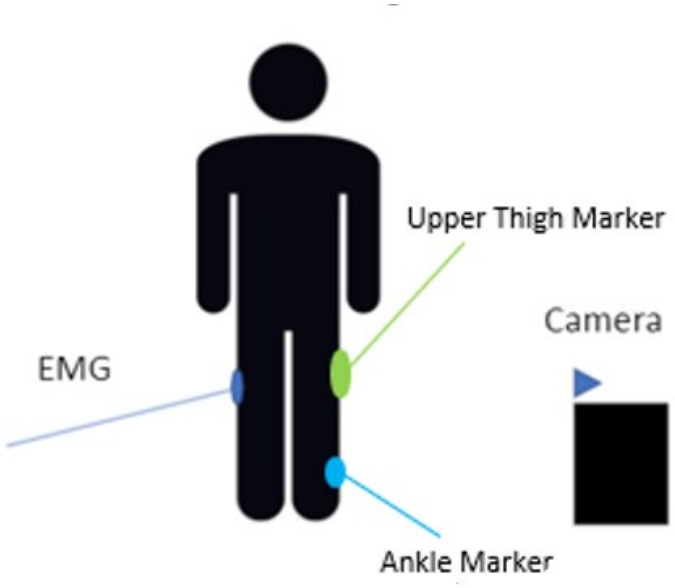


Figure 22: Camera placement with respect to the participant and the markers. Camera is facing the markers.

The green marker should be placed on the upper thigh and the blue marker should be attached to the ankle as shown in Figure 18 and Figure 22. For better performance of the camera, there is a need to ensure the camera is not pointing at any bright light sources (e.g. window). Therefore, closed curtains with indoor lighting is used to ensure that the camera can fully detect the markers and reduce the risk of detecting non-desirable spots. The detection of the markers must be tested prior to running the exergame. Any object mistaken as the markers should be removed or hidden to not distract the tracking process.

3.8.3 Heartbeat Monitoring

The maximum allowable and the average heartbeat rates were calculated for each participant as explained in 3.2.2 and can be found in Table 1. The heartbeat monitoring was done using the Fitbit app on the mobile.

3.8.4 Game explanation and scaling the tracker

The last step before starting the experiment is presenting the game environment to the participant and explain how to play the game and ensue, they know exactly how they should play the exergame. Also, participants should have enough space to do the squats. Before starting the experiment, there is a need to calibrate the tracker by prompting the participant to perform one squat to record the position of their full squat and full stand.

3.8.5 Perceived Fatigue

After doing all the discussed steps and preparations, the experiment was started. Every 30 seconds the researcher asked the participant to rank their level of tiredness form 1-10 (10 being the most tired) as discussed in 3.3 (The template of the form is available in Appendix A). The experiment was continued until the participant got tired and did not want to continue with more squats.

3.9 Experimental Trial

3.9.1 Introduction

Two goals were set for the experimental trials:

1. Develop the data analysis methods to detect muscle fatigue.
2. Determine the muscles that are appropriate for fatigue detection by observing the variations in the EMG data that represent muscle activity.

Based on the studies reviewed in literature review, it is expected to find a decreasing slope in MNF and MDF of the EMG signals collected from muscles as they become fatigued. The test procedure as discussed in Section 3.8 was conducted on eight participants to achieve these goals. All EMG data, game data, and the perceived data were collected to develop an appropriate post-processing

method for the experiment. For the preliminary investigation, the 8-channel EMG system was used to collect data from 8 different muscles in the thigh and the shank. Note that this research aims to develop a tool which can be used with rehabilitation robots by physiotherapists. It is important to find the minimal set of required EMG channels for fatigue detection in order to prevent time consuming preparation, and measurement for the attachment of the sensors at each session.

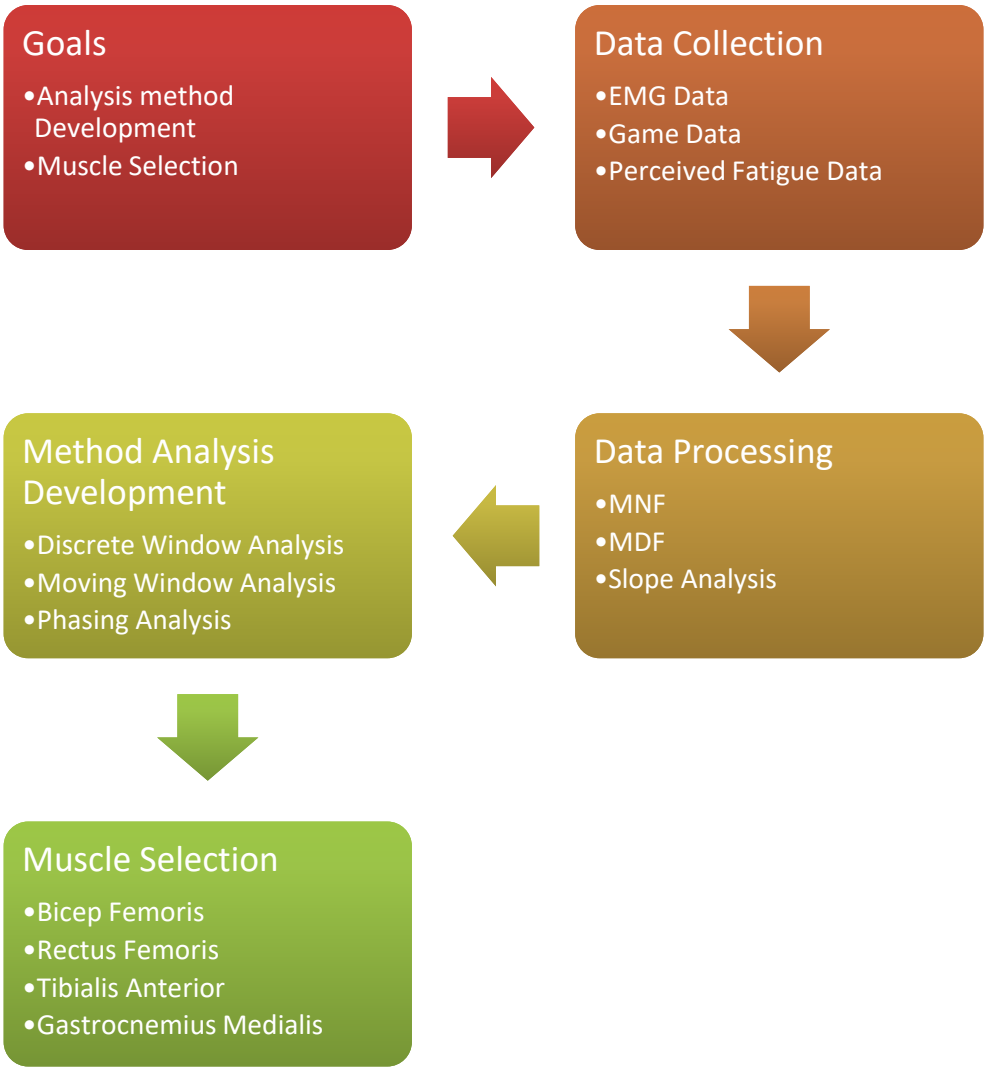


Figure 23: Summary of the experimental trial using 8 muscles

3.9.2 Method

Eight muscles from the lower limb (the shank and thigh) were selected for collecting data: Tibialis Anterior (TA), Gastrocnemius Lateralis (GL), Gastrocnemius Medialis (GM), Soleus from the shank, and Biceps Femoris (BF), Semitendinosus, Quadriceps Femoris (Vastus Medialis), Quadriceps Femoris (Vastus Lateralis) from the thigh which their locations are shown in Figure 24.

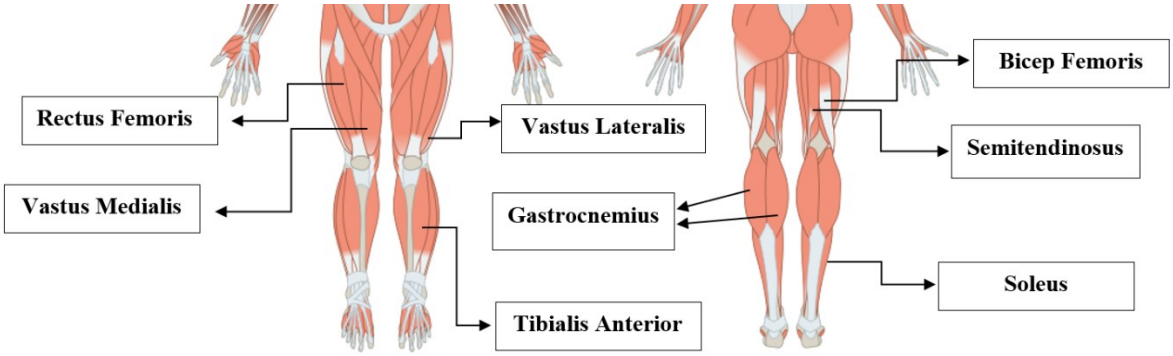


Figure 24: Schematic of lower limb muscles used in the preliminary analysis (labels are added to the base image from [108])

After the test preparations and procedure were done and the data were collected from the participants, there is a need to process the data since the EMG data has a lot of noise and needed to be filtered as discussed in Section 2.3. The EMG signals recorded from the muscles was filtered by applying a bandpass filter with a range of 20 Hz to 400 Hz which were implemented by Butterworth's filter in MATLAB. The sampling frequency of 1000Hz was used to have a Nyquist rate of 500 Hz. Moreover, the 60 Hz noise from the powerline was also removed using a band stop filter with a cutoff frequency of 58 to 62 Hz, 118 to 122 Hz, and 178 to 182 Hz. After applying the filters, the signal was rectified by taking the absolute value and subtracting it from the mean of the

filtered EMG data. An envelope of the EMG signal was created using the RMS function as defined below:

$$RMS = \sqrt{\frac{1}{n}\sum_{i} x_i^2}$$

Where, RMS is root mean value, n is the population of the data points, x_i is the sample data value, and i represents the sample data index. The scheme for selecting origin and end of each envelope was chosen based on the type of the analysis. Once the envelope was obtained, MDF (Median Power Frequency), and MNF (Mean Power Frequency) of the data were determined since they are important parameters in recognizing fatigue as discussed in Section 2.4.

MNF and MDF slope is used for recognizing the muscle fatigue of participants based on the literature in fatigue recognition. Three different types of analysis were developed in this step to determine the most appropriate method for fatigue recognition.

Note that EMG signal normalization is a reliable basis for observing the activity of a muscle in different days or comparing muscle activity in different individuals [109], [110]. In this work, observing the muscle activity trend during the experiment is of more importance, therefore, EMG normalization has been omitted. Moreover, collecting data with EMG sensors may include different types of noises. Cable motion artifacts and movement artifacts are one of the noises which are considered in some of the studies that used sEMG sensors [111]. With the use of modern electronic technologies, companies like Delsys have almost eliminated cable motion artifacts. Furthermore, movement artifacts are generated when muscle moves underneath the skin, or when a force impulse travels through the muscle causing the skin to move. This noise can be mostly

filtered out with a high-pass filter of 20 Hz which may remove some muscle activity data. This has not been studied in this research.

3.9.2.1 Discrete Window Analysis

The average time of doing one squat for a healthy subject is about three seconds and the EMG system records one thousand data in one second. Furthermore, reliable data analysis on each EMG data envelope requires enough data. As a result, different window length sizes were examined to realize the suitable length for our analysis based on the muscle activity trends provided in the literature. In this method, the data was divided into discrete windows with the length of nine seconds (equivalent of three squats). For each window (interchangeably referred to as bin), the MNF and MDF is calculated and represented as a single data for the window. A window length of 9 seconds includes enough data to have consistent MNF and MDF trends, however, there will be no sudden jumps in the values for individual bins.

The overall analysis for the discrete windows analysis was developed and resulted in Figure 25 and Figure 26 as examples. Each blue dot represents the MNF or MDF of the corresponding muscle for each bin. The red line was fitted through the entire data indicating the change in the muscle activity as the time passed. Furthermore, the perceived fatigue data was added to the graphs for comparison between muscle activity and perceived fatigue. The original perceived fatigue data (1 to 10, least to most tired) was reverted and scaled to a 10 to 0 scheme where 10 is the least tired condition.

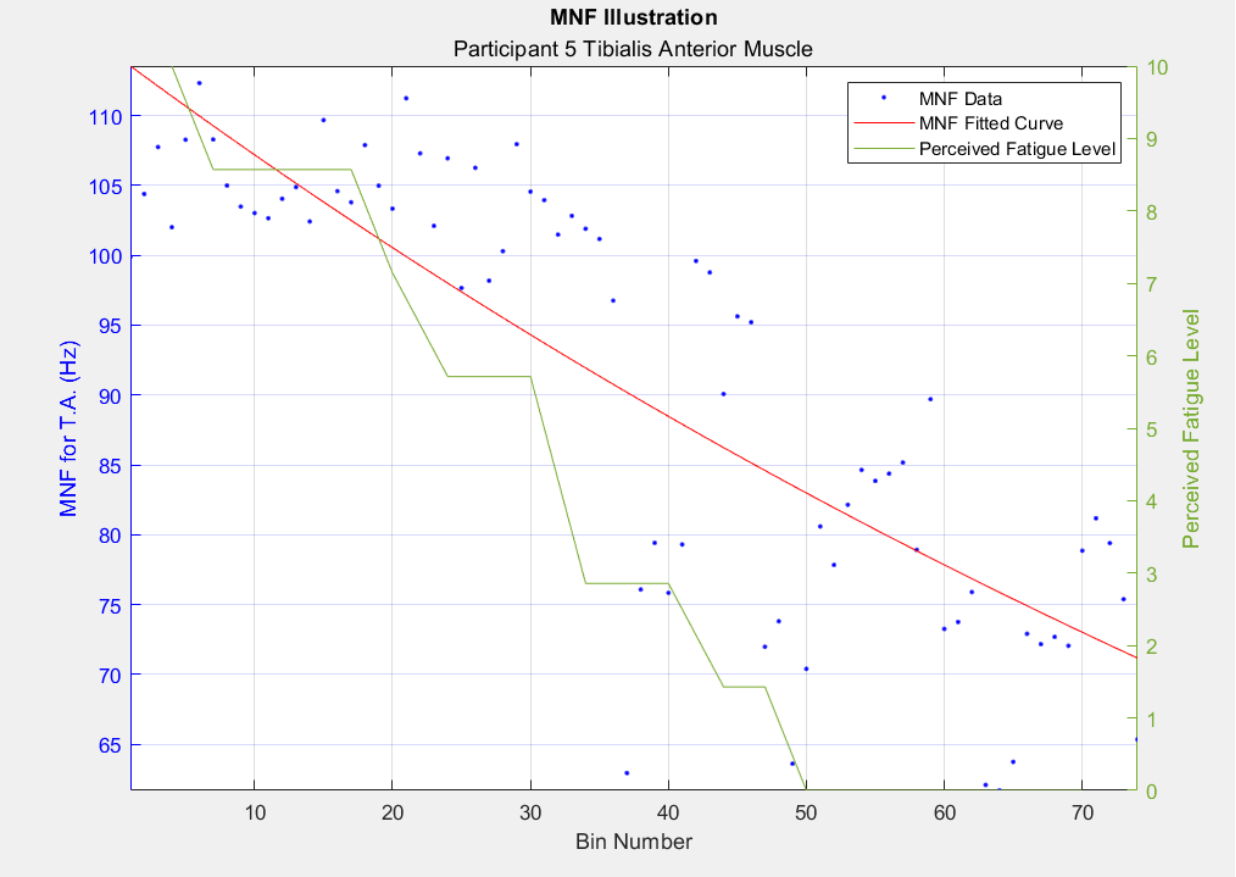


Figure 25: Graph of the MNF discrete window analysis of 9 seconds in experimental trial for Tibialis Anterior muscle in participant 5

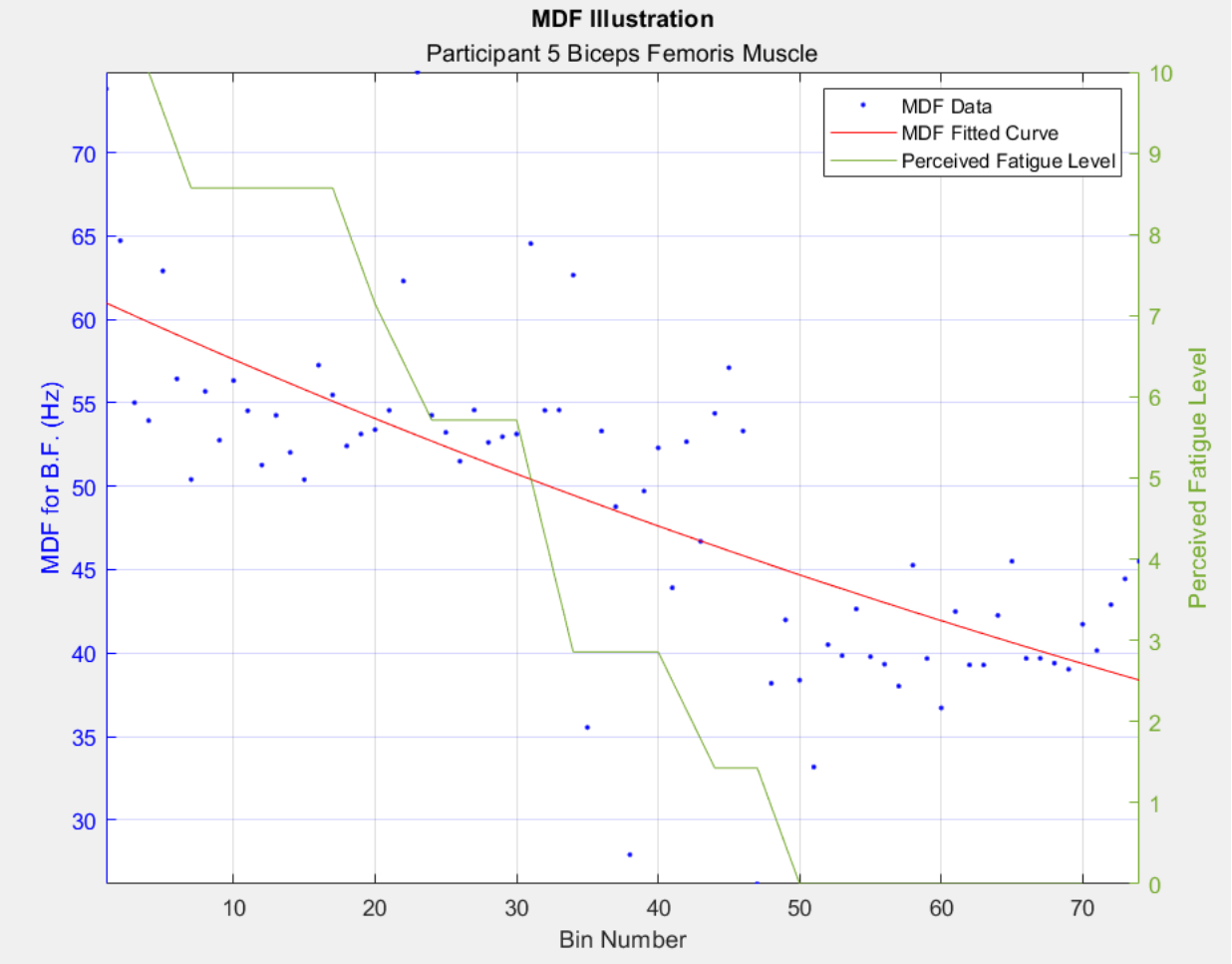


Figure 26: Graph of the MDF discrete window analysis of 9 seconds in experimental trial for the Bicep Femoris muscle in participant 5

Figure 25 is an appropriate illustration of the fatigue detection using the muscle activity data. The blue dots in the first 30 bins indicate a relatively linear behavior, and a similar trend is also seen for the last 25 bins. However, there is a downward trend from the 30th to the 50th bins. In addition to the negative slope of the fitted line through the data, the overall value of the MNF is decreasing which indicates muscle fatigue. The perceived fatigue data passes the half-way point between the 30th and the 50th bins; this is yet another confirmation that the MNF and MDF values can help in determining muscle fatigue.

3.9.2.2 Continuous Window Analysis

In addition to the expected slope decrease in the discrete analysis of MDF and MNF, observing the muscle activity information in relation to the previous squats can be useful. Therefore, this method is developed to observe the muscle activity data in relation with the previous squats. In this method, two parameters are defined as the step size and the moving window size. MNF and MDF analysis were done based on the step size, however, the start points for the corresponding bin changes based on the moving window size. In other words, the start points for the data used in each bin is identified by the moving window size, and the amount of the data used for each bin is based on the step size. Figure 27 represents an illustration of the continuous window analysis using a bin size of 9 s and a moving window size of 2 s. The first 8 bins for this analysis are shown where the red area indicates the interval at which the MNF and MDF values of the EMG signal would be calculated.

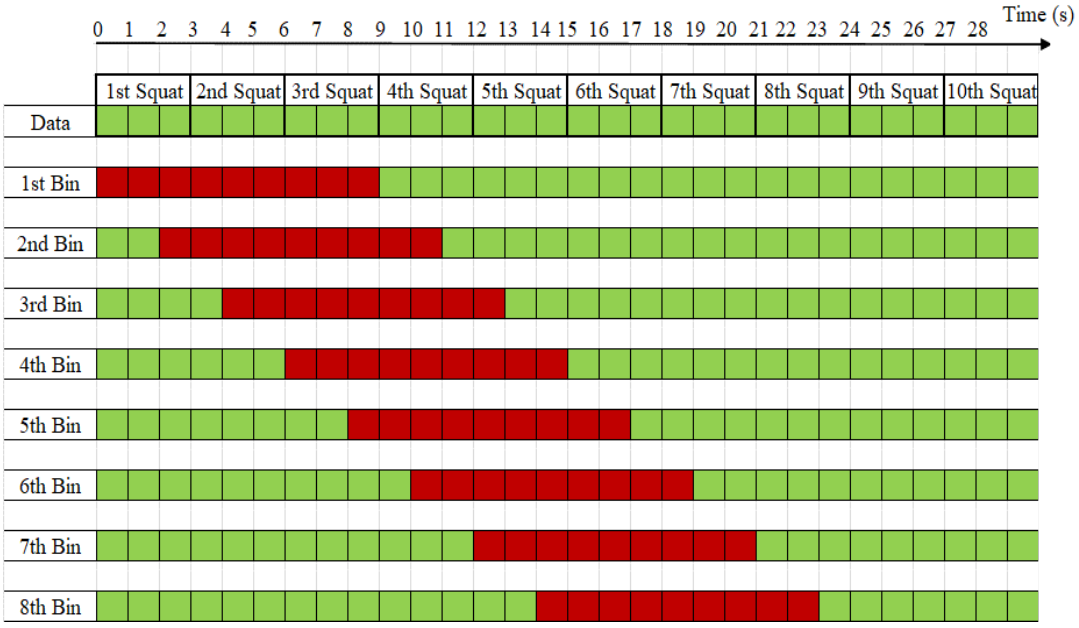


Figure 27: Visualization of the moving window method based on step size of 9 s and moving window size of 2 s

After further trial and errors, the nine seconds step size and the moving window length of two seconds (as shown in Figure 28 and Figure 29) performed well. Similar to the discrete window analysis, the 9 s bin size provided enough data for achieving a consistent trend with small variations in the calculated value. The 2 s moving window was selected strategically to have a low number of consecutive squat sets (the interval covers consecutive squats and no overlap with other squats, similar to 1st, 4th, and 7th bin in Figure 27) while getting a large number of bins that help with observing the muscle fatigue. In this case, each squat data is calculated in more than one discrete window and most calculations considered some part of the previous squats.

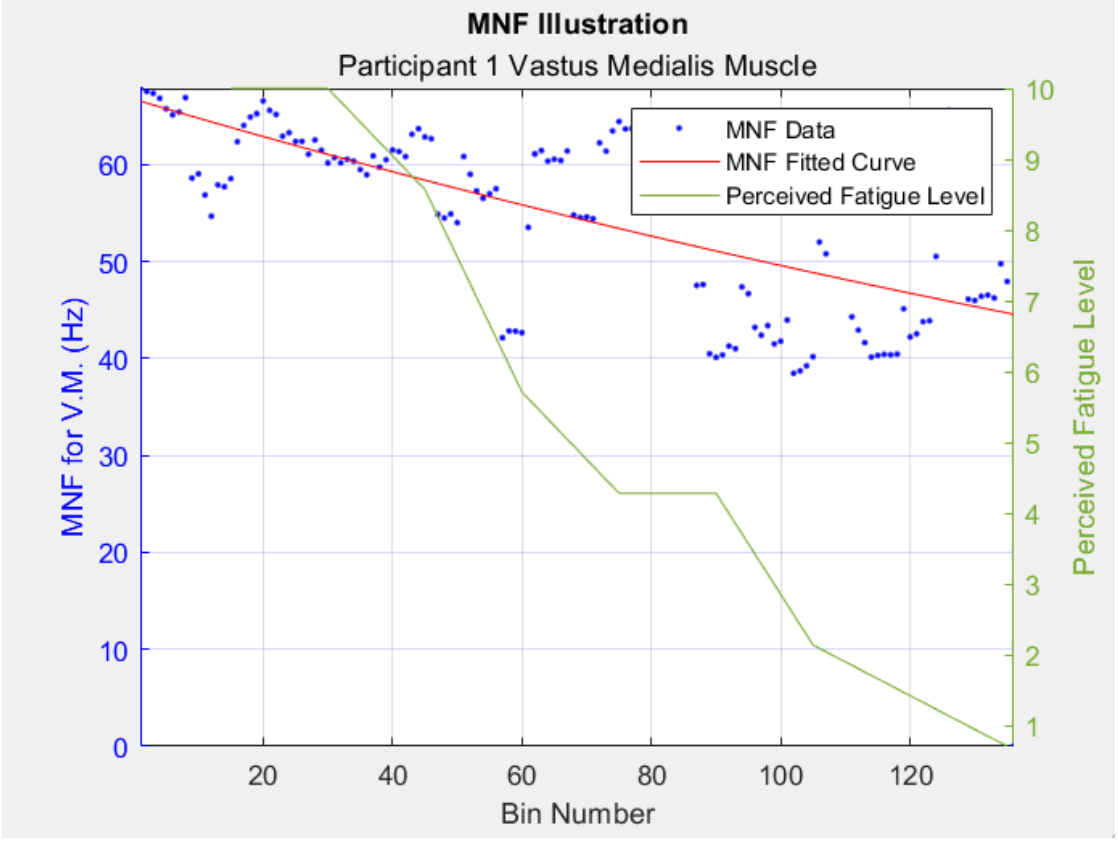


Figure 28: Graph of the MNF continuous window analysis (step=9s, Moving Window=2s) in experimental trial for Vastus Medialis muscle in participant 1

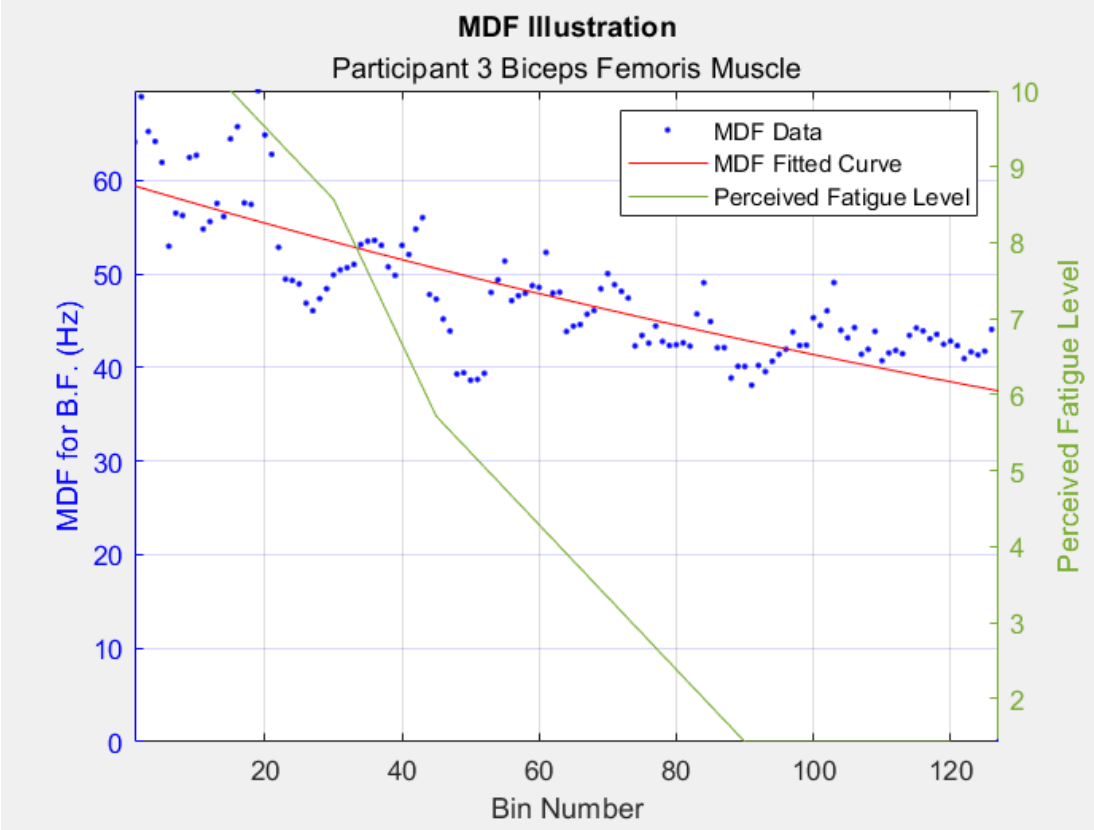


Figure 29: Graph of the MNF continuous window analysis (step=9s, Moving Window=2s) in experimental trial for Bicep Femoris muscle in participant 3

3.9.2.3 Phasing Analysis

Squat motion can be divided into four phases based on the joints motion:

1) Full stand
2) Moving down from full stand towards full squat (Sitting down)
3) Full squat
4) Moving up from full squat towards full stand (Standing up)

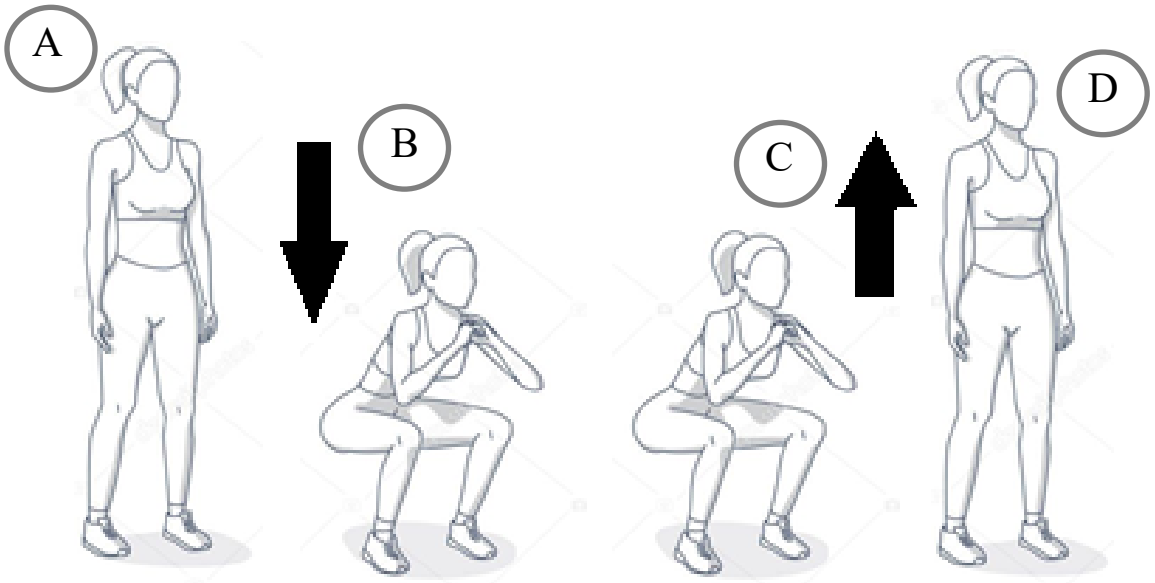


Figure 30**: A: Full Stand Position, B: Full Stand to Full squat, C: Full squat to Full Stand, D: Full stand position** [112]

During the squat, different muscles are utilized in the motion. In each of the moving phases, some muscles are more active than the others resulting in a faster inception of fatigue for those muscles. Furthermore, after the experiment, participants claim that they felt differently during each of the phases. For example, one of the claims was that the pressure in the muscles made it most difficult to perform the third phase (moving up from full squat to full stand) which implied the muscles in charge of this motion were fatigued earlier. Therefore, the phasing method was developed to detect muscle fatigue in different phases of the squat.

The game data including the tracking information provided an opportunity to find the participant relative position at any point in the experiment which can be aligned with the EMG data. A complete squat has a relative **position** between 0 (full squat) and 1 (full stand) which may be a little off due to the participant lack of engagement during calibration compared to the dynamic

scenario during the exergame. Furthermore, as the time passed and participant became more tired while being excited to play the exergame, the full squats were shortened as their squat position did not show the exact range of 0 to 1.

Position data as a function of time was utilized in order to determine the phases of the squats and align the EMG data. Peaks and valleys represent the full stand and full squat respectively, therefore, there is a need to find each peak and each valley. The built-in MATLAB function for finding the peaks was utilized to find the peaks in the data with appropriate thresholds. The valleys were also identified with the same method on the inverted data. Figure 31 illustrates the results associated with detecting the full squat and standing up phases for participant 1. Figure 32 is the enlarged graph of the interval 125s to 160s on Figure 31. As mentioned previously, the position of a participant is recorded from 0 to 1 (full squat to full stand) based on the calibration data. The blue line in Figure 31 and Figure 32 is the position profile during the experiment, and the green lines indicate the beginning of full squat to the beginning of full stand phase in each squat. In this research, all the phases and their respective combinations have been examined, and the interval that starts from the beginning of full squat phase and ends at the beginning of the full stand phase provided the best result for muscle fatigue detection (aligned with the participants comments).

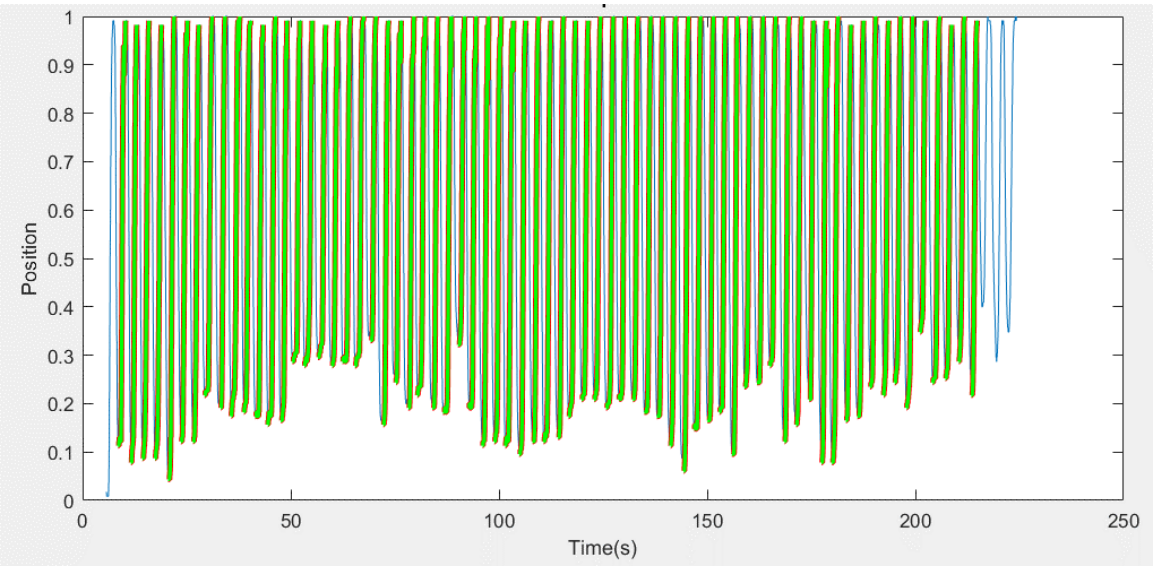


Figure 31: Detection of peaks and valleys of the position data for participant 1 in order to identify the full squat and standing up phases.

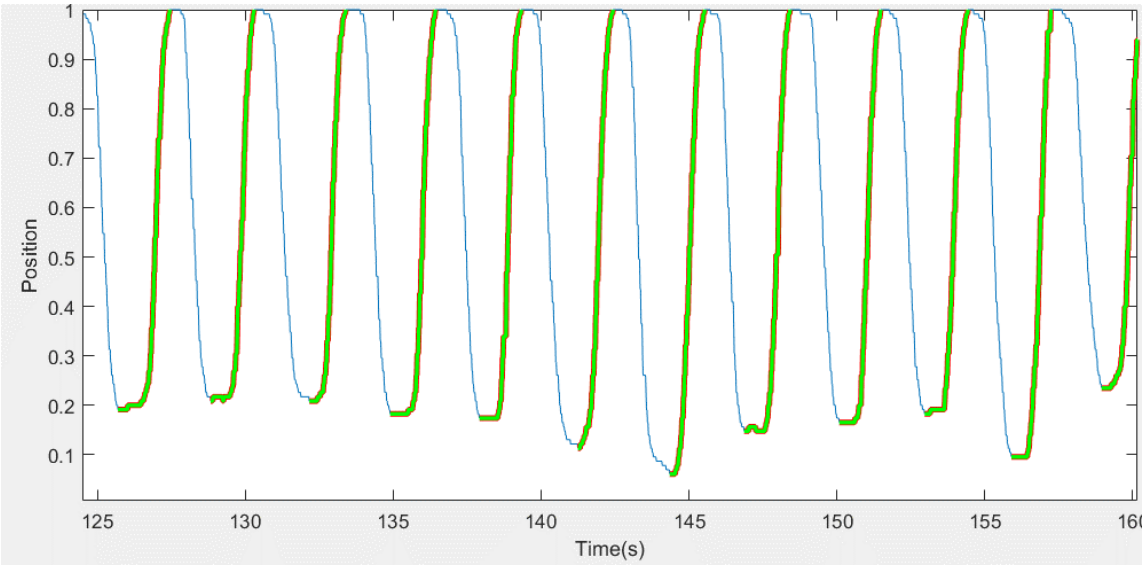


Figure 32: Peaks and valleys of the position data from 125s to 160s for participant 1

Associated times for the beginning and end of each phase were determined from game data, and they can be synchronized with the EMG data timestamps for the calculation of the MNF and MDF

for any phases. Therefore, by using this method moving down motion and moving up motion data is used in the process of detecting fatigue separately.

Similar to the previous methods, the MNF and MDF of the data were calculated and a line was fitted to the data. Figure 33 and Figure 34 represent examples of the phasing analysis for the combination of the full squat and standing up phases where each of the blue dots represent the associated analysis for a single squat.

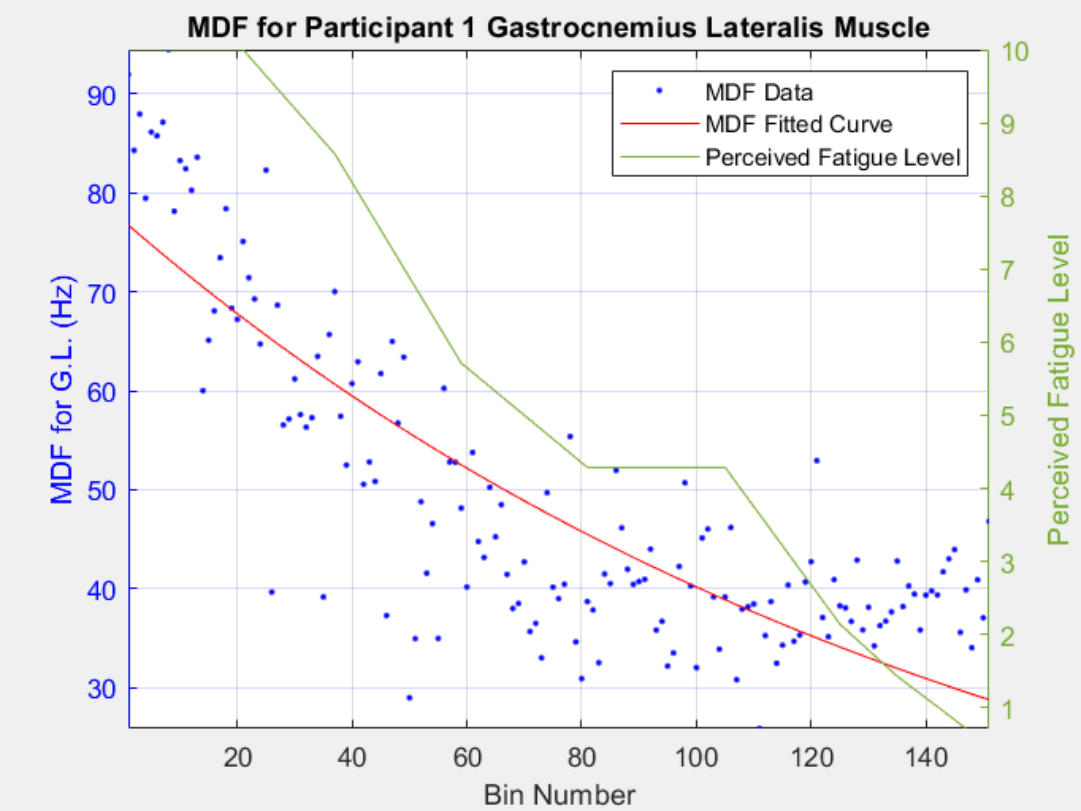


Figure 33: Graph of the MDF phasing analysis in experimental trial for the Gastrocnemius Lateralis muscle in participant 1

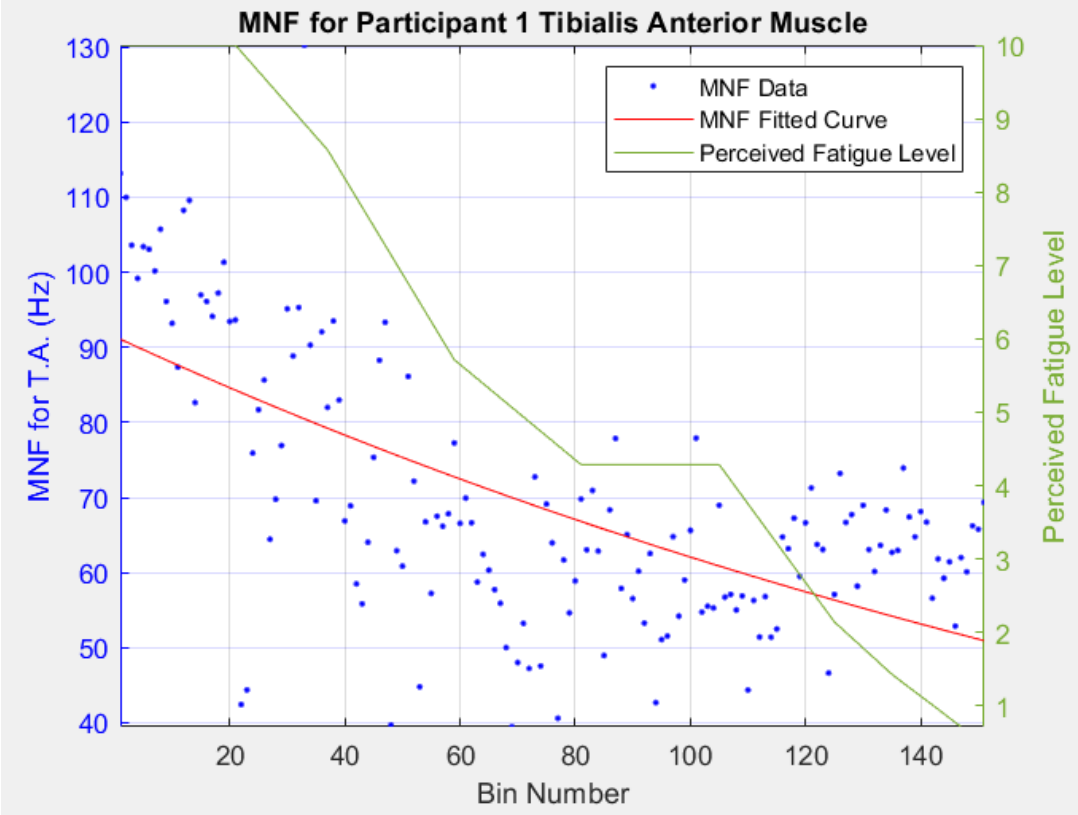


Figure 34: Graph of the MNF phasing analysis in experimental trial for the Tibialis Anterior muscle in participant 1

3.9.3 Muscle selection

As explained in Section 3.9.1, the goal of the experimental trial was to first develop methods which assist in detecting fatigue in muscles and second, to determine which muscles are more active (reach fatigue sooner) and the appropriate method for analysis. A 4-channel EMG system, as compared to a 8-channel system, is desirable to be used along with a rehabilitation robot during physiotherapy sessions due to the following reasons:

1. The 4-channel system needs lower time for setting up the EMG sensors and preparing patients since it has lower number of electrodes which require attachment and measurement for accurate data collection.
2. When sensors are in proximity, the muscle activity data associated to one muscle may be picked up by a sensor which was not intended for that muscle. Therefore, limiting the number of electrodes which would be used in an experiment is preferable.
3. For collecting the muscle data precisely, there is a need to place the electrodes on an exact location over the muscle. Finding the exact position needs plenty of measurements rendering it time-consuming and prone to error. In addition, some muscles are not easily accessible and finding the desired positions for placing the electrodes may be difficult. Therefore, selection of more accessible muscles is preferable.

Due to the above reasons, only four muscles were selected for the primary trial.

The squat motion is used for simulating the sit-to-stand motion where the most active muscles are those that are responsible for knee flexion. Experimental trial resulted in selection of **Tibialis Anterior (TA – Muscle 4)**, **Gastrocnemius Medialis (GM – Muscle 3),** which are the bigger muscles and more accessible from the back and front of the shank for the primary trial. Furthermore, **Biceps Femoris (BF – Muscle 2)**, Quadriceps Femoris (Vastus Medialis), and Quadriceps Femoris (Vastus Lateralis) from the thigh muscles provided acceptable results. However, instead of the Vastus Medialis and Vastus Lateralis, **Rectus Femoris (RF – Muscle 1)** was selected since it is part of the same muscle group (Figure 35), and the participants reported most pain during the standup phase from the vicinity of this muscle.

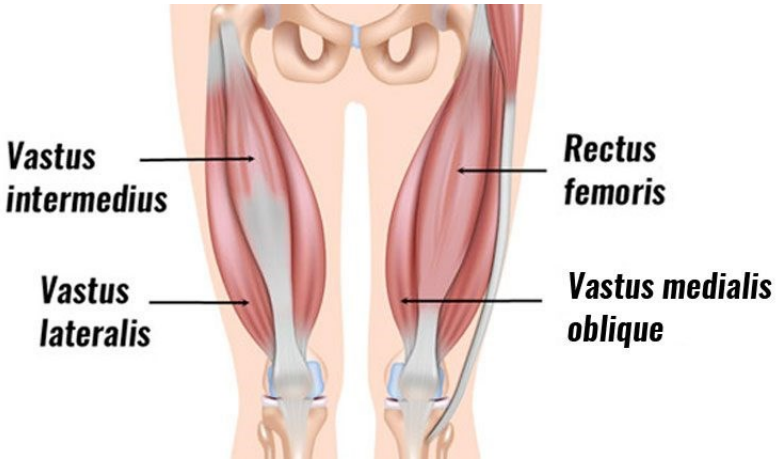


Figure 35: Vastus Lateralis, Vastus Medialis and Rectus Femoris muscle placement. Vastus intermedius is beneath the Rectus Femoris muscle [113]

Selection of these four muscles provided the opportunity to observe the muscle fatigue in both the thigh and the shank. Moreover, choosing muscles from both back and front of the shank and thigh give the opportunity to have a more inclusive observation of muscles from the entire lower limb. Monitoring these muscles could potentially show the order at which the muscles get tired.

3.10 Primary Trials

After completion of the experimental trials and finalizing the analysis method, the most appropriate muscles were selected for the experiment that can be used to effectively study muscle fatigue. A new set of experiments were conducted to achieve the following goals:

1. Validating that the selected muscles can be utilized for fatigue detection during an exergame involving the squat motion while collecting EMG, game, and perceived fatigue data.
2. Confirming the developed procedure to be effective (e.g. low noise), accessible (e.g. easy to setup), and efficient (e.g. timely preparation) for this exergame in a clinical setting.

3. Investigate the fatigue detection from the EMG data, and at the same time, prepare the data for further analyses such as machine learning.

Figure 36 illustrates the overall block diagram for this part of the research, where the main data is collected and analyzed to reach the objectives of the research. The utilization of the 4-channel EMG system distinguished the experimental trials from the current primary trials.

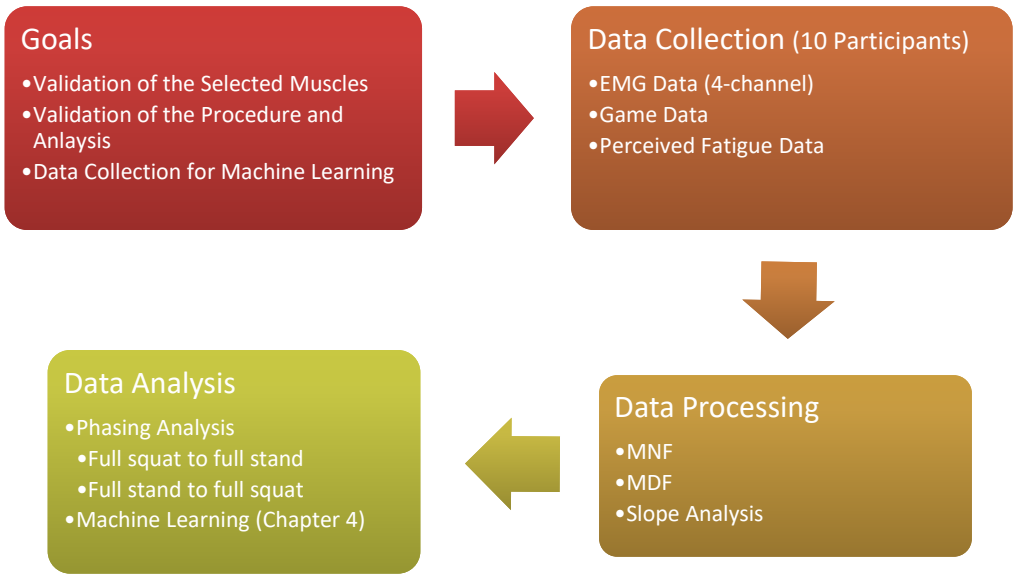


Figure 36: Schematic of the goals and steps taken for the primary trials

Other than the lower number of muscles, the primary trials involved the same procedure as the experimental trials. Before getting into the discussions and results associated with the primary trials, some details regarding the participants should be provided.

3.10.1 Participant Details

The sample population was composed of 10 healthy participants (aged 25 to 30, mean age of 26.9, mean body mass of 70 kg, mean stature of 172.1 m, and mean body mass index of 23.38 kg/m2). Table 1 provides the summary of the information about the participants who attended primary trial. All the procedures and eligibility criteria were similar to the experimental trial, as explained from Section 3.2 through 3.8.

Table 1: The information on the subjects who participated in the primary trial

	Gender	Height (cm)	Weight (Kg)	BMI (Kg/m^2)	Age	Maximum Allowable Heart Rate (bpm)	Average Heart Rate50-70% (bpm)
Participant 1	Female	164	63	23.4	26	194	97 - 135
Participant 2	Female	162	54	20.6	28	192	96 - 134
Participant 3	Male	178	73	23.0	27	193	96 - 135
Participant 4	Female	168	65	23.0	30	190	95 - 133
Participant 5	Male	183	85	25.4	27	193	96 - 135
Participant 6	Male	185	82	24.0	27	193	96 - 135
Participant 7	Female	160	50	19.5	25	195	97 - 136
Participant 8	Male	174	82	27.1	27	193	96 - 135
Participant 9	Male	170	77	26.6	26	194	97 - 135
Participant 10	Female	176	61	19.7	26	194	97 - 135

3.10.2 Procedural Validation and Expandability

Furthermore, the procedure developed for the EMG and exergame can be adopted by physiotherapists to recognize muscle fatigue with their patients. A sensor attachment guide has been developed for the appropriate muscles in Appendix E. The duration for skin preparation, measurements, and attachment of the electrodes was less than 5 minutes, indicating the improved efficiency of using four electrodes. The wire for the electrodes did not interfere with the sensors

during the squat motion. In addition to the ease of attachment, identifying the selected four muscles was straightforward. The uncomplicated nature of the procedure provides the means for expandability to a clinical setting.

3.10.3 EMG Data Analysis Results and Discussion

The phasing method provided means of analysis without considering the speed of each squat, which allows analysis of the muscle activity at times that the muscle is in motion. For comparison's sake, some examples of the discrete and moving window analysis results will be provided for the participants in the primary trials. However, the main focus will be on the phasing analysis as the results are essential to Chapter 4.

Figure 37 through Figure 48 are the results based on three analysis methods. Note that only a sample of the discrete window method and a sample of the moving window method have been shown (Figure 37 and Figure 38, respectively) here. The remaining results are attached to for the phasing method; however, all results can be found in this section as it is the basis for the next chapter.

Analysis Methods Comparison

The main advantage of the moving window analysis over discrete window analysis is the larger number of data points. The moving window analysis uses each squat data in multiple bins, unlike the discrete window analysis, which considers each of the squats only in one bin. The larger quantity of data points allows for better visualization of the data, which can be highlighted by comparing Figure 37 and Figure 38. For example, the data for participant 10 in Figure 37 is not as

clear as the same data in Figure 38 with the moving window analysis where it shows that the participant has had some low muscle activity in the middle but slightly recovered afterward. Note that both of these analysis methods show promising results for detecting fatigue considering the slope of MNF and MDF of muscles, but the discrete window analysis shows a more scattered data distribution.

The phasing method analysis for the same muscle (Figure 44) for participant 10 similarly shows the dip in the middle; however, the data is much smoother than the discrete window analysis and the moving window analysis and indicates a more clear distribution of the data even though it has a fewer number of data point in comparison to moving window analysis. The phasing method only considers the data at which the muscles are more active with respect to resting position while no repeated data is used. The moving window analysis accounts for the entire length of squats where it includes some resting times for muscles when the participant is in the standing up position. Moreover, in the phasing analysis with all the phases considered in Figure 44 for participant 10, muscle 4 of this participant does not show a positive slope in the line of the best fit. This is yet another indication that the phasing analysis provides a more accurate representation of the muscle activity compared to the discrete window analysis and the moving window analysis. The reporting from participants regarding their upward motion provided considerable improvement to the results and the development of the analysis method. Note that in the phasing data analysis in these analyses, it is possible that the participants perform a squat faster or slower than the game, which allows undesirable EMG data (related to the times that the participants find the motion difficult) to be analyzed. The phasing method has been adopted as the main analysis method for this study due to the abovementioned reasons.

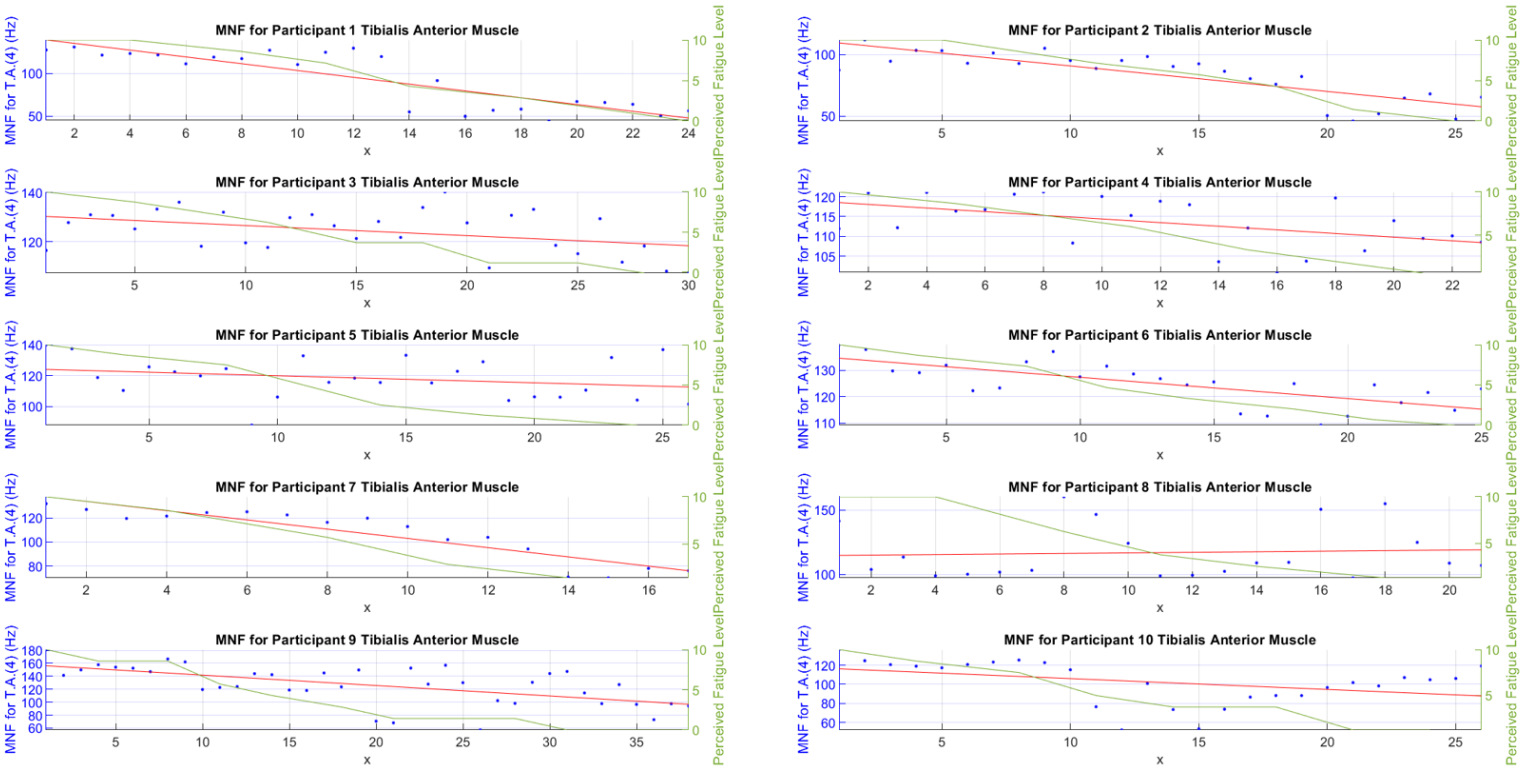


Figure 37: MNF of Tibialis Anterior Muscle (Muscle 4), Discrete Window Analysis (Step=9s, and x represents the bin number)

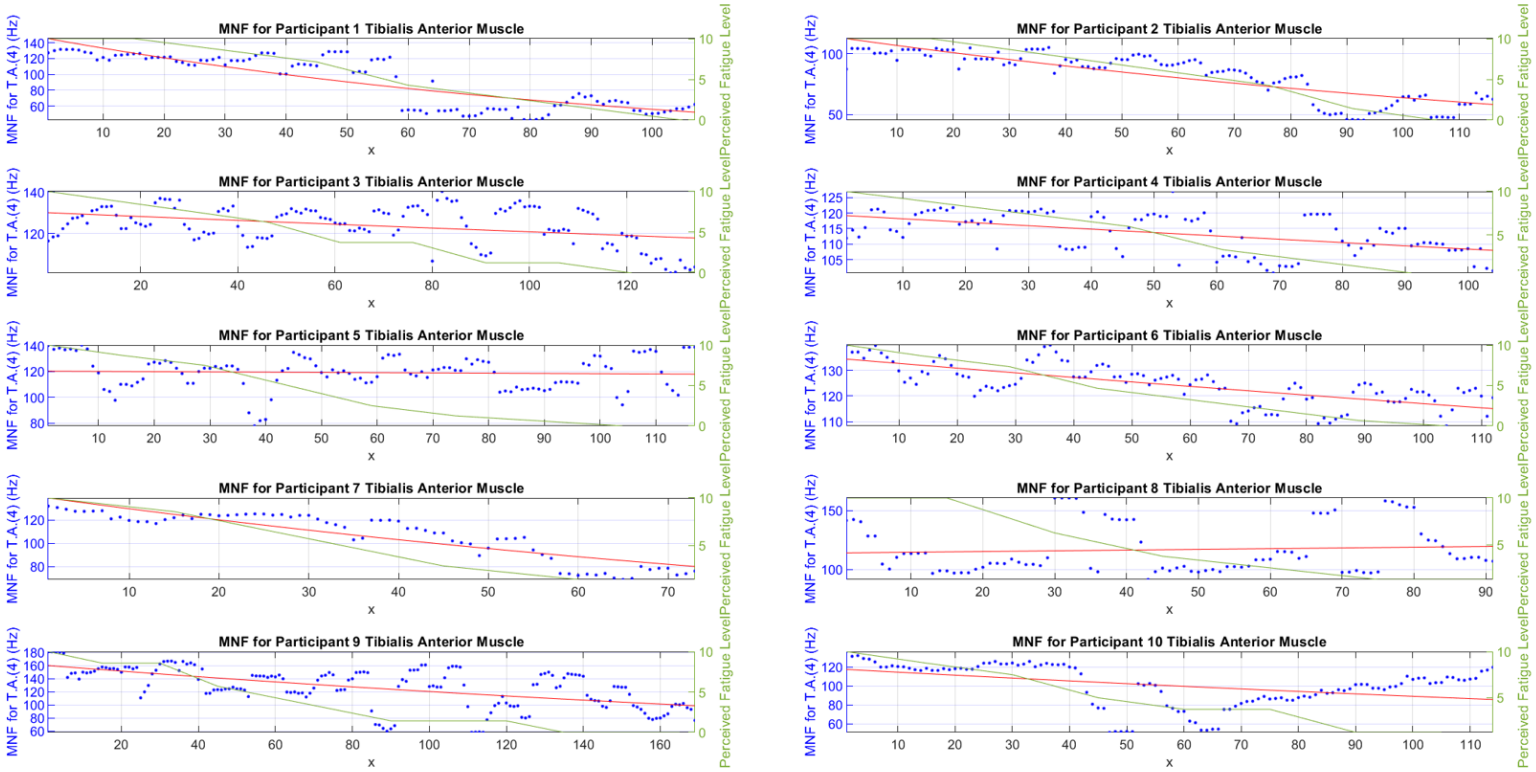


Figure 38: MNF of Tibialis Anterior Muscle (Muscle 4), Moving Window Analysis (Step=9s, Moving Window=2s, and x represents the bin number)

Gender and Muscle Activity

All participants exhibit an expected downward trend in their muscle activity in at least one of their muscles. It is believed that the main reason is the unconscious change in the utilization of the muscles by changing their motion, the center of gravity, or the speed of the exercise. Furthermore, the downward trend in the MNF and MDF of the experiments is not correct in at least one of the muscles in participants 3, 5, 6, 8, 9, and 10. The male gender is a common characteristic in these participants except participant 10, who was a female. Even though there have been no studies to confirm this point, not all muscles in the more muscular male anatomy become tired during exercise. Further investigation into participant 10 indicates that her BMI is low and could potentially mean an active participant. Figure 39 and Figure 40 represent the separated male and female muscle activity graphs which indicates the male muscle activity does not follow the expected decrease in all muscles.

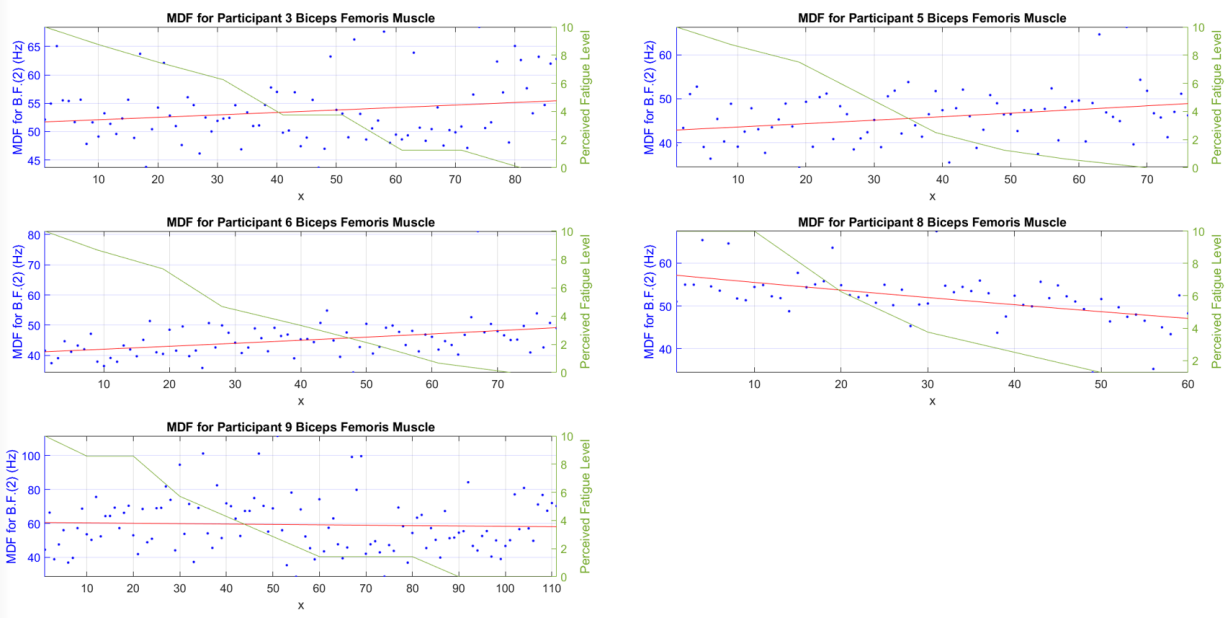


Figure 39: MDF of Bicep Femoris Muscle for Male Participants (Phasing Method Analysis – from the beginning of full squat to the beginning of full stand position- x represents the bin number)

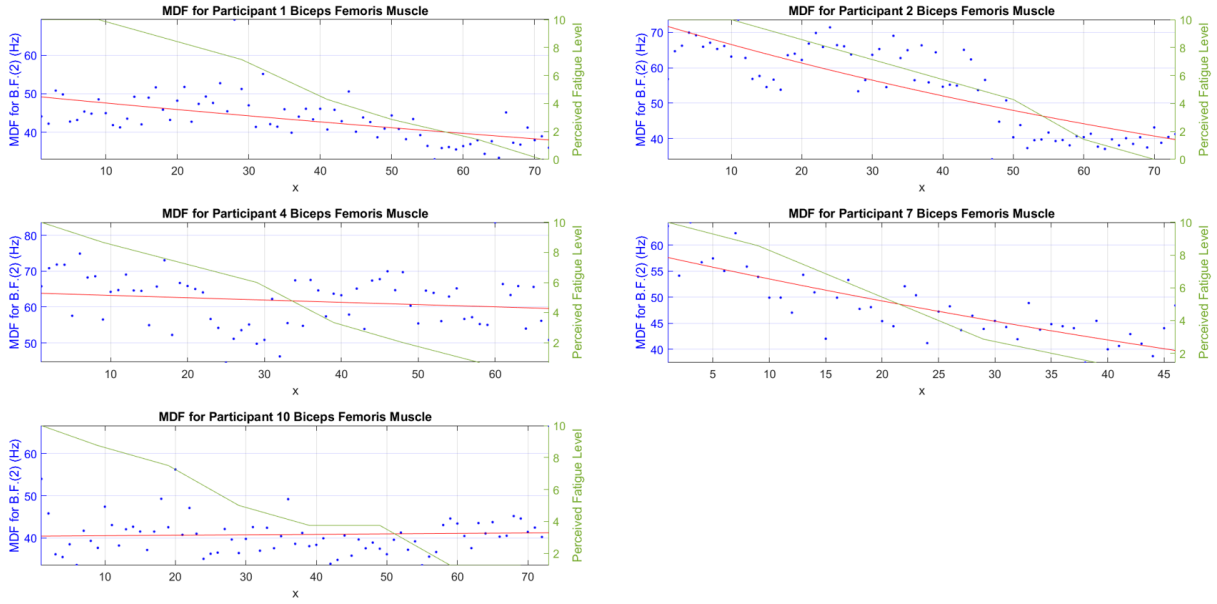


Figure 40: MDF of Bicep Femoris Muscle for Female Participants (Phasing Method Analysis – from the beginning of full squat to the beginning of full stand position- x represents the bin number)

Phasing Approach Comparison

Two main phasing analysis approaches have been shown in Figure 41 through Figure 48: full squat and standing up, full stand and sitting down. Both methods provided reliable data for individual fatigue detection. It is realized that most of the female participants showed a downward trend in their MNF (and MDF) analyses for both approaches. On the other hand, some male participants show that their muscles rest during one of the phases. For example, in muscle 4, participant 6 shows a downward trend in the standing up phase (Figure 44); however, the trend in the sitting down phase (Figure 48) is almost neutral. One can argue that participant 6 in Figure 48 is showing signs of fatigue in the middle of the dataset, which is consistent with the observation (slight recovery after the fatigue has begun) with participant 10, muscles 2, and 4, among others.

The current data set has provided a good set of observations; however, no conclusion can be drawn due to the small set of datasets. This lack of data is evident since no trend can be detected for

generalization. Table 2 represents the muscles Decreasing trends as desired, from the beginning of full squat to the beginning of standing up phase and the beginning of full stand to full sitting down phase. As it can be seen, both approaches are representing a good result, but the full squat to full stand phase is showing more decreasing trends in participants' muscle activity.

Table 2: Comparing the activity muscles which shows a decreasing trend in both Full squat to Full stand phase and Full stand to Full squat phase

Muscle Name	Full squat to Full stand	Full stand to Full squat
Rectus Femoris	1,2,4,5,6,7,10	1,2,4,5,7,8,10
Bicep Femoris	1,2,4,7,8,10	1,2,4,6,7,9
Gastrocnemius Medialis	1,2,4,7,8,10	1,2,4,7,9,10
Tibialis Anterior	1,2,4,5,6,7,9,10	1,2,3,5,7,9,10

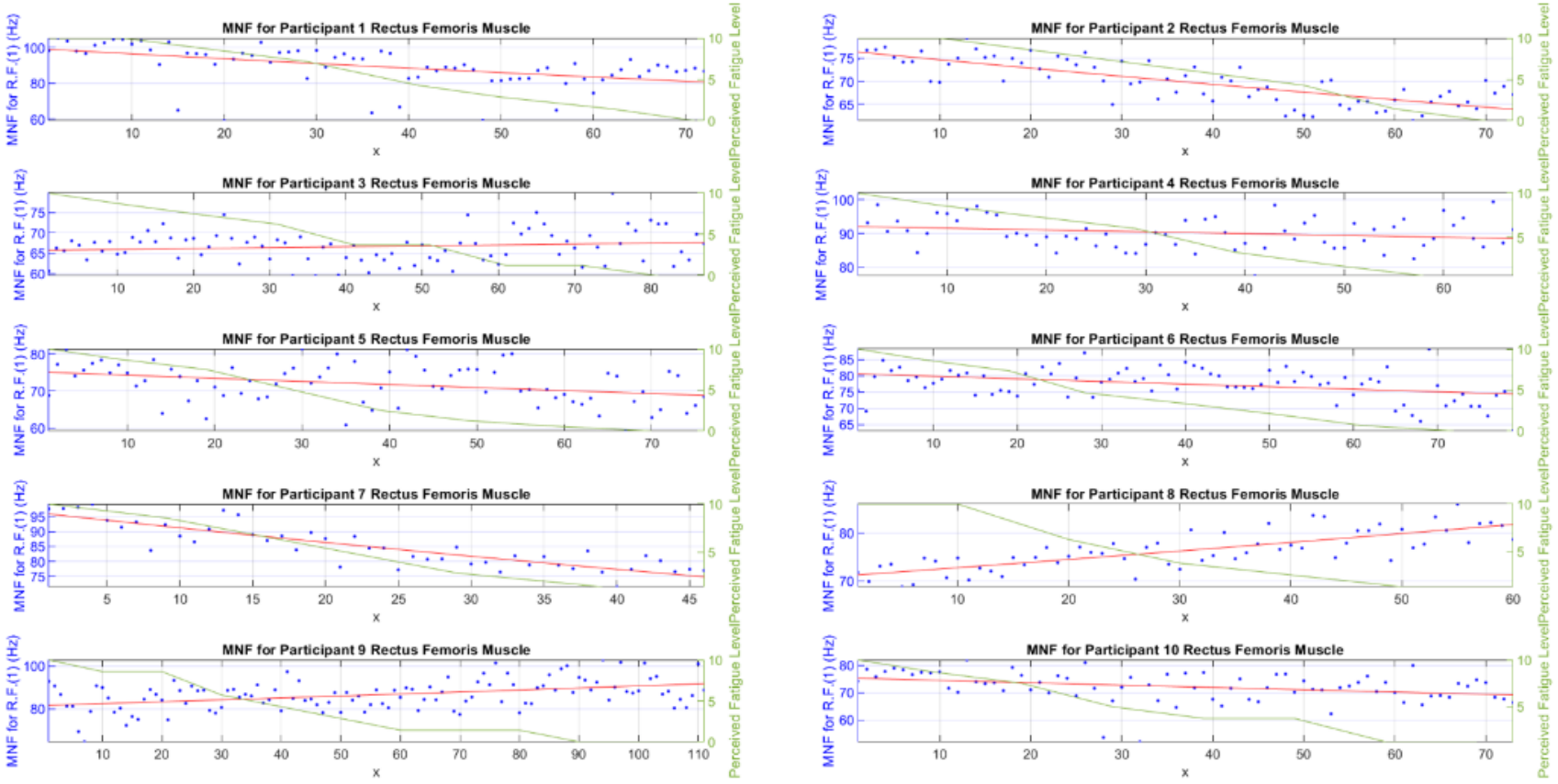


Figure 41: MNF of Rectus Femoris Muscle (Muscle 1), Phasing Method (from the beginning of full squat to the beginning of full stand position- x represents the bin number)

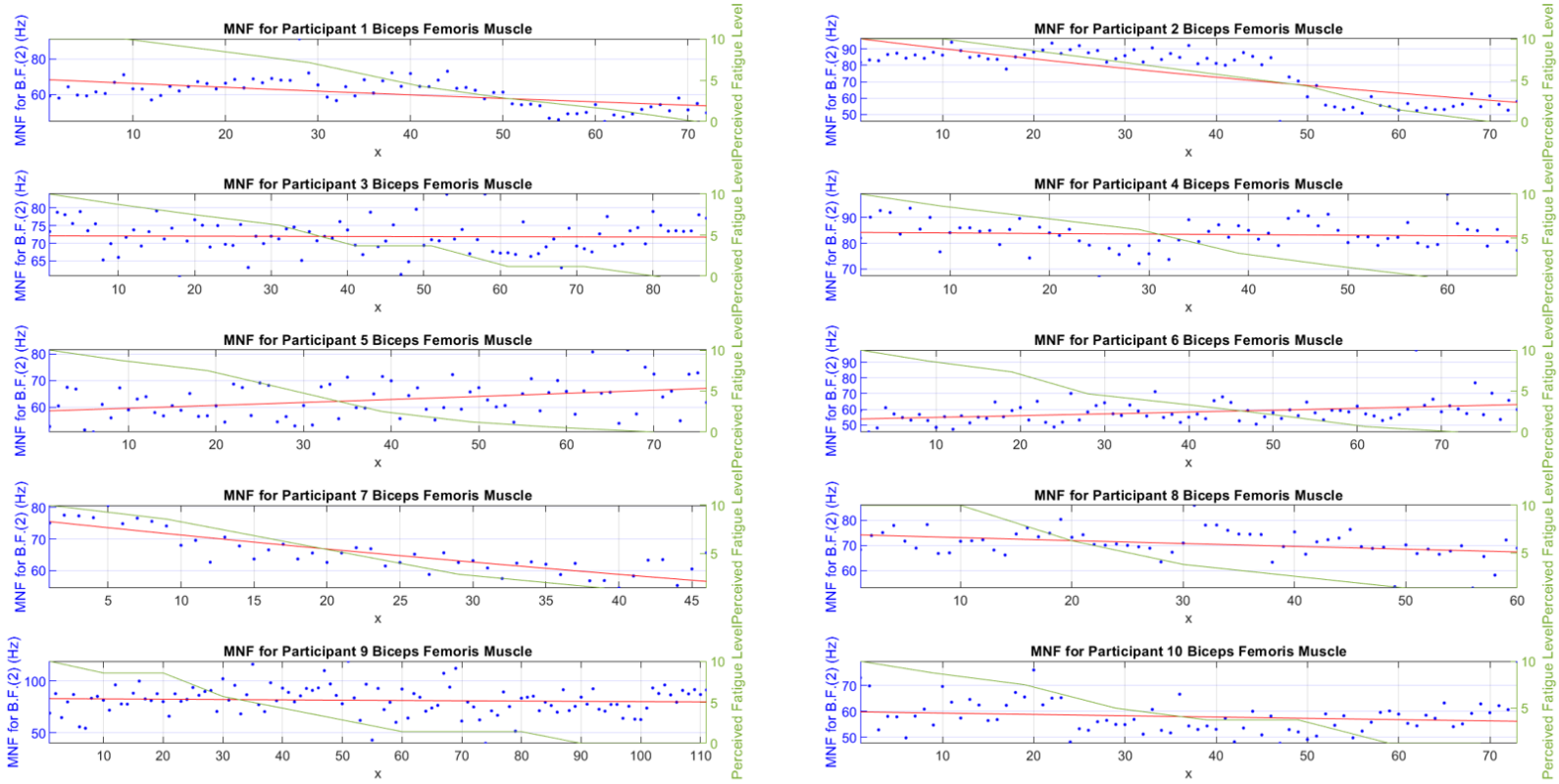


Figure 42: MNF of Bicep Femoris Muscle (Muscle 2), Phasing Method (from the beginning of full squat to the beginning of full stand position- x represents the bin number)

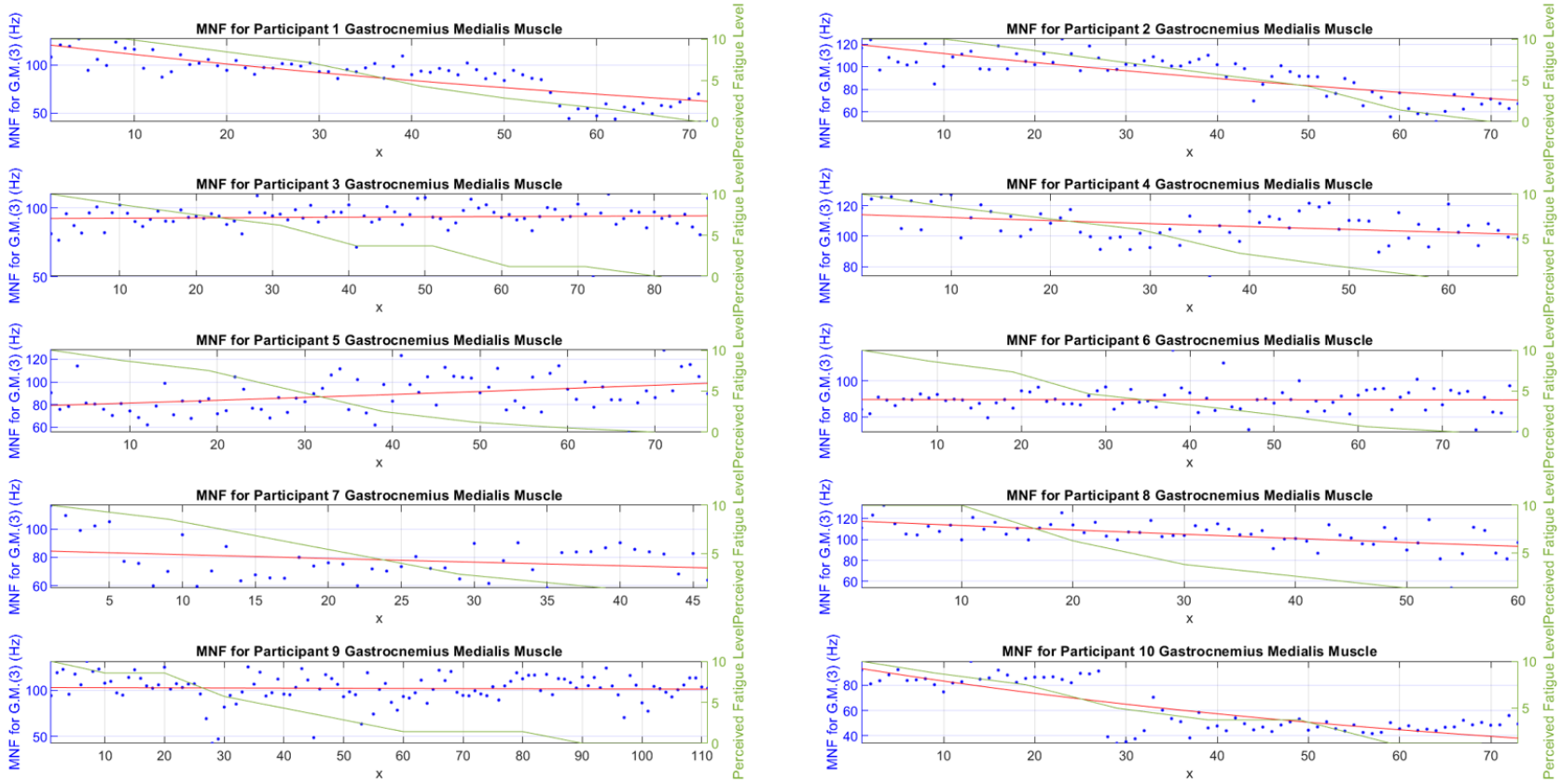


Figure 43: MNF of Gastrocnemius Medialis Muscle (Muscle 3), Phasing method (from the beginning of full squat to the beginning of full stand position- x represents the bin number)

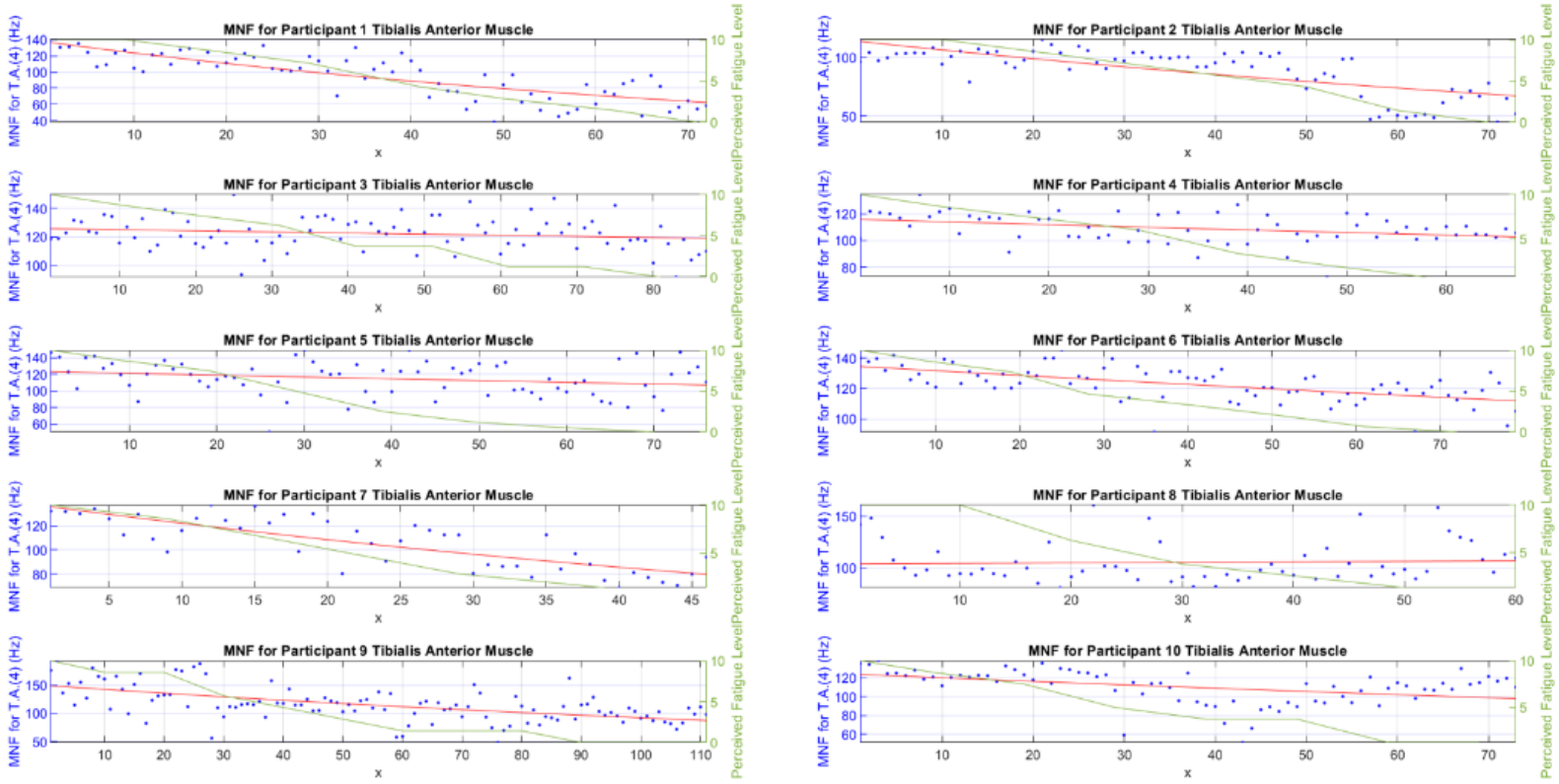


Figure 44: MNF of Tibialis Anterior Muscle (Muscle 4), Phasing Method (from the beginning of full squat to the beginning of full stand position- x represents the bin number)

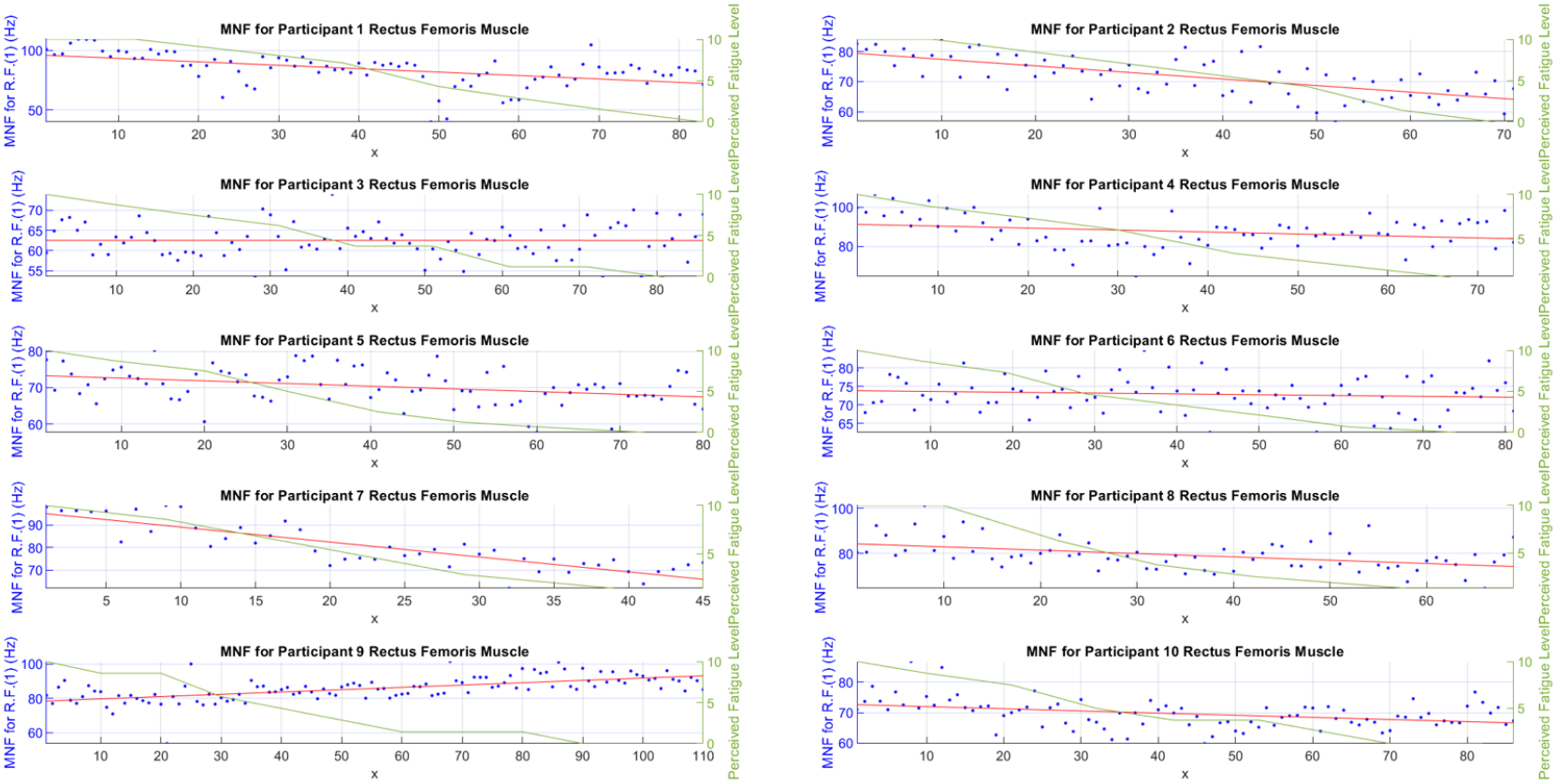


Figure 45: MNF of Rectus Femoris Muscle (Muscle 1), phasing method (full standing to the beginning of full squat- x represents the bin number)

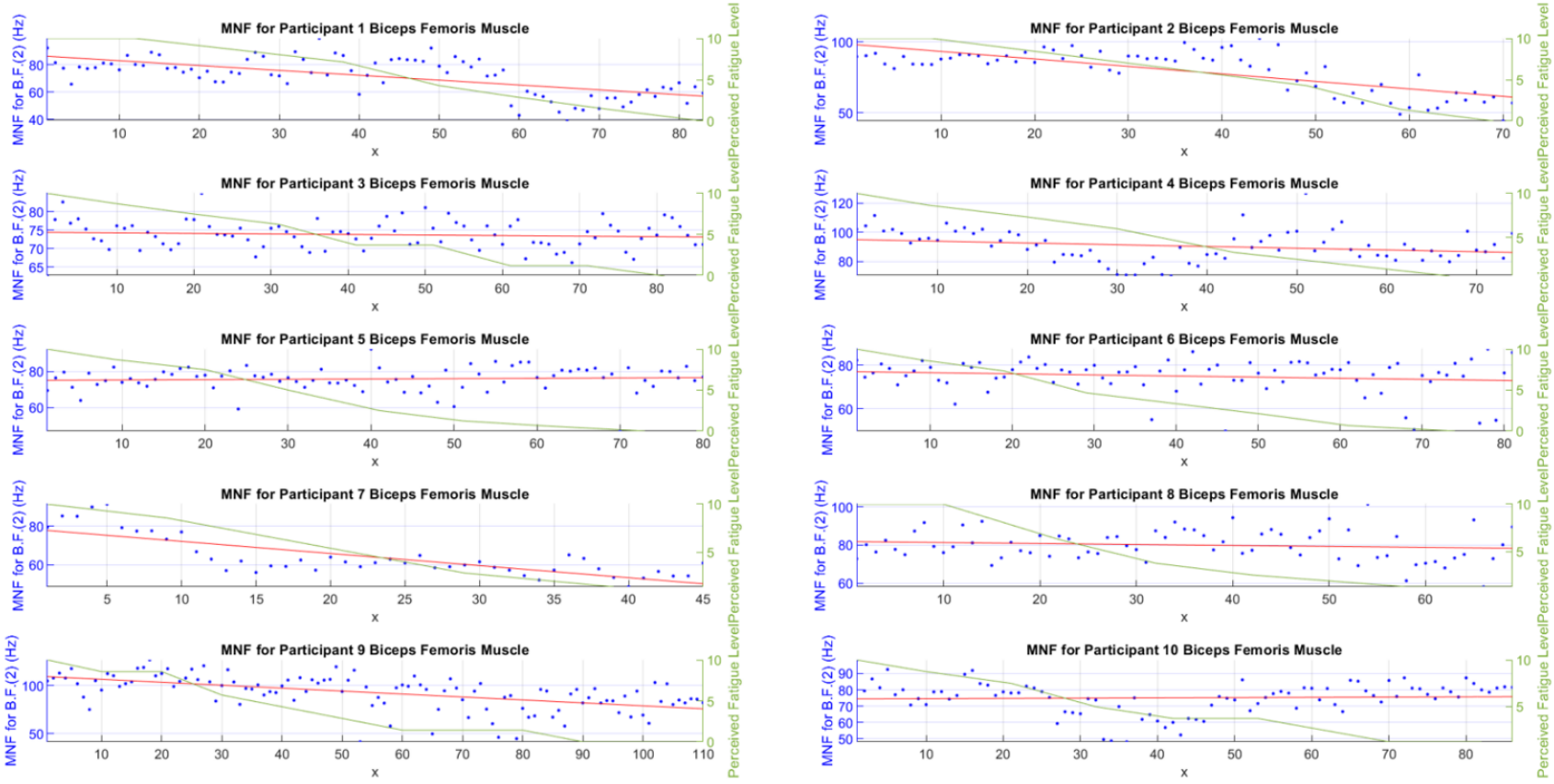


Figure 46: MNF of Bicep Femoris (Muscle 2), Phasing Method (full standing to the beginning of full squat- x represents the bin number)

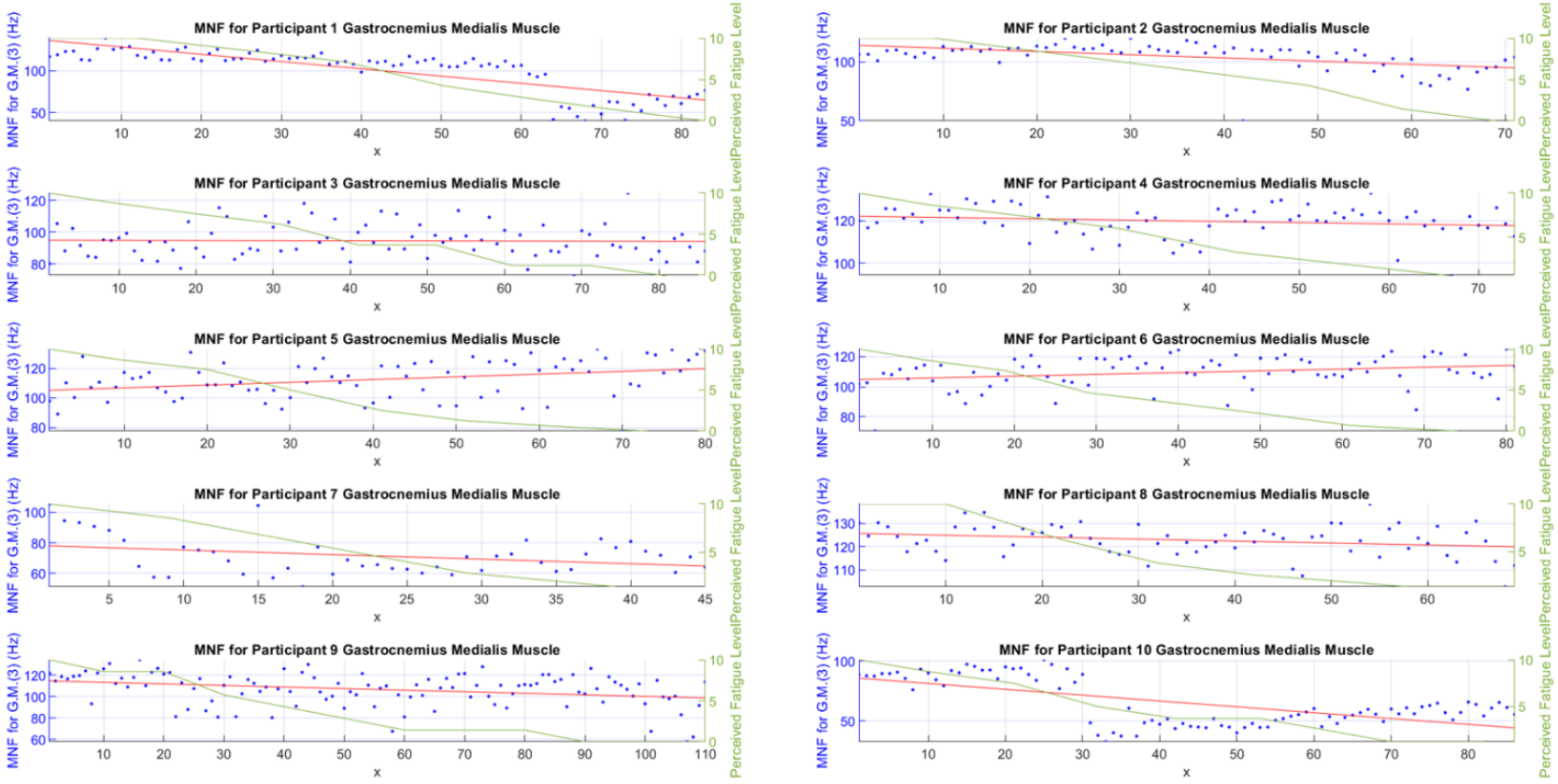


Figure 47: MNF of Gastrocnemius Medialis Muscle (Muscle 3), Phasing Method (full standing to the beginning of full squat- x represents the bin number)

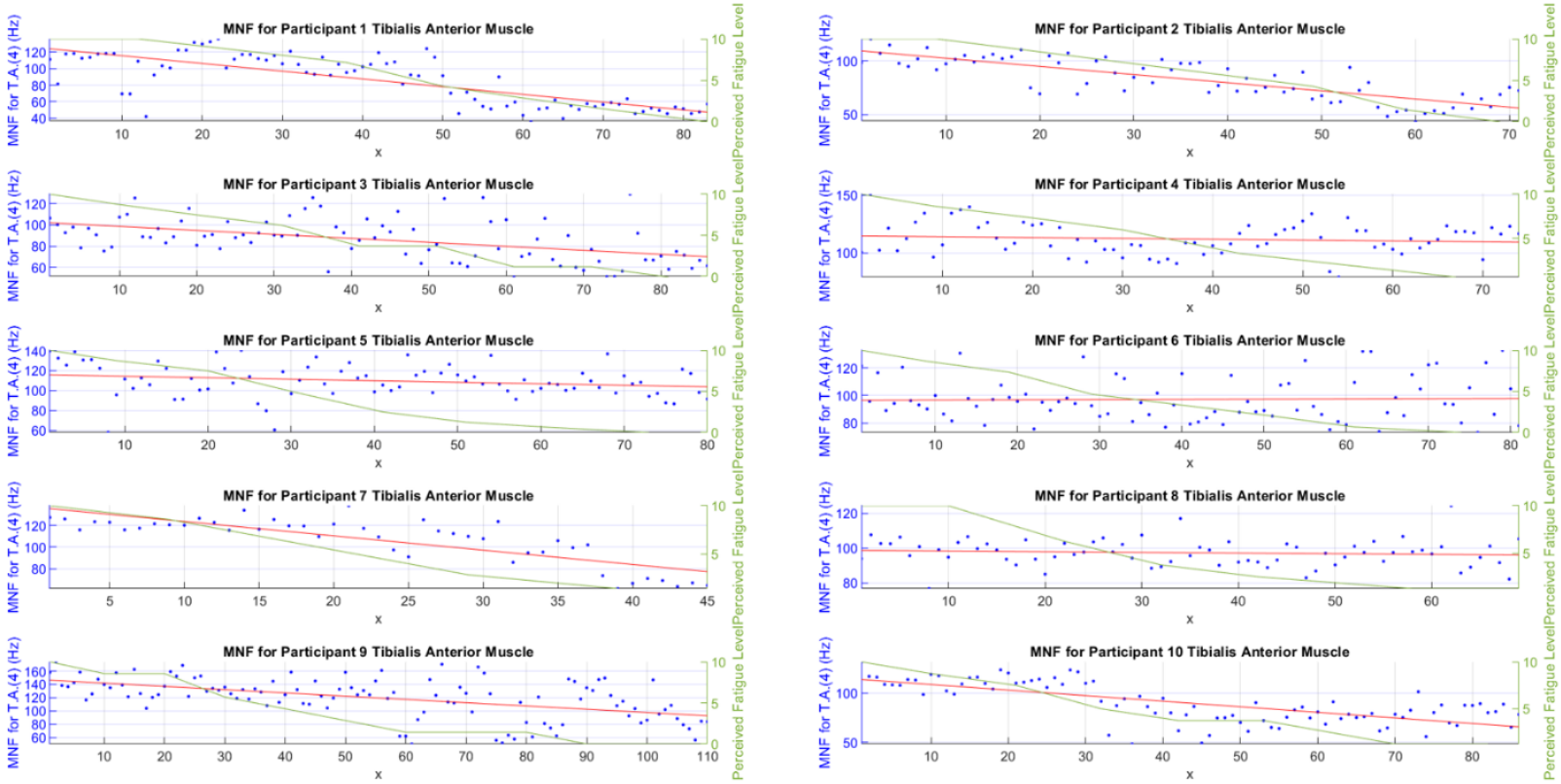


Figure 48: MNF of Tibialis Anterior Muscle (Muscle 4), Phasing Method (full standing to the beginning of full squat- x represents the bin number)

3.10.4 Summary of major findings

- A minimal adequate set of muscles were identified for EMG data collection in the experimental trial: Rectus Femoris (RF) and Biceps Femoris (BF) on the thigh, Gastrocnemius Medialis (GM) and Tibialis Anterior (TA) from the shank.
- Among the developed analysis methods in the experimental trial, the phasing method was a more reliable method. It allowed analyzing the muscle activity during different phases of squat when the muscles are in motion and showed a smoother trend compared to other analysis methods.
- The expected downward trends were exhibited in at least one of the muscles of the participants, but it was not reliable as it was not consistent in all muscles. This inconsistency in results led the research to develop different machine learning algorithms as well as the findings on the impact of participant information on physical fatigue.
- The developed procedure in the experimental trial for preparing and testing participants were examined during the primary trial. It was proved that this procedure can be expanded to be used in clinical applications as long as all possible steps for preventing noisy EMG data were taken.

Chapter 4

Muscle Fatigue Predictions using Machine Learning

4.1 Introduction

The main goal of this work is to prepare a tool which assists in recognizing muscle fatigue when a stroke patient is using the rehabilitation robots. As discussed in Chapter 1, it is essential to recognize muscle fatigue because (i) continuing exercise in fatigue conditions would harm the patient, (ii) adjusting the exercise parameters could help in faster recovery, (iii) excessive fatigue could unconsciously alter the target muscles that the patient utilizes for the exercise. For reaching this goal, a method is needed which recognizes the fatigue (and potentially the fatigue level) in the muscles that are targeted by the robot being efficient and easy to use in a way that it can be employed for consecutive physiotherapy sessions. The analysis method of chapter 3 can help recognize the start period of muscle fatigue in the patient after a physiotherapy session. However, there were some issues that led us to use machine learning for the rest of the project:

1) The real-life applications, such as a robot-assisted physiotherapy session, require fatigue detection as soon as the muscle is fatigued. Detection of fatigue onset is necessary for

various stages of rehabilitation to prevent depreciating damages such as muscle inflammation [114].

2) It is desirable to utilize muscle fatigue level while the patient undergoes robot-assisted rehabilitation since appropriate assistive measures (e.g., changing the resistive force) can be taken to elongate the exercise. No consistent trend for determining fatigue level from the muscle activity data was found. The method developed in this study attempts to estimate the immediate fatigue level during the therapy, which allows adjustment of the rehabilitation parameters of the robots.
3) The results from Chapter 3 did not indicate a consistent trend among the muscle activity levels for different muscles. A lack of negative slope in every muscle activity graph is not necessarily an indication of muscle fatigue. Therefore, there is a need for a model which provides a single conclusion considering all the muscle activity information.
4) As discussed in 3.10.3, participants gender affected their muscle activity and, consequently, the fatigue level. This observation led to the belief that there are various factors that may affect muscle fatigue. Different parameters are considered that seem to be related to participants' mental and physical fatigue. The muscle activity, the participant information, and the game data should be considered simultaneously for effective fatigue detection.

Considering the above reasons, the versatility of machine learning allows for finding a unified method that accounts for all the data at the same time. Machine learning has begun to appear in the medical and engineering applications such as diagnosis and prediction [115]–[117]. Chan et al. [117] developed a cardiac arrest detection model for smart devices using recordings from agonal breathing samples (recorded from 9-1-1 calls). Furthermore, gestational diabetes during pregnancy

has been successfully predicted using a machine learning model developed by researchers at the Weizmann Institute of Science [115]. This chapter aims to develop a supervised machine learning model that captures the patient information and the activity levels to detect the fatigue onset using classification and perform fatigue level recognition by regression. Different supervised machine learning algorithms were investigated to determine the best method for the purpose of this work. The method section in this chapter contains the necessary procedures and associated information that was utilized to develop both the classification and regression models. A thorough analysis of the findings will be provided in the results and discussion section, along with appropriate parameters for the final models.

For future experiments, the classifier and regressor with the best performance will be used in a real-time experiment with new participants. For this purpose, there is a need to add an assistive controller to the game which can change intensity of the exercise during the experiment. For both algorithms the required input information would only be the participant information; the rest of the features will be calculated as the experiment is run. The results of new real-time experiments will show the benefits of fatigue prediction such as increasing the duration of the exercise.

Figure 49 represents the summary of the steps taken in this research to develop a suitable machine learning model for the two objectives. The features were labeled based on the objectives.

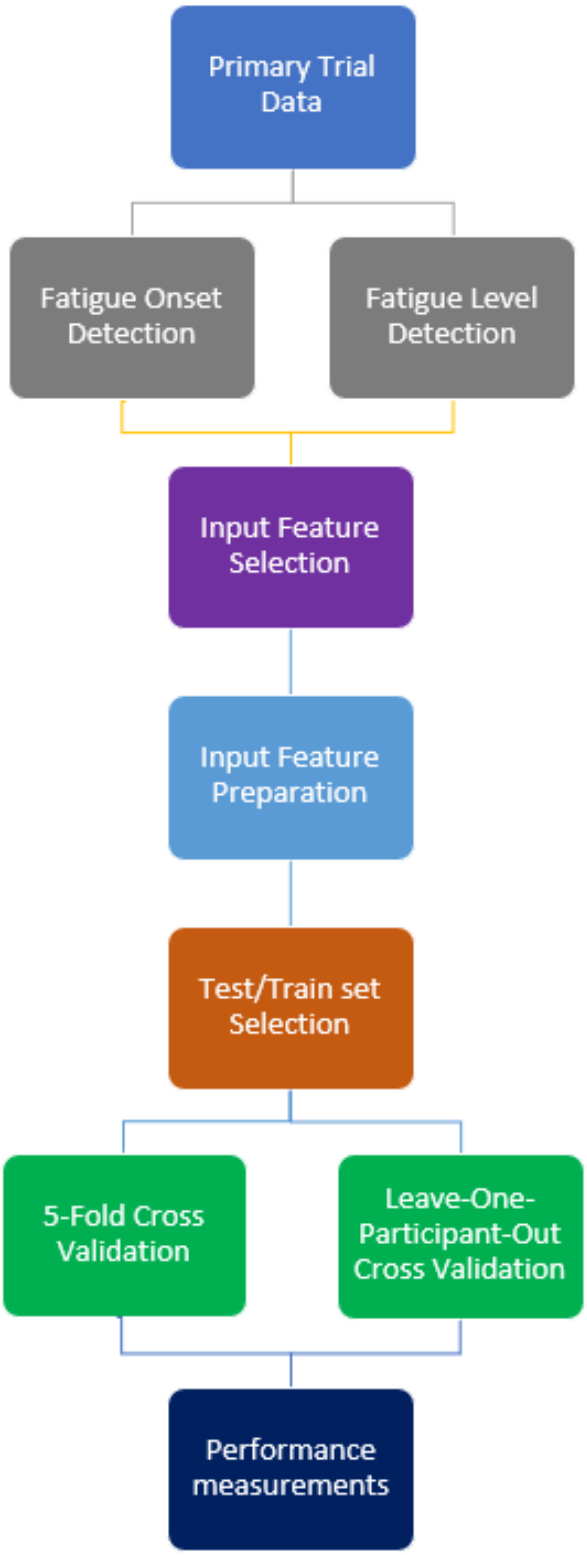


Figure 49: Summary of the steps taken for developing machine learning models for onset fatigue detection and fatigue level detection

4.2 Method

Figure 50 represents the overall steps that were taken to prepare dataset and analyze them in order to detect fatigue onset and quantify fatigue level. According to the Section 2.5.11, the following steps are required for developing the machine learning model:

1. **Defining objectives**: Taking a broad look at the problem leads to defining the objectives of the research. Here, fatigue onset detection and level of fatigue prediction are defined as the research objectives.
2. **Selecting input features**: For the development of machine learning models, some suitable input features should be selected as the algorithms utilize them in the learning process. In this research, input features are selected from the preprocessed EMG data (muscle activity), the game data, and the participants data.
3. **Input feature preparation**: Firstly, the selected input features need to be visualized as a whole to ensure machine learning is appropriate by taking the following two steps: 1. Inspecting the distribution of each feature and ensuring the data behaves expectedly 2. Examining the correlation among the features to ensure no reliable relationship exists among the features. Secondly, there is a need to ensure that all the features are numerical (e.g., “Male” and “Female” should be replaced with numerical equivalent) and equally contribute to the machine learning algorithms (e.g., a feature with a range of 0-100 skews the algorithms more compared to a feature with a range 0-1). Feature scaling and encoding were performed in this work as gender is a selected input feature for our models.

4. **Grid Search**: Grid search is one of the most important steps which improves the performance of machine learning models. During the training process, grid search performs a sensitivity analysis on the parameters associated with the operation of the models (also known as **Hyperparameters**) based on a specified metric.
5. **Test/Train set Selection**: After preparing input features and setting up the grid search, it is necessary to select a suitable train and test set from the entire dataset to ensure the calculated metrics represent the dataset. The scheme used for setting up the test/train set can be repeated across the entire dataset to ensure a generalized model. This repetition is also known as **cross-validation**. Two different selection was chosen to be used in this work for both objectives: (i) 5-Fold cross-validation, (ii) Leave-one-participant-out cross-validation, which will be explained further in Section 4.2.5.
6. **Performance Measurement**: After applying all these steps and running different classifiers and regressors, the performance of the models should be measured and compared to find out the best model for this research. The algorithm validation metrics are discussed in 2.5.7.

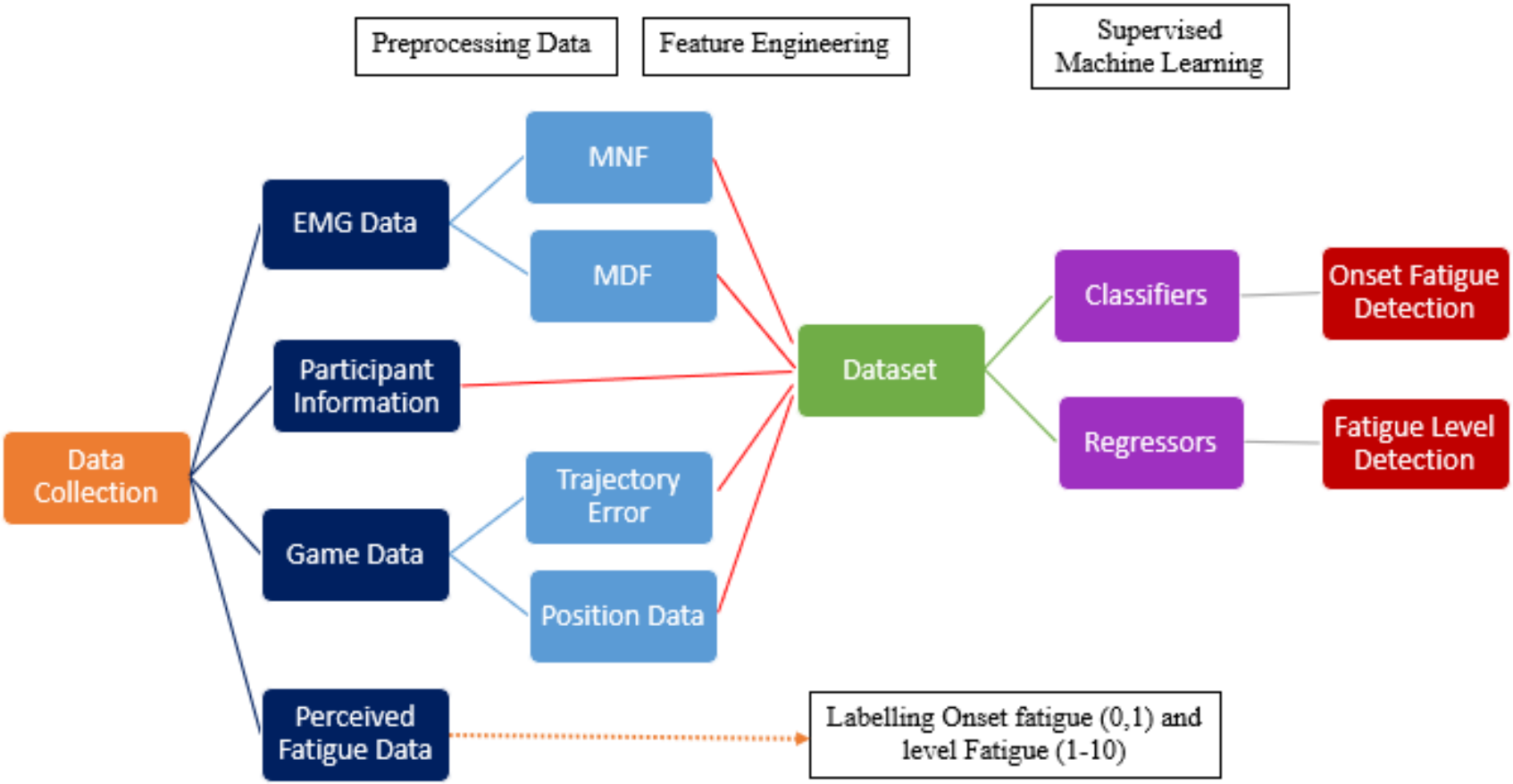


Figure 50: The block diagram of the stages taken for the development of classification and regression model

4.2.1 Input Features

In this section, the list of input features selected from EMG data, game data, and perceived fatigue data is provided, along with the necessary explanations. Thirty-eight features were selected to use as input features for developing the machine learning models. The onset fatigue detection dataset was labeled in a binary fashion, whereas the fatigue level detection was labeled numerically based on the perceived fatigue data.

Bin

As explained in the chapter 3, each participant performs several squats while they are playing the game. Individual squat signals and associated phases were recognized from the game data as it provides the position data of participants doing squats, and then these data were synchronized with EMG data to figure out each squat data as part of the preprocessing, which is explained in section 3.9.2.3. Input features were grouped into bins corresponding to a specified number of squats; the phase and number of squats are selected based on the study at hand. Each set of squats are referred to as a bin (e.g., a bin size 3 means the data for three squats at the duration of a phase under the study will be used)

Bin Start

The bin start indicates the time stamp of the appropriate phase of the first squat in the bin. This data was extracted from synchronizing the position data extracted from the game data and EMG data, which is explained in 3.9.2.3.

Bin End

The Bin End feature shows the time at which the appropriate phase of the last squat is finished. The reference time for both Bin Start and Bin End features is the start of the game.

Average MNF

As explained in Section 2.4, MNF is one of the most important features for detecting muscle fatigue [39], [118]. This feature is the mean of the MNF values for all the appropriate phases in the bin. All four muscles in this study have their respective features in the dataset: AvgMNFm1, AvgMNFm2, AvgMNFm3, AvgMNFm4.

Average MDF

MDF is also one of the most important fatigue detection features vastly used in the literature[39], [119]. Similar to MNF, each phase of the squat has its own MDF value, which is calculated in the preprocessing part. Based on the size and phase of the bin, the average MDF value is calculated for each bin. All muscles are allocated with their own features in the dataset: AvgMDFm1, AvgMDFm2, AvgMDFm3, AvgMDFm4.

Slope and Intercept for MNF and MDF

The trend of the MNF and MDF can be captured by monitoring the slope and intercept of a fitted line for each bin. The slope and intercept values for each fitted line provide an independent set of values that could show fatigue. The fitting operation is applied for the data on all the muscles and the following features: MNFSlopem1, MNFInterceptm1, MNFSlopem2, MNFInterceptm2, MNFSlopem3, MNFInterceptm3, MNFSlopem4, MNFInterceptm4, MDFSlopem1, MDFInterceptm1, MDFSlopem2, MDFInterceptm2, MDFSlopem3, MDFInterceptm3, MDFSlopem4, MDFInterceptm4.

Gender

After processing the data in chapter 3, the results indicate that fatigue act differently when it comes to different genders. As discussed in section 3.10.3, all muscles in the male are not fatigued at the same time and showed a different trend in various muscles, but in most of our female participants,

all muscles were fatigued at the end of the experiment. This result shows that it is important to consider gender as one of the input features for developing a machine learning model. In this column, the gender of the participants is entered as F for Female and M for Male.

Height

For each participant, there is a column which shows their height. Before the test, each participant's height is measured by asking them to stand against the wall, and the examiner places a flat object on their head and marks the wall accordingly. Then their height is measured by measuring the distance of the marked line to the ground.

Weight

A digital scale was placed on a flat surface in order to measure the weight of each participant.

BMI

Body Mass Index indicates a scale for body size and is calculated from dividing the weight of a person in kilograms by the square of the height of that person in meters [120]. Even though BMI has been widely used for categorizing obesity, the experiment performed by Wellens et al. [121] concluded that the correlation between body fat and BMI varies based on gender and physical characteristics. The correlation between age and BMI has also shown difficulties in diagnosing obesity [122]. However, along with weight, height, and age, BMI is a helpful input for comparison between different participants.

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

Leg Length

Another feature that is found effective in the squat movement is the length of the leg, as squat motion requires the force of muscles and the torque generated from the joints. This feature measures the total length of the hip to the ankle of each participant.

Thigh Length

In addition to what was discussed for leg length, in the experimental trial, the consensus was that the thigh muscles were among the first muscles which started to feel fatigued. Therefore, having a thigh length as an input feature is helpful. The thigh length is the measurement from the hip to the knee for each participant.

Age

The physical condition of individuals is affected by age [123]–[126]. Having a diverse age range would be great to see the effects of age on muscle fatigue. But unfortunately, because of the COVID-19 pandemic and quarantine situation, the participants were in the same age range from 25 to 30.

Phase

As explained in Section 3.9.2.3, there is a need to consider different muscles activities and signals based on the upward or downward motion. Depending on the upward (1) or downward (0) phase, appropriate intervals within the bins are used for calculating the EMG and game-related features.

Squat Depth

Another helpful information for the machine learning algorithms, which is realized from game data, is the participants' squat length, as it is expected to decrease when the participant becomes fatigued. As explained in the Section 3.8.4, each participant performed one squat in front of the camera to calibrate the full-standing (1) and full-squat (0) positions before starting the experiment. The relative length of each squat is the difference between the two extremes of the calibrated squat position. Ideally, each squat depth should be 1; however, as time goes by, the depth decreases as the participant gets fatigued (e.g., after 3 minutes of the experiment, participant 4 only completed

80% of the original squat length). This information has been averaged out for each bin and placed as an input feature column in the table.

Average Error

In the game, there is a predetermined trajectory which is expected to be followed by the participants in order to collect most coins. As they get tired, there may be a chance of missing the desired trajectory. The difference between the trajectory that participants follow, and the predetermined trajectory is calculated and represented as Error. The average error is calculated and used as practical information for the input feature table for each bin.

Max Error

As the participant gets more tired, it is expected that the game performance decreases, and the trajectory error increases. Observing the maximum error of each bin can be helpful, and therefore, it was determined for each bin and used as another input feature.

Fatigue

The last column of the input feature table is called target. When only fatigue detection is studied, classification is utilized where the target value indicates whether the corresponding features in the row refer to fatigued (1) or not fatigued (0). Labeling of the data was based on the perceived fatigue data of more than 6 in a 1-10 (least to most fatigued) range. On the other hand, if the fatigue level is the point of the study, regression analysis is performed based on the perceived fatigue level.

4.2.2 Data Visualization

After preparing the input feature table for the algorithm, reviewing the data can be useful as it can give a better understanding of the trends, the relationships, or potential problems in the data. For classification, the distribution of each input feature with respect to the target values is used.

Moreover, the correlation between the features is helpful to be studied for both classification and regression. Here, the two different approaches that were used for getting a better vision from the data is explained.

4.2.2.1 Input Feature Distribution

In this section, the distribution of each input feature with the fatigue/not-fatigue categories are plotted and can represent various attributions of the features with the labels. For instance, it will show that how much of the data is labeled as each of the categories.

Figure 51 is the histogram of the Bin End feature for the entire dataset (both phases) isolated based on the label. The distribution of the not-fatigued data and the fatigued data are a good representation of the experiment in a way that the likely chance of fatigue is greater towards the third minute of the experiment, whereas the not-fatigued is expected to be skewed to the right. The small blue tail of the histogram from the 150 to 180 second shows that a small number of instances had passed the 150^{th} second without feeling fatigued. Further investigation into participants indicated that only participant 4 had not become tired until the 180^{th} second of the game.

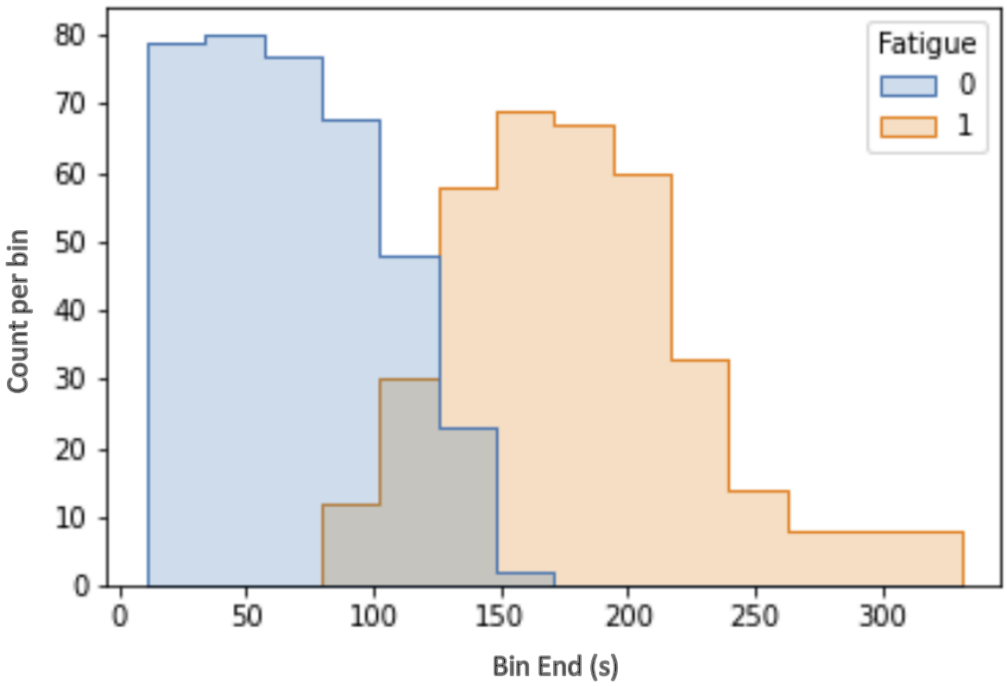


Figure 51: Bin End distribution histogram for all the participants' data distinguishing the data into fatigued and not fatigued

In another example, Figure 52 indicates that there are more bins with lower MDF values for the Muscle 4 when it is fatigued. This verifies the findings from the literature indicating that muscle activity (represented by MDF) is decreased when a muscle is fatigued, therefore, the data follows a trend that is expected. The machine learning analysis will assist in including all the features for classification.

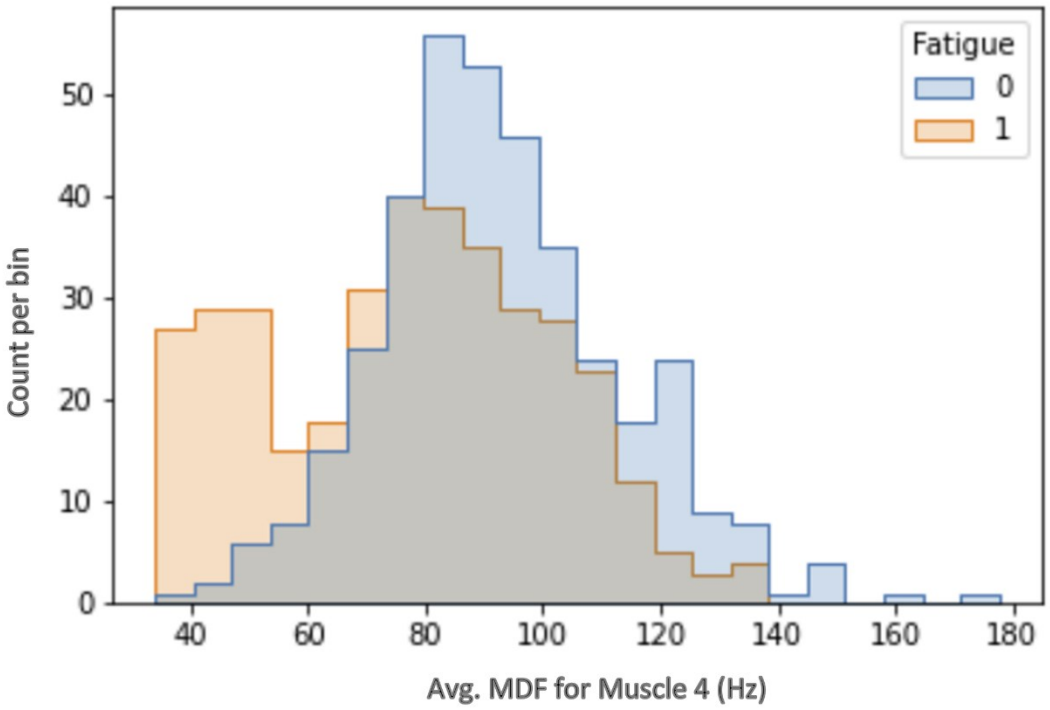


Figure 52: Average MDF for muscle 4 distribution histogram of all participants' data separating the data into fatigued and not fatigued

4.2.2.2 Input Feature Correlation

Knowing the correlation between different input features can help to find out if there is any relationship between input features and if so, how strong is the correlation between them. The correlations between features may affect the training process, and these correlations need to be taken into account when interpreting the results. The Correlation matrix between the input features is shown on Figure 53 where linearity between the variables is measured (referred to as Pearson r Correlation). It is expected that the correlation between the features to be near zero as the features expected to be independent. The highest correlations occur between the slopes and intercepts of the fitted lines on MNF and MDF data of a muscle.

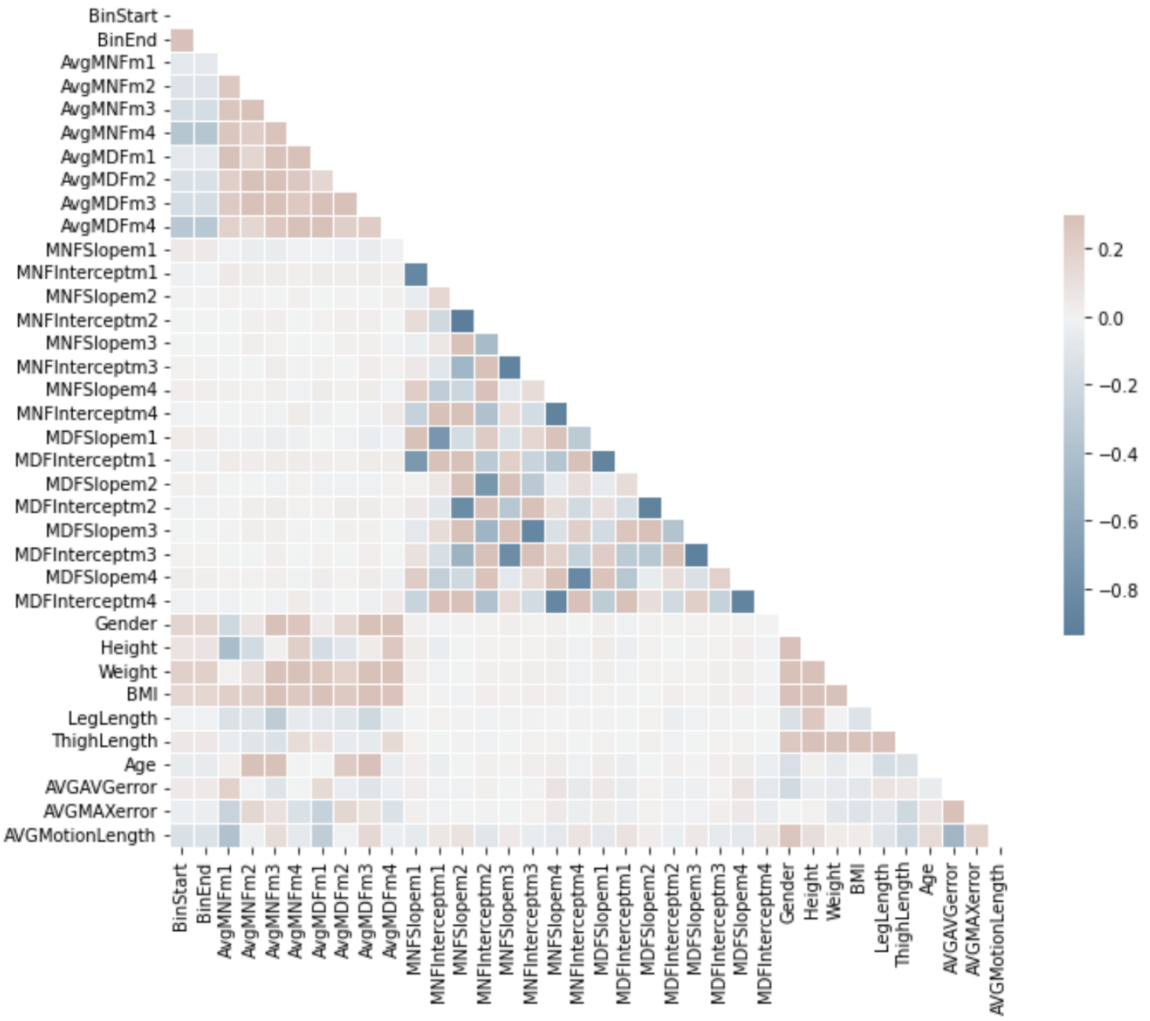


Figure 53: Correlation matrix for all the input features

4.2.3 Feature Scaling and Encoding

There is a need to ensure all the features are numerical. For the analysis in this work, the gender of participants is required to be converted to a numerical value from the original M or F (male or female). The problem is the numerical significance of 0 or 1 which affects the training process as they have a specific value meaning for the machine learning algorithms. The solution is to use a binary representation of the numbers with same values, also known as encoding, 01 for Female

category, and 10 for Male category. Furthermore, encoding must be applied to the "phase" feature as the significance of the number is not important, rather, the feature represents a category.

As explained in Section 2.5.4, scaling the input features before the training is essential as it helps to reduce the errors. Therefore, Min-Max Scaling was applied to all the input features except phase and Gender which fall into categorical data. Please note that the variables that will be fed to the models should be scaled based on Table 3:

Table 3: Minimum and maximum values utilized for scaling of the features studied in this work; future datasets must be scaled based on this table.

Feature	BinStart	BinEnd	AvgMNFm1	AvgMNFm2	AvgMNFm3
MIN	5.02E+00	1.13E+01	4.68E+01	4.41E+01	3.57E+01
MAX	3.23E+02	3.31E+02	1.09E+02	1.18E+02	1.30E+02

Feature	AvgMNFm4	AvgMDFm1	AvgMDFm2	AvgMDFm3	AvgMDFm4
MIN	4.64E+01	3.35E+01	3.36E+01	3.00E+01	3.40E+01
MAX	1.80E+02	1.00E+02	1.11E+02	1.22E+02	1.78E+02

Feature	MNFSlopem1	MNFInterceptm1	MNFSlopem2	MNFInterceptm2	MNFSlopem3
MIN	-9.06E+00	-1.35E+03	-4.75E+01	-1.74E+03	-2.85E+01
MAX	7.07E+00	1.06E+03	8.03E+00	9.62E+03	1.18E+01

Feature	MNFInterceptm3	MNFSlopem4	MNFInterceptm4	MDFSlopem1	MDFInterceptm1
MIN	-2.28E+03	-1.78E+01	-8.31E+03	-9.86E+00	-3.43E+03
MAX	5.83E+03	4.18E+01	4.07E+03	1.74E+01	1.39E+03

Feature	MDFSlopem2	MDFInterceptm2	MDFSlopem3	MDFInterceptm3	MDFSlopem4
MIN	-2.07E+01	-1.99E+03	-3.64E+01	-2.76E+03	-2.08E+01
MAX	1.06E+01	4.21E+03	1.40E+01	7.39E+03	4.78E+01

Feature	MDFInterceptm4	Height	Weight	BMI	LegLength
MIN	-9.53E+03	1.60E+02	5.00E+01	1.95E+01	8.00E+01
MAX	4.39E+03	1.86E+02	9.00E+01	2.70E+01	9.50E+01

Feature	ThighLength	Age	AVGAVGerror	AVGMAXerror	AVGMotionLength
MIN	3.90E+01	2.50E+01	2.82E-02	1.39E-01	3.71E-01
MAX	4.80E+01	3.00E+01	3.79E-01	9.23E-01	1.00E+00

4.2.4 Grid Search

In this research, the grid search is chosen for model tuning as it is more effective for the purpose of this work. The grid search process was implemented during the training process to find the best hyperparameters. As explained in Section 2.5.5, grid search is essential for training the models effectively. Table 4 represents the variables that were chosen for grid search in both classification and regression using the python language. F1 in classification and mean-squared-error for regression were the metrics that the grid search algorithm used for the finding the best parameter set at each iteration of the cross validation. For instance, the choices yielded 360 iterations of SVM analysis for each fold of a 5-fold cross-validation (1800 in total), the selected model by the grid search provided the maximum F1 score among all the iterations.

Table 4: The models and parameters utilized in the grid search process

Classification [127]		
Model	***Hyperparameter***	***Choices***
Logistic Regression	Inverse of Regularization Strength (C)	1e-3, 1e-2, 1e-1, 1, 1e1, 1e2, 1e3, 1e5
	Model Penalty	l1, l2, elasticnet
	Class weight	None, Balanced
LDA	Tolerance	1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 1e-0
QDA	Tolerance	1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 1e-0
Gaussian NB	Variance smoothing	1e-8, 1e-9, 1e-10
SVM	Inverse of Regularization Strength (C)	1e3, 5e3, 1e4, 5e4, 1e5
	Gamma	0.0001, 0.0005, 0.001, 0.005, 0.01, 0.1
	Class weight	None, Balanced
	Tolerance	1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 1e-0
Random Forest	Bootstrap	TRUE, False
	Max depth	10, 20, 30, 50, 90, None
	Max features	auto, sqrt
	Min samples leaf	1, 2, 4, 8
	Min samples split	2, 5, 10
	N estimators	100, 200
Neural Network	Hidden layer sizes	min_layer=3, max_layer=8, min_node=5, max_node=20, node_interval=2
	Alpha	0.0001, 0.001, 0.01, 0.1
Regression [127]		
Neural Network	Hidden layer sizes	min_layer=3, max_layer=8, min_node=5, max_node=20, node_interval=3
	Alpha	0.0001, 0.001, 0.01, 0.1
Random Forest	Bootstrap	TRUE, False
	Max depth	10, 20, 30, 50, 90, None
	Max features	auto, sqrt
	Min samples leaf	1, 2, 4, 8
	Min samples split	2, 5, 10
	N estimators	100, 200
Decision Tree	Criterion	squared_error, absolute_error
	Splitter	best, random
K-Nearest-Neighbors	N neighbors	4, 5, 6
	Weight	uniform, distance

4.2.5 Test/Train set selection

There are many ways of selecting Test/Train sets for the cross validation as explained in 2.5.6. For this research, two different approaches for selecting the Test/Train sets were used for determining which machine learning algorithm suits the application that will be highlighted below.

4.2.5.1 5-Fold Cross Validation (5F-CV)

For the first Test/Train set selection, the entire data set will be randomly divided into 5 different groups and, the training was performed 5 times as one of the groups is selected for testing while the rest are used for training. This process is repeated four more times with a different group as a testing set. The validation parameters as explained in 2.5.7 will represent the result. The metrics for each of the folds along with appropriate visuals will be represented for comparison and further analysis. The 5F-CV is an efficient method when it comes to the training time; no matter the number of participants, the training process is only performed 5 times. However, the distribution of the test set must be considered in order for it to cover a representative data.

4.2.5.2 Leave-One-Participant-out Cross Validation (LOPO-CV)

This approach in cross validation utilized the entire data for all but one of the participants in the training process, and the model is tested against the participant data that was left out. The process of training on nine groups and then testing on the selected group as the test set will be repeated ten times and each time one of the groups will be the test set. A major advantage of the LOPO-CV is the special perspective on the effect of a participant's data on the training. Moreover, the time for

the training of this cross-validation approach increases as the number of participants rises. Similar to the 5F-CV approach, the results and appropriate visual will be provided.

4.3 Results and Discussion

A combination of EMG data, exergame performance metrics, and patient information were adopted into machine learning models to improve the quality of care and facilitate the rehabilitation process. Fatigue onset recognition and fatigue level prediction are the tasks that were undertaken.

4.3.1 Fatigue Onset Recognition with Machine Learning Classification

In this method, the goal is to find a machine learning algorithm that defines whether the participant is fatigued or not. As explained previously, the target was labeled based on the reported perceived fatigue of larger than six. The rest of the data were gathered from the preprocessing step. This work is supervised learning, and a classification algorithm should be used here. The classifiers utilized in this study are:

- LDA
- QDA
- Neural Network
- Random Forest
- SVM
- Gaussian NB
- A Baseline classifier (a coin toss model) was also used for comparing the other classifiers.

A 5-fold cross-validation and a leave-one-participant-out cross-validation have been utilized in this section to allow the best method to be chosen by the developers. The 5-fold cross-validation

is more efficient when more data is added to the dataset, whereas the leave-one-participant-out cross-validation is preferred with the current number of data points.

4.3.1.1 5-Fold Cross-Validation (5F-CV)

Table 5 provides the average and the standard deviation of the performance metrics for the best set of parameters found with the grid search algorithm for each fold. The analysis **accuracy** for all the models was in an acceptable range of 0.80±0.16 and 0.94±0.02. The RF classifier has performed better due to a higher value of accuracy (0.94) and a smaller deviation (±0.02) which means the true fatigued predictions and true not-fatigued predictions of the RF classifier with respect to the entire dataset is higher than the other classifiers. **Recall** metric represents the proportion of the incidents that the participant was actually fatigued and was correctly predicted by the model. The best recall score was for the RF model with the value of 0.93±0.06, while the least performing classifier was the Gaussian NB (0.80±0.10). **Specificity** is the third most important metric showing how well the model performs against not-fatigued predictions. The specificity for the RF classifier was 0.94±0.05, whereas the next value was almost tied between Gaussian NB and LDA (0.90±0.09 and 0.90±0.03, respectively). The proportion of correctly predicted fatigued with respect to all the positive predictions is described by the **precision** metric. The most precise classifier was the RF (0.94±0.05), and the least precision was achieved by the QDA (0.81±0.17). The RF classifier shows a larger standard deviation in the precision metric than most classifiers, but it performed better than the rest without any outlier. **F1 score** was implemented to consider both precision and recall as a single metric for the comparison of different classifiers regarding the true positive predictions (the fatigued label was correctly predicted). The

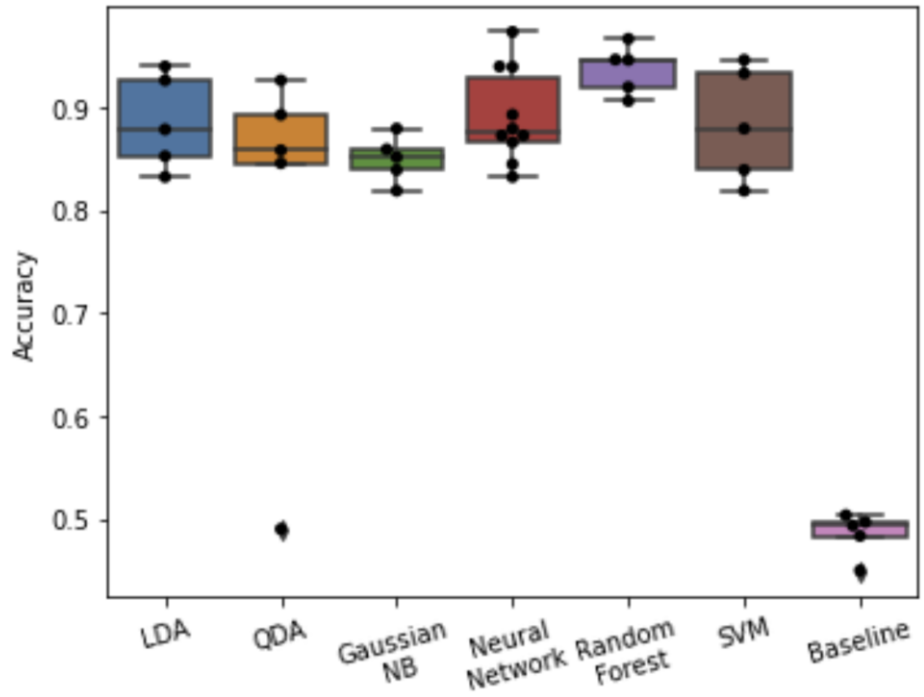
range of the F1 score is between 0.94±0.02 and 0.83±0.09, where the RF classifier has performed best.

Note that the neural network analysis was performed twice to ensure the effects of repeatability on the data. It indicated that repeating the analysis does not change the range of the outcomes. The neural network was chosen since it has the least number of randomly selected parameters that affect the results. Furthermore, the QDA analysis has only been reported to show its limited nature; the data set was not optimized for this classifier, and some expected collinearity warnings were observed.

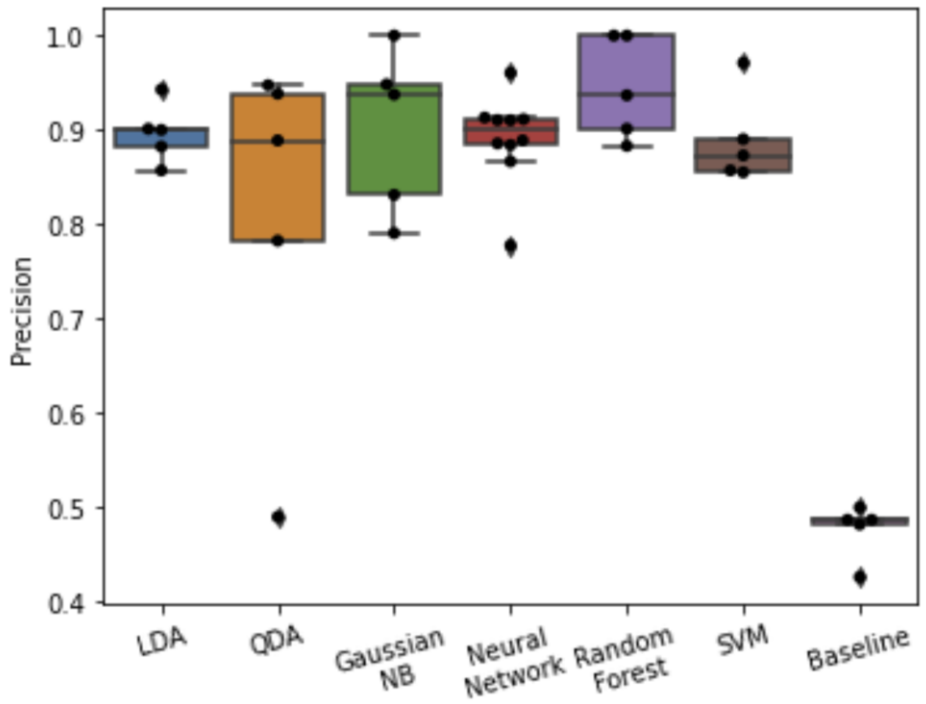
Table 5: Summary of the metrics for the 5-fold cross validation in the fatigue onset recognition.

Classifier	Accuracy	Recall	Specificity	ROC-AUC	F1 score	PR-AUC	Precision
Gaussian NB	0.85 ± 0.02	0.80 ± 0.10	0.90 ± 0.09	0.85 ± 0.02	0.84 ± 0.03	0.81 ± 0.03	0.90 ± 0.08
LDA	0.89 ± 0.04	0.87 ± 0.09	0.90 ± 0.03	0.89 ± 0.04	0.88 ± 0.05	0.84 ± 0.04	0.90 ± 0.03
QDA	0.80 ± 0.16	0.90 ± 0.11	0.70 ± 0.36	0.80 ± 0.15	0.83 ± 0.09	0.77 ± 0.14	0.81 ± 0.17
Neural Network	0.87 ± 0.04	0.88 ± 0.08	0.86 ± 0.07	0.87 ± 0.04	0.87 ± 0.04	0.82 ± 0.04	0.87 ± 0.05
Random Forest	0.94 ± 0.02	0.93 ± 0.06	0.94 ± 0.05	0.94 ± 0.02	0.94 ± 0.02	0.91 ± 0.03	0.94 ± 0.05
SVM	0.88 ± 0.05	0.87 ± 0.09	0.89 ± 0.04	0.88 ± 0.05	0.88 ± 0.06	0.84 ± 0.06	0.89 ± 0.04
Baseline	0.49 ± 0.02	0.50 ± 0.07	0.47 ± 0.04	0.49 ± 0.02	0.48 ± 0.05	0.49 ± 0.01	0.48 ± 0.03

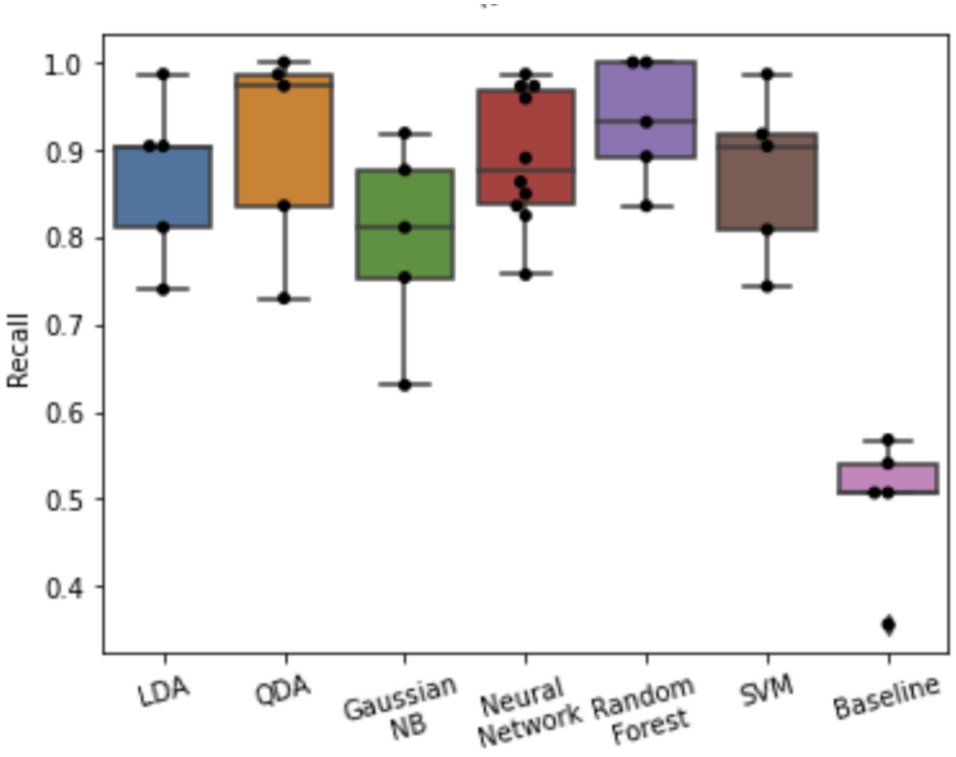
Additionally, Figure 54 illustrates the metrics as a box plot for better comparison. The results of each cross-validation fold for the respective classifier are shown with a black dot on the diagrams. The box indicates the quartile range (first to the third quartile), where the median is shown as a gray line. The whiskers illustrate the minimum and maximum values which have filtered out the outliers (single dots away from the rest of the data).


Accuracy
0.9
0.8
0.7
0.6
0.5
LDA
QDA
Gaussian NB
Neural Network
Random Forest
SVM
Baseline

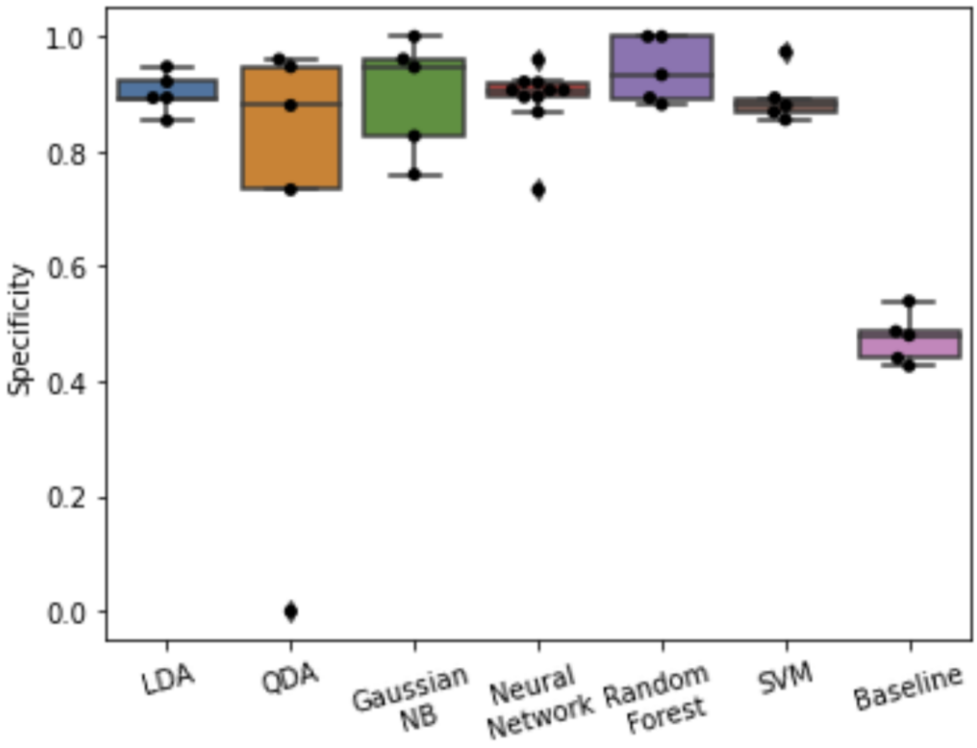
(a)


Precision
1.0
0.9
0.8
0.7
0.6
0.5
0.4
LDA
QDA
Gaussian NB
Neural Network
Random Forest
SVM
Baseline

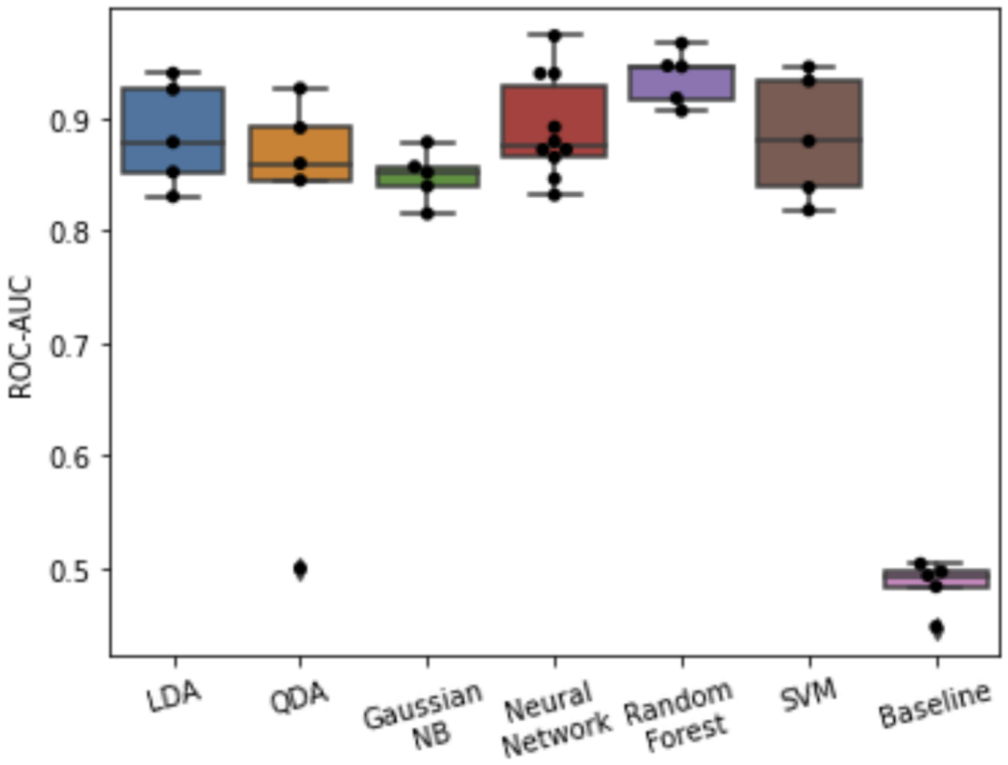
(b)


Recall
1.0
0.9
0.8
0.7
0.6
0.5
0.4
LDA
QDA
Gaussian NB
Neural Network
Random Forest
SVM
Baseline

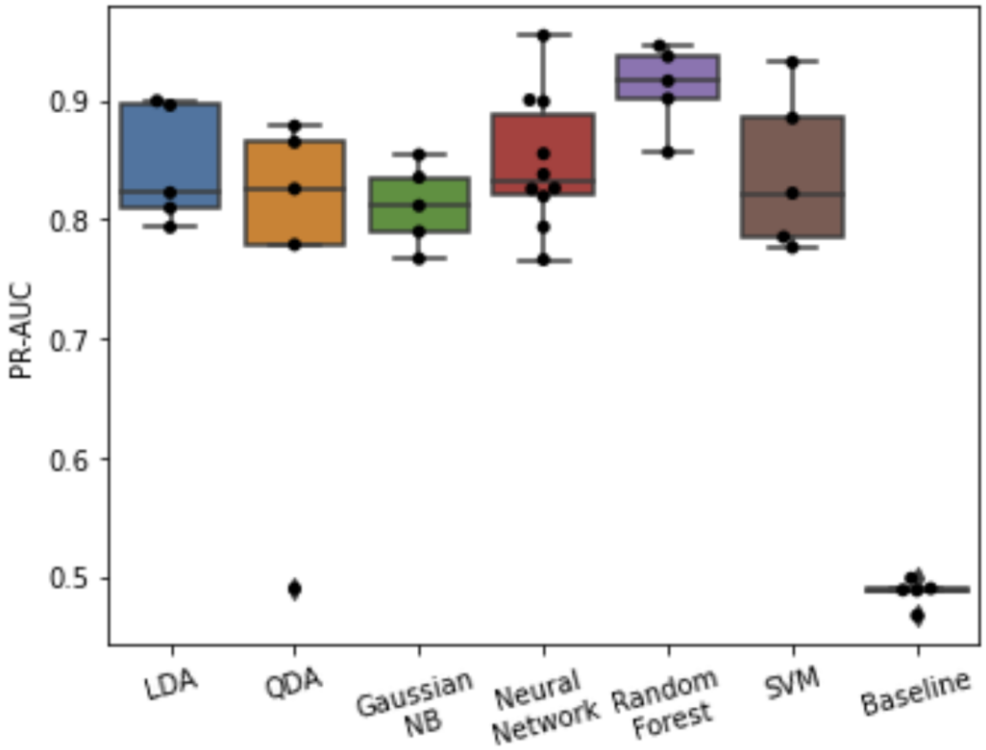
(c)


Specificity
1.0
0.8
0.6
0.4
0.2
0.0
LDA
QDA
Gaussian NB
Neural Network
Random Forest
SVM
Baseline

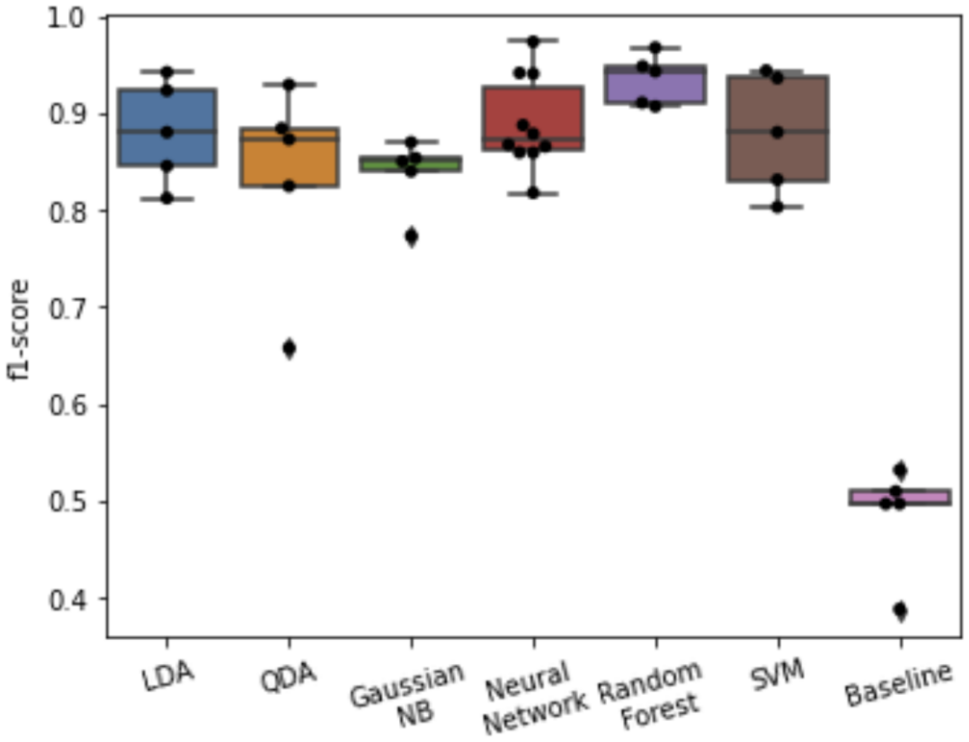
(d)


ROC-AUC
0.9
0.8
0.7
0.6
0.5
LDA
QDA
Gaussian NB
Neural Network
Random Forest
SVM
Baseline

(e)


PR-AUC
0.9
0.8
0.7
0.6
0.5
LDA
QDA
Gaussian NB
Neural Network
Random Forest
SVM
Baseline

(f)



(g)

Figure 54: 5-fold cross-validation metrics for all the classifiers in the fatigue onset recognition a) Accuracy, b) Precision, c) Recall, d) Specificity, e) ROC-AUC, f) PR-AUC, and g) F1 Score

Figure 55 represents the ROC and PR curves which are helpful in selecting the best classifier. The lines for each fold of the classifier have been drawn on the curve with light hues; however, the average has been plotted with a darker tone while the area where the classifier may fall into has been shaded accordingly. The RF classifier had the best performance based on the curves as the ROC-AUC is 0.94±0.02 (closets value to 1), while its PR-AUC is 0.91±0.03, which had the largest area under the curve among the other classifiers. The next best performance is for the LDA classifier with the ROC-AUC and PR-AUC of 0.89±0.04 and 0.84±0.04, respectively.

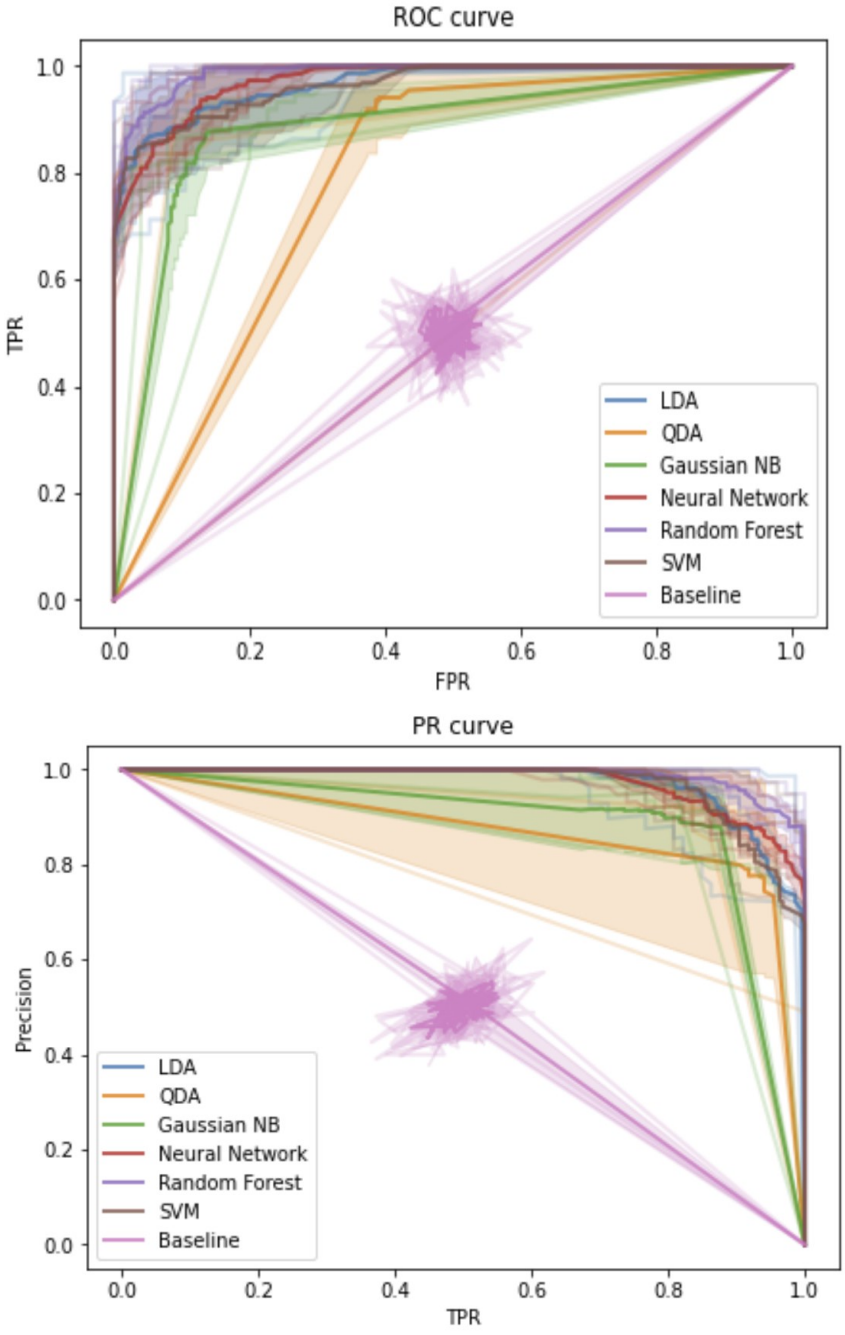


Figure 55: Appropriate curves for comparing the classifiers, (a) the ROC curve, (b) the PR curve.

Figure 56 represents the confusion matrices for all the classifiers; it includes the TP, TN, FP, and FN values discussed in Section 2.5.7.1. The RF classifier includes the greatest number of

correctly predicted instances compared to the other classifiers. LDA and Gaussian NB classifiers also provided an acceptable FN; however, the FN values are high resulting in diminished rehabilitation. Note that it is preferable to have a FP (premature fatigue prediction) rather than a FN (patient is fatigued, but the model indicates not fatigued) as there will be no serious harm in case of ending the rehabilitation process early; however, too many FP detections become annoying and interruptive in the process. Therefore, a reasonable balance is required between these two parameters.

In summary, the RF classifier consistently performed well among the validation metrics as shown in Table 5. It has the best accuracy and F1 score, indicating how well this classifier performs on our dataset. Figure 55 is another indication that the RF classifier is a suitable choice for the purpose of the research given the larger area under the curves and proximity to the optimal point as per Sections 2.5.9 and 2.5.10. Furthermore, the confusion matrices in Figure 56 indicated that the RF classifier had performed the best in the most critical scenario when the detection of fatigue is failed, and the exercise was elongated. The parameters associated with the best fold of the RF classification analysis using Python are as follows:

- 'bootstrap': False
- 'max_depth': None
- 'max_features': 'sqrt'
- 'min_samples_leaf': 2
- 'min_samples_split': 5
- 'n_estimators': 100

These parameters can be utilized for replicating the model for adaptation with rehabilitation robots.

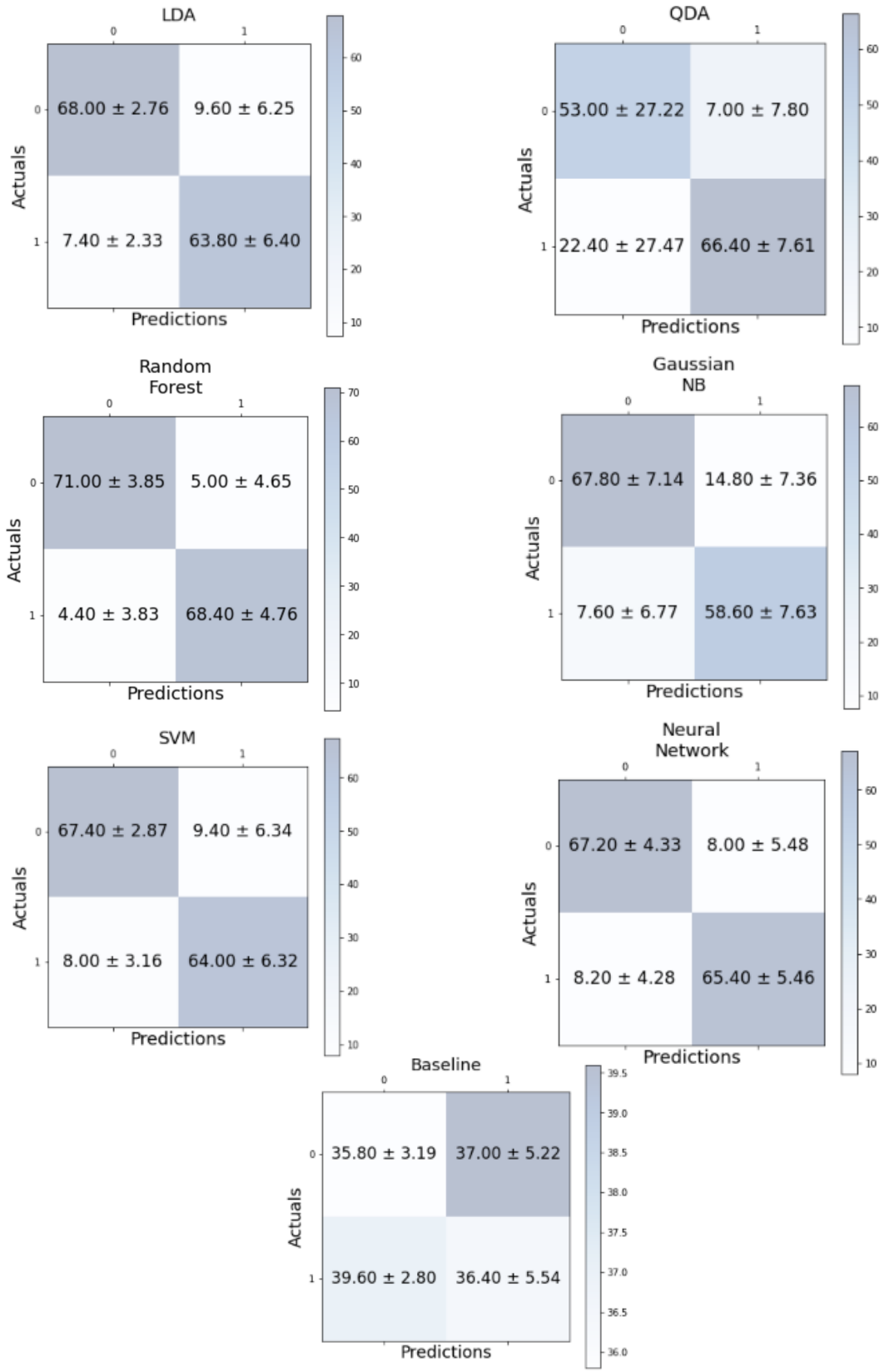


Figure 56: Confusion matrices for the 5-fold cross-validation analysis using all the classifiers

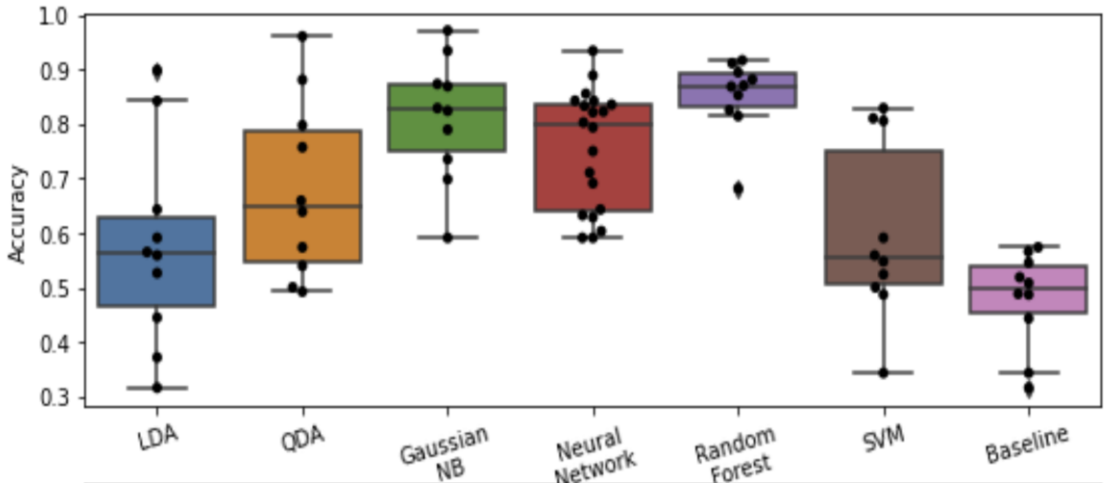
4.3.1.2 Leave-One-Participant-Out Cross-Validation (LOPO-CV)

Table 6 and Figure 57 are the numerical and illustrative representations of the analysis results, respectively.

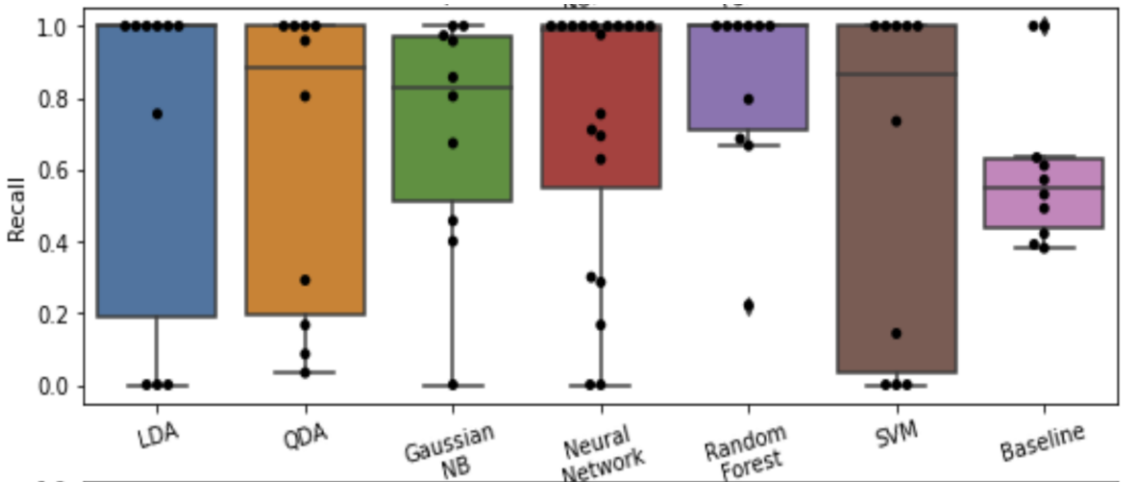
Starting with **accuracy**, the RF classifier provided the highest performance (0.85±0.06) while the LDA provided the least performance (0.58±0.18), which is only 0.10 higher than the baseline. The **recall** range is from 0.59±0.46 to 0.84±24, which are performed by SMV and RF, respectively. The large standard deviation in the recall is perhaps due to the low testing set population. Additionally, the deviation of the recall from the baseline with respect to the QDA, LDA, Neural Network, and SVM is lower than the RF and Gaussian NB classifiers. Gaussian NB classifier had the best performance of **specificity** with the value of 0.91±0.13 closely followed by RF and Neural Network classifier (0.87±0.12 and 0.85±0.16, respectively). LDA, QDA, and SVM classifiers did not perform well while having a large standard deviation. Moreover, the precision metric for the RF classifier was 0.87±0.12, which was the best performance, while LDA and SVM had lower performances than baseline (the baseline precision value was 0.50±0.11). Finally, using the F1 score for comparing the classifiers, the RF classifier had the best score (0.81±0.16), and the SVM and LDA had the worst F1 scores with large standard deviations (0.46±0.35 and 0.49±0.35).

Table 6: Summary of the metrics for the leave-one-participant-out cross validation in the fatigue onset recognition.

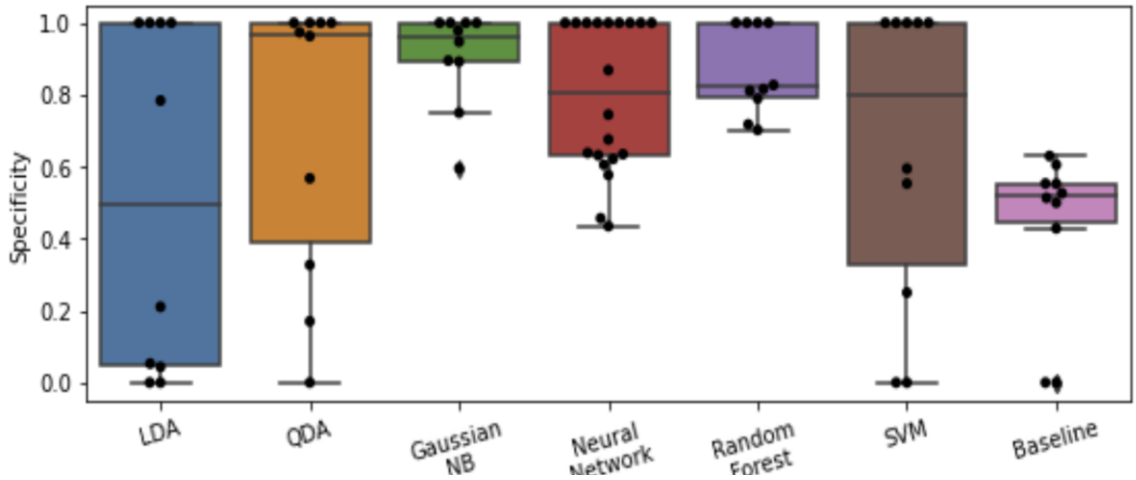
Classifier	Accuracy	Recall	Specificity	ROC-AUC	F1 score	PR-AUC	Precision
Gaussian NB	0.81 ± 0.11	0.71 ± 0.32	0.91 ± 0.13	0.81 ± 0.13	0.73 ± 0.27	0.77 ± 0.15	0.81 ± 0.29
LDA	0.58 ± 0.18	0.68 ± 0.45	0.51 ± 0.46	0.59 ± 0.15	0.49 ± 0.35	0.55 ± 0.19	0.42 ± 0.33
QDA	0.68 ± 0.15	0.63 ± 0.41	0.70 ± 0.38	0.67 ± 0.15	0.57 ± 0.30	0.62 ± 0.18	0.82 ± 0.22
Neural Network	0.78 ± 0.12	0.66 ± 0.40	0.85 ± 0.16	0.76 ± 0.15	0.65 ± 0.35	0.72 ± 0.17	0.70 ± 0.36
Random Forest	0.85 ± 0.06	0.84 ± 0.24	0.87 ± 0.12	0.85 ± 0.08	0.81 ± 0.16	0.79 ± 0.11	0.87 ± 0.12
SVM	0.60 ± 0.15	0.59 ± 0.46	0.64 ± 0.40	0.61 ± 0.14	0.46 ± 0.35	0.56 ± 0.18	0.48 ± 0.37
Baseline	0.48 ± 0.08	0.60 ± 0.21	0.43 ± 0.22	0.52 ± 0.04	0.51 ± 0.06	0.50 ± 0.10	0.50 ± 0.11



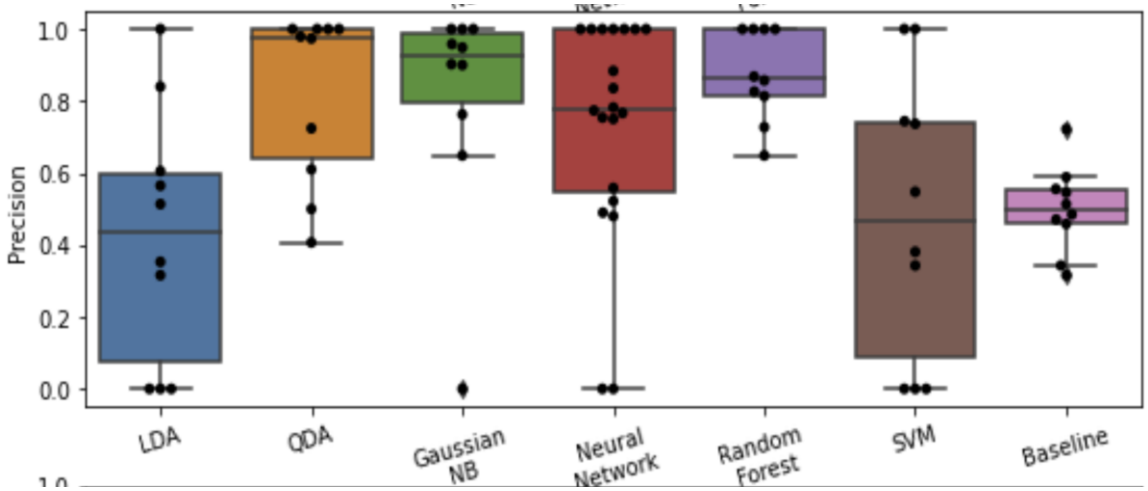
(a)


Recall
1.0
0.8
0.6
0.4
0.2
0.0
LDA
QDA
Gaussian NB
Neural Network
Random Forest
SVM
Baseline

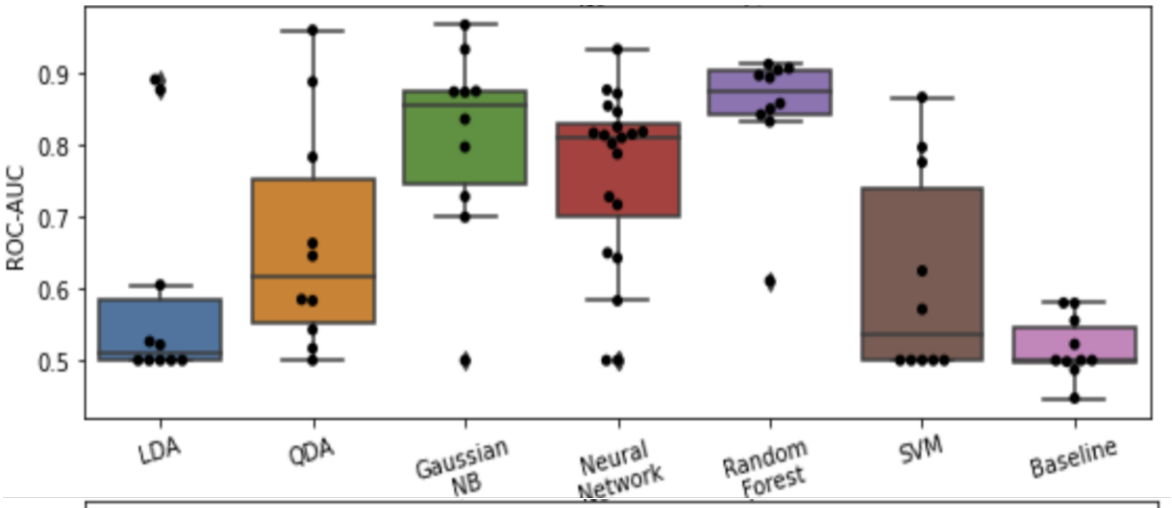
(b)


Specificity
1.0
0.8
0.6
0.4
0.2
0.0
LDA
QDA
Gaussian NB
Neural Network
Random Forest
SVM
Baseline

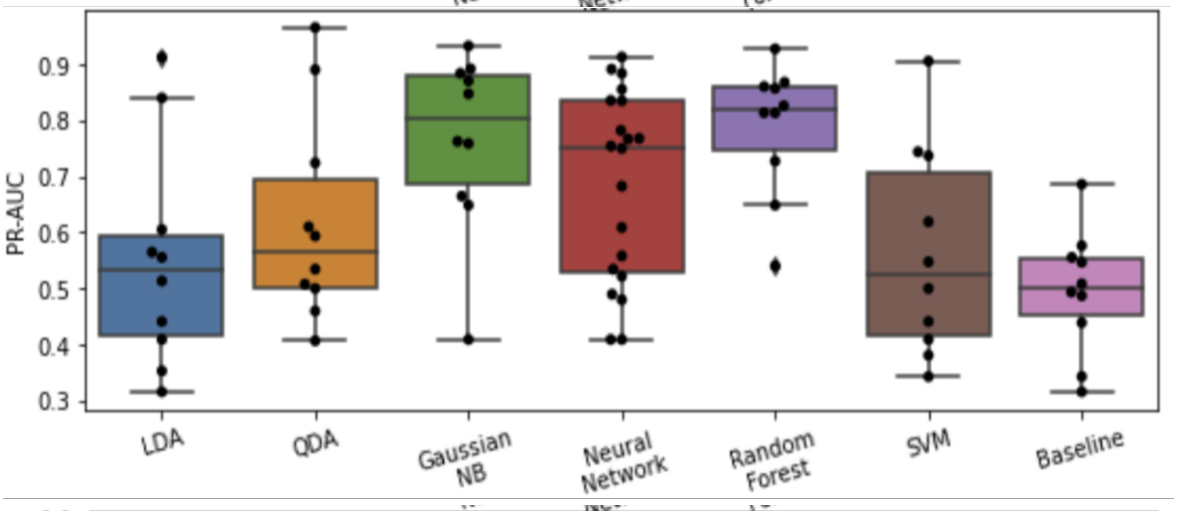
(c)


Precision
1.0
0.8
0.6
0.4
0.2
0.0
LDA
QDA
Gaussian NB
Neural Network
Random Forest
SVM
Baseline

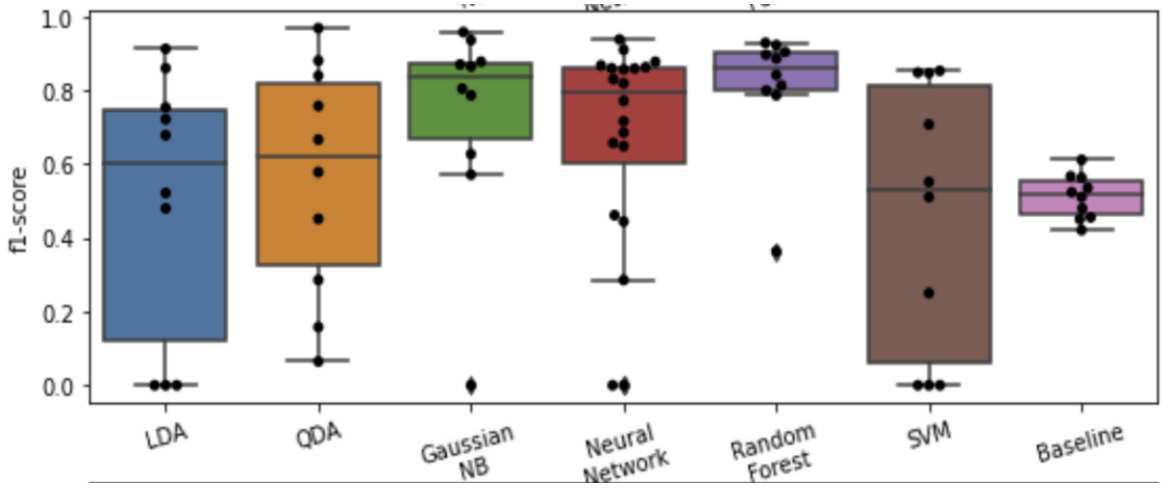
(d)



(e)



(f)



(g)

Figure 57: The associated box plots representing the metrics for the leave-one-participant-out cross validation in the fatigue onset recognition analysis. a) Accuracy, b) Recall, c) Specificity, d) Precision, e) ROC-AUC, f) PR-AUC, and g) F1 Score

The superiority of the RF classifier is evident in Figure 58, where the solid purple line in both curves is closer to the optimal points. The ROC-AUC for RF classifier was 0.85±0.08, and the PR-AUC of this classifier was 0.79±0.11. Even though the Gaussian NB line on the ROC curve is closer to the optimal point in an interval between FPR of 0.150 and 0.175, the RF has performed better for the rest of the range. One factor that may have caused this issue is the outlier in the folds (which is participant data). Further investigation indicated that the outlier corresponds to participant 7.

Additionally, the RF classifier analysis with leave-one-participant-out cross-validation without participant 7 data was investigated. Table 7 represents both sets of data which expectedly show improvements; however, these metrics have not reached the 5-fold cross-validation level. This confirms that the population of the dataset influences the results. Switching the participant 7 dataset with a new participant would have potentially improved the metrics.

Table 7: Metrics for random forest leave-one-participant-out cross-validation analysis when considering participant 7 data as an outlier.

	Accuracy	Recall	Specificity	ROC-AUC	F1-score	PR-AUC	Precision
With Participant 7	0.85 ± 0.06	0.84 ± 0.24	0.87 ± 0.12	0.85 ± 0.08	0.81 ± 0.16	0.79 ± 0.11	0.87 ± 0.12
Without Participant	0.88 ± 0.05	0.90 ± 0.13	0.88 ± 0.14	0.89 ± 0.06	0.88 ± 0.06	0.84 ± 0.09	0.88 ± 0.12

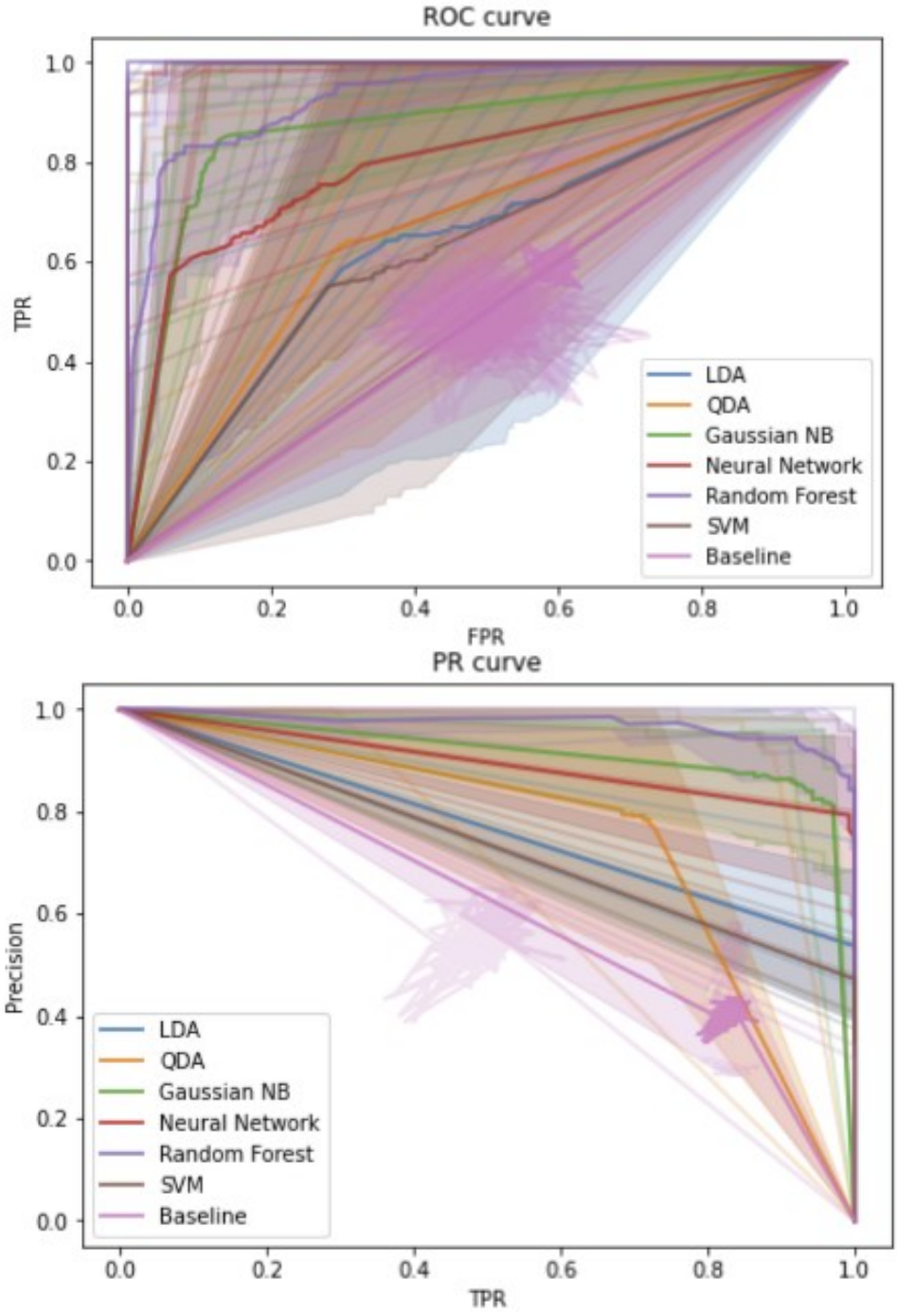


Figure 58: Associated ROC and PR curves for the leave-one-participant-out cross validation involving all the classifiers.

The confusion matrices for all the classifiers are provided in Figure 59. The RF classifier has performed better than the rest of the classifiers. However, it indicates that almost 18% of the time, the participant has been fatigued, but the prediction was not correct. Further investigation into participant 10, as an example, showed that the model first predicted fatigue while the participant was not fatigued. This is an indication that the RF is a dependable classifier that can operate well with some failsafe provisions to notify the therapist of a potential fatigue detection. Moreover, the goal of this analysis is to detect the onset of fatigue which is the beginning of a period when the muscle activation is reduced, and therefore, there is a leeway to change the exercise accordingly.

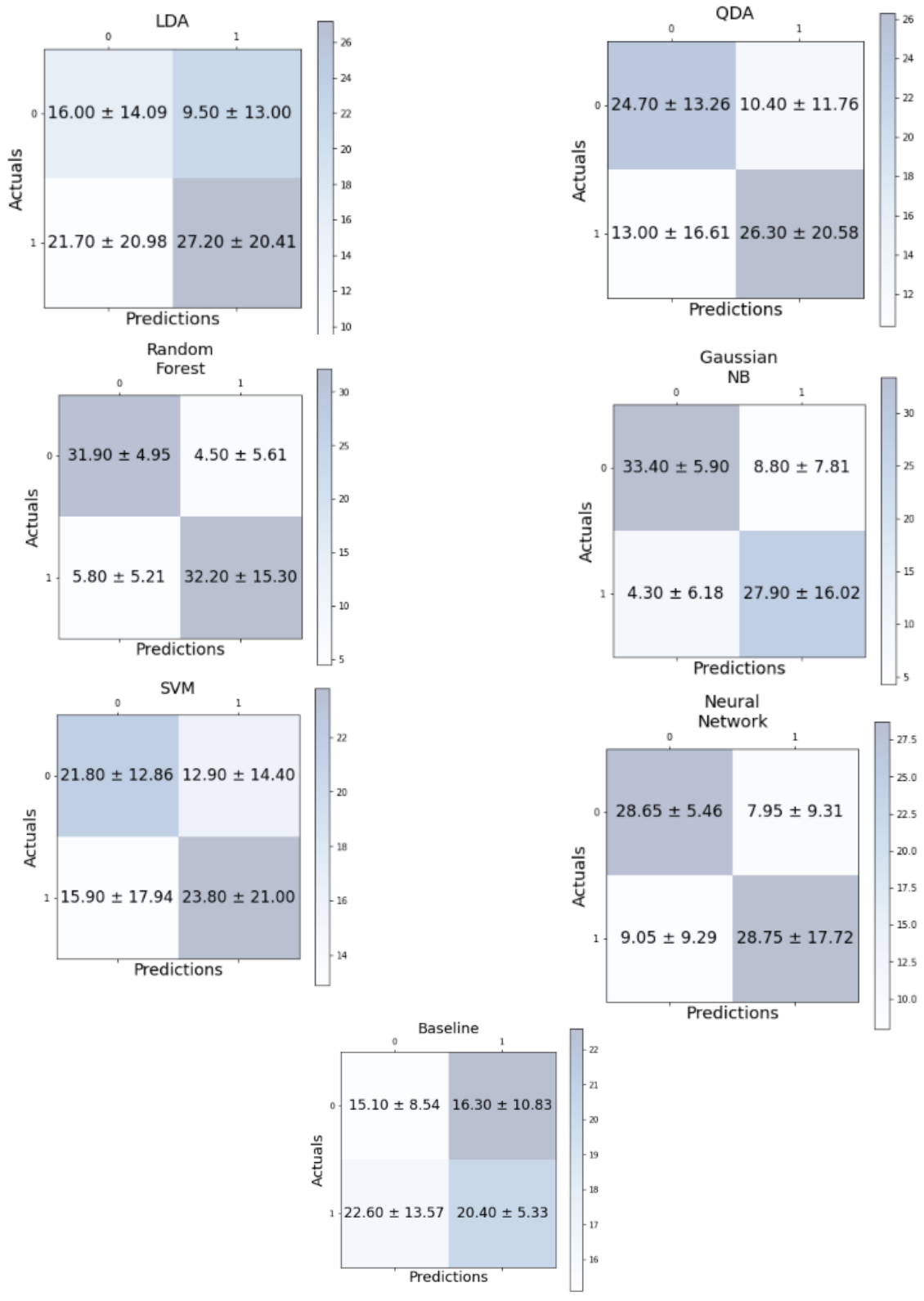


Figure 59: Confusion matrices for the leave-one-participant-out cross-validation analysis using all the classifiers

The RF classifier has once again performed better than the other classifiers when considering the distribution of the folds in all the metrics. Except for an outlier, the rest of the folds are densely packed within the quartile range (colored boxes) compared to the other classifiers, as shown in Figure 57. The most appropriate set of parameters for replicating the RF analysis are:

- 'bootstrap': False
- 'max_depth': 50
- 'max_features': 'sqrt'
- 'min_samples_leaf': 1
- 'min_samples_split': 2
- 'n_estimators': 200

4.3.1.3 **Fatigue Onset Detection Methods Comparison**

- The RF classifier has performed better and more consistently compared to the rest of the classifiers in both cross-validation methods.
- The sensitivity of the classification is increased with LOPO-CV analysis compared to the 5F-CV based on the ratio of the standard deviation and the mean (from the confusion matrices) since the LOPO-CV was a more conservative approach in preparing the model. The testing data set used for calculating the metrics in LOPO-CV is fewer than the 5F-CV, which affects the sensitivity of the values.
- Unlike the 5F-CV, the metrics of LOPO-CV are closer to the baseline classifier and indicate that the models do not perform as well. The main reason for the discrepancy is the lack of randomized selection of the training set in the 5-fold cross-validation. As mentioned earlier, the effect of random test/train selection in the 5F-CV is reduced when more data is available. Consequently, 1/5 of the data contain more variation of the features resulting in metrics that better reflect the performance of the model.

4.3.2 Fatigue Level Recognition with Regression

Rehabilitation robots such as VIGGR light are capable of adjusting their operating parameters through a controller with the goal of elongating the rehabilitation to maximize the effects. Furthermore, rehabilitation robots can be paired with exergames to prevent patient boredom and improve psychological impediments [128]. However, the patient gets tired as the therapy progresses if the robot's speed and impedance are not changed. The requirement for the robot controllers to effectively change the parameters is the availability of fatigue levels during the rehabilitation process. The controllers would then adjust the parameters according to the fatigue level and the predetermined therapy plan. The following regressors have been selected to perform the analysis based on the preprocessed data:

- Linear regression
- Neural network
- Random Forest
- Decision Tree
- K-Nearest Neighbors

The labels for the bins were added based on the fatigue level reported by the participants. Based on this application, the population of the dataset is low; however, the analysis was developed to better implement the capabilities of regression. The same cross-validation approaches were used for the regression analysis.

4.3.2.1 5-Fold Cross-Validation (5F-CV)

The RF and DT regressors have shown the best performance in recognizing the fatigue level, as seen in Table 8 and Figure 60. The comparison between MAE and MSE in RF and DT regressors

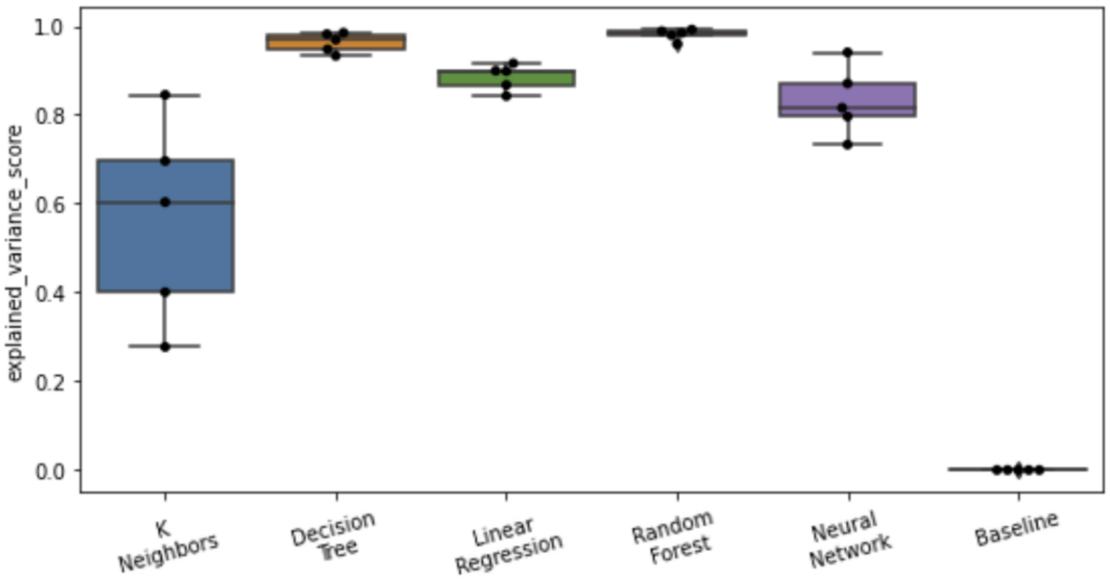
indicates that the RF yields fewer outliers compared to the DT. The MAE of both regressors is in the same range (0.26±0.08 and 0.21±0.11 for the RF and DT, respectively); however, the MSE (which is more susceptible to outliers) of the DT regressor (0.39±0.22) is almost twice the MSE of the RF (0.23±0.15). The KNN regressor had the worst performance score of MAE and MSE with large standard deviations among all the regressors with the value of 1.68±0.45 and 5.48±3.30, respectively. The presence of outliers is also evident from the maximum error metric. In the worst-case scenario, the DT regressor recognized the fatigue level incorrectly by almost 1 point compared to the RF regressor (2.97±0.97 and 1.93±0.77, respectively). The KNN regressor has the highest Max Error value of 6.12±1.29, which is higher than the baseline regressor. The coefficient determination, r^2 score, for the RF regressor has shown an exceptional performance followed by the DT regressor. The r^2 score for RF and DT regressor is 0.02 and 0.04 away from the perfect score of 1.

The following represents the best set of parameters that were used for the RF regressor as the best model for this application:

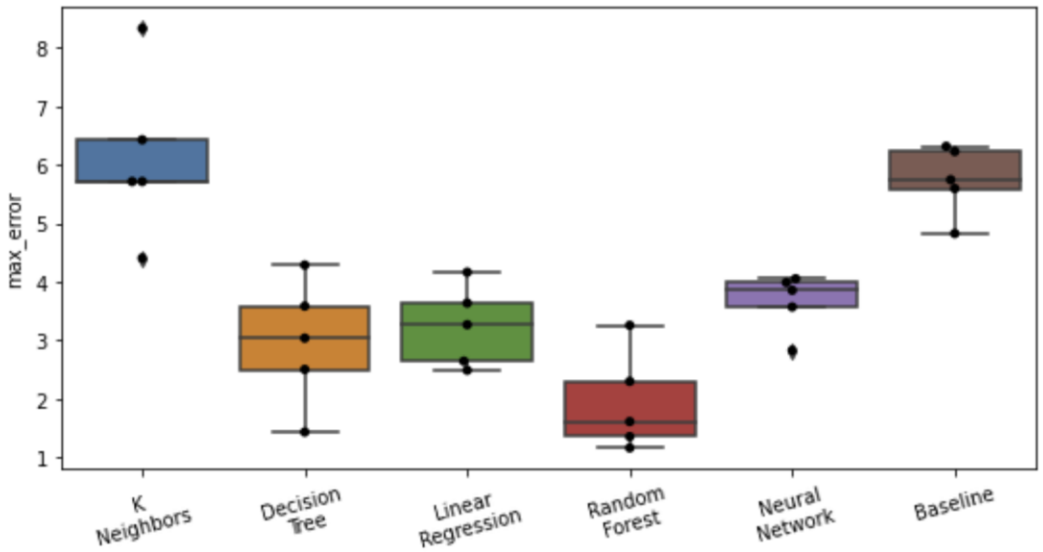
- 'bootstrap': True
- 'max_depth': None
- 'min_samples_leaf': 1
- 'min_samples_split': 2
- 'n_estimators': 200

Table 8: The metrics for the regression analysis for 5-fold cross validation.

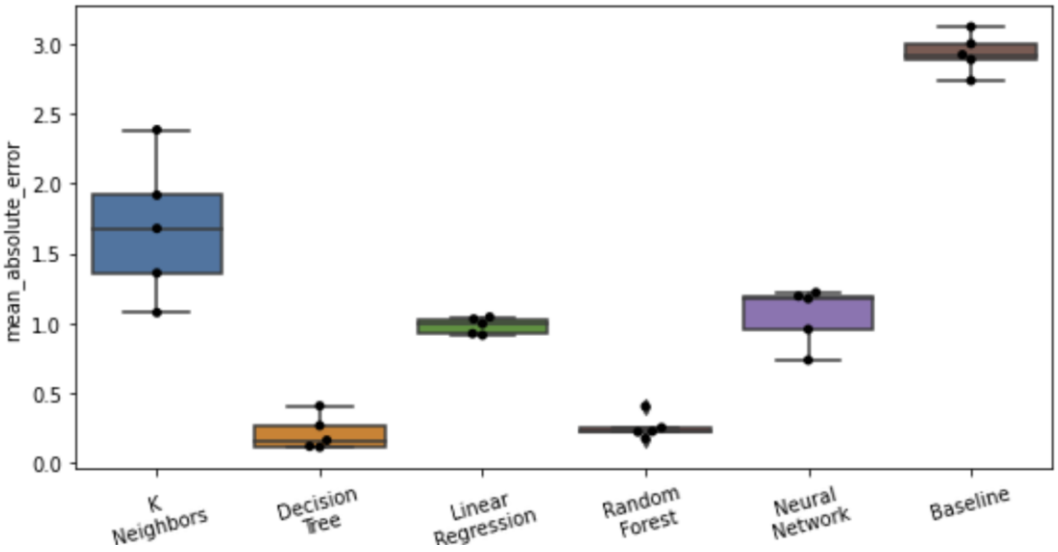
Regressor	Explained Variance Score (EVS)	Max Error	Mean Absolute Error (MAE)	Mean Squared Error (MSE)	r^2 score
Random Forest	0.98 ± 0.01	1.93 ± 0.77	0.26 ± 0.08	0.23 ± 0.15	0.98 ± 0.01
Decision Tree	0.96 ± 0.02	2.97 ± 0.97	0.21 ± 0.11	0.39 ± 0.22	0.96 ± 0.02
Linear Regressor	0.88 ± 0.03	3.24 ± 0.62	0.98 ± 0.05	1.48 ± 0.19	0.86 ± 0.02
Neural Network	0.83 ± 0.07	3.66 ± 0.45	1.06 ± 0.19	1.79 ± 0.52	0.82 ± 0.06
K_NearestNeighbors	0.56 ± 0.20	6.12 ± 1.29	1.68 ± 0.45	5.48 ± 3.30	0.47 ± 0.31
Baseline	0.00 ± 0.00	5.74 ± 0.53	2.93 ± 0.13	10.64 ± 0.76	-0.03 ± 0.03



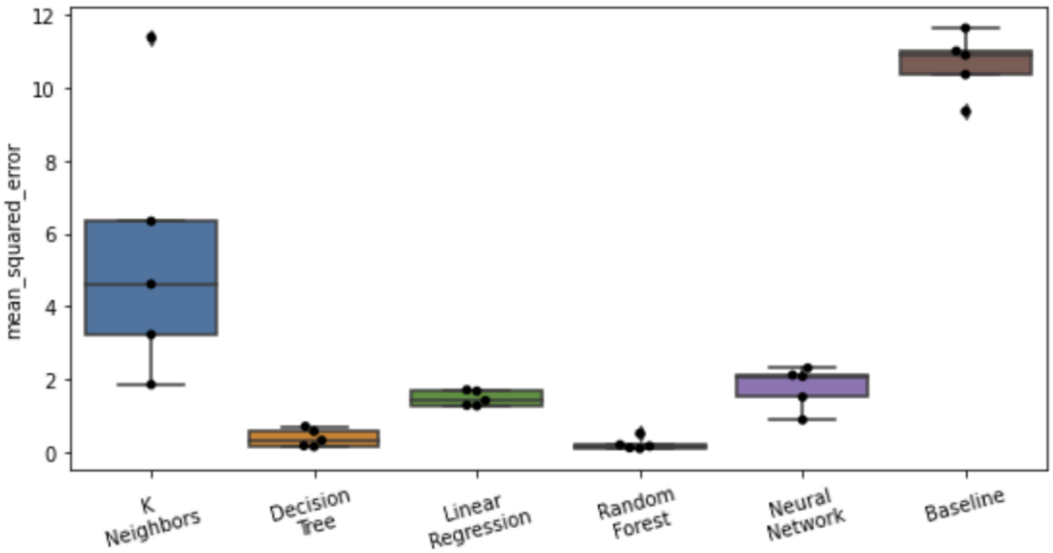
(a)


max_error
8
7
6
5
4
3
2
1
K Neighbors
Decision Tree
Linear Regression
Random Forest
Neural Network
Baseline

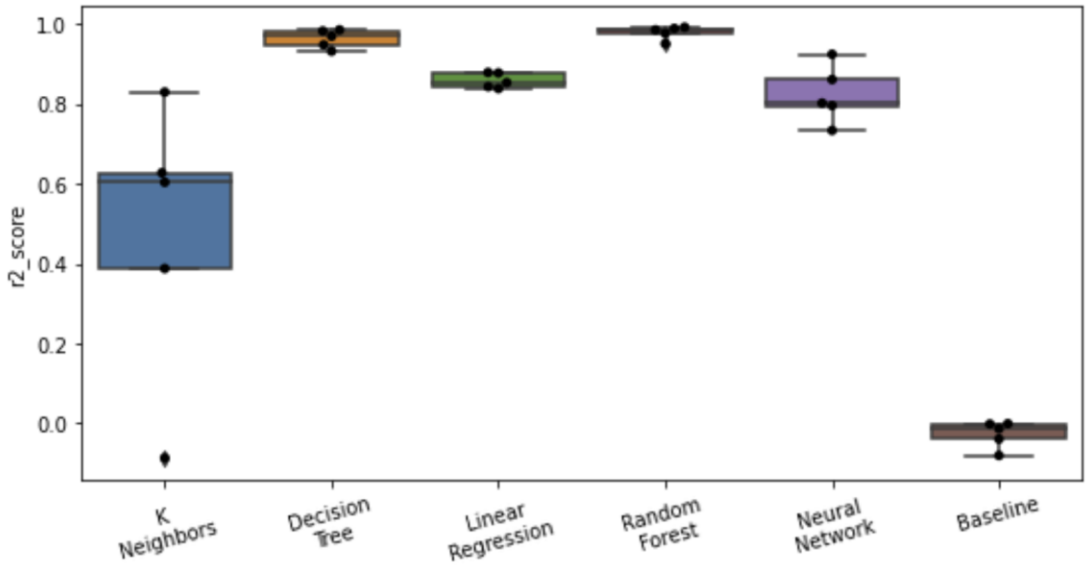
(b)


mean_absolute_error
3.0
2.5
2.0
1.5
1.0
0.5
0.0
K Neighbors
Decision Tree
Linear Regression
Random Forest
Neural Network
Baseline

(c)


mean_squared_error
12
10
8
6
4
2
0
K Neighbors
Decision Tree
Linear Regression
Random Forest
Neural Network
Baseline

(d)



(e)

Figure 60: Associated box plots for the 5-fold cross-validation of all the regressors used for predicting the fatigue level. a) EVS, b) Max error, c) Mean absolute error, d) Mean squared error, e) r^2 score.

4.3.2.2 Leave-One-Participant-Out Cross-Validation (LOPO-CV)

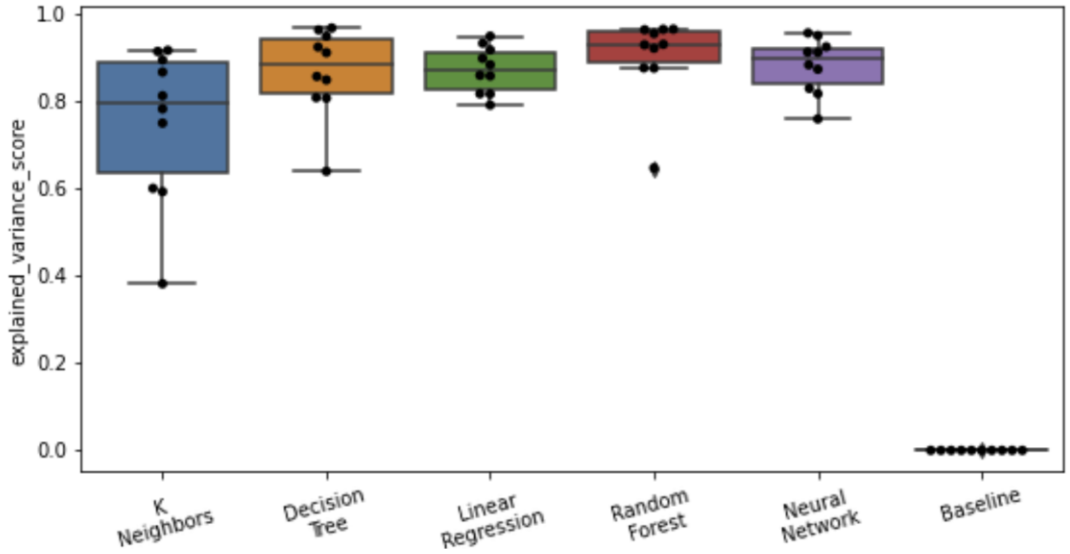
As shown in Figure 61 and Table 9, MAE and MSE of the RF regressor were closer to zero (0.85±0.39 and 1.52±1.47, respectively) followed by DT (1.85±1.46 and 0.93±0.35, respectively) compared to the rest of the regressors, whereas the EVS and the r^2 score of RF were closer to one (0.90±0.09 and 0.81±0.22, respectively). The DT regressor's EVS and r^2 score displayed the second-best performance after RF, with the values of 0.87±0.10 and 0.78±0.21, respectively. The LR regressor earned the worst results of MAE and MSE with a large standard deviation (4.49±2.46 and 27.30±25.00, respectively) which were higher than the baseline regressor performance. The RF and DT regressors had the lowest maximum error of 2.38±0.76 and 2.73±0.90, respectively, and the LR regressor had the highest maximum error of 6.38±2.93. The best performance is once again for the RF regressor. Replicating the RF regression requires the following set of parameters:

- 'bootstrap': True

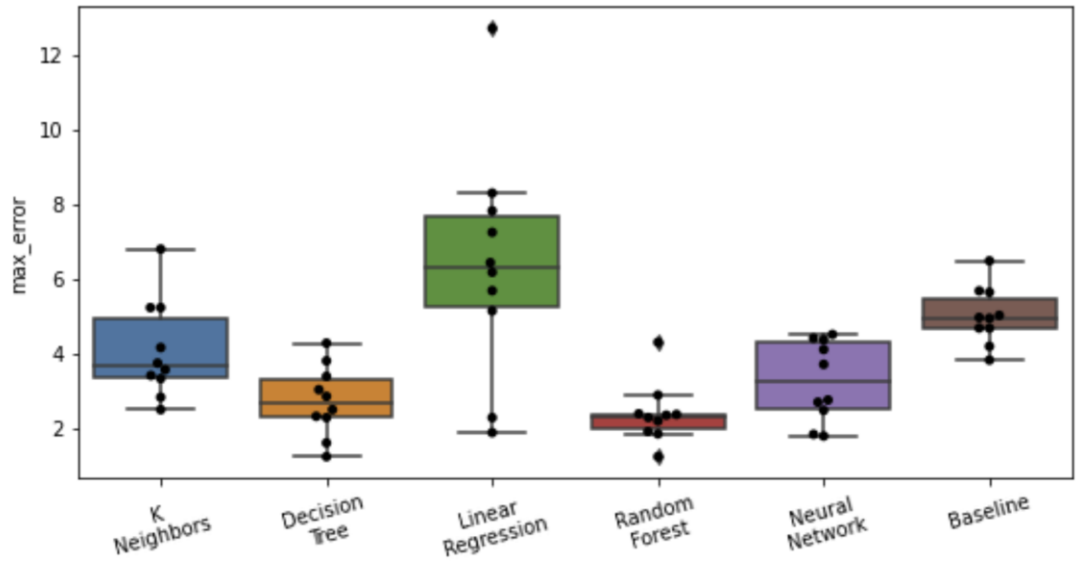
- 'max_depth': 30
- 'min_samples_leaf': 1
- 'min_samples_split': 5
- 'n_estimators': 100

Table 9: The metrics for the regression analysis for leave-one-participant-out cross validation.

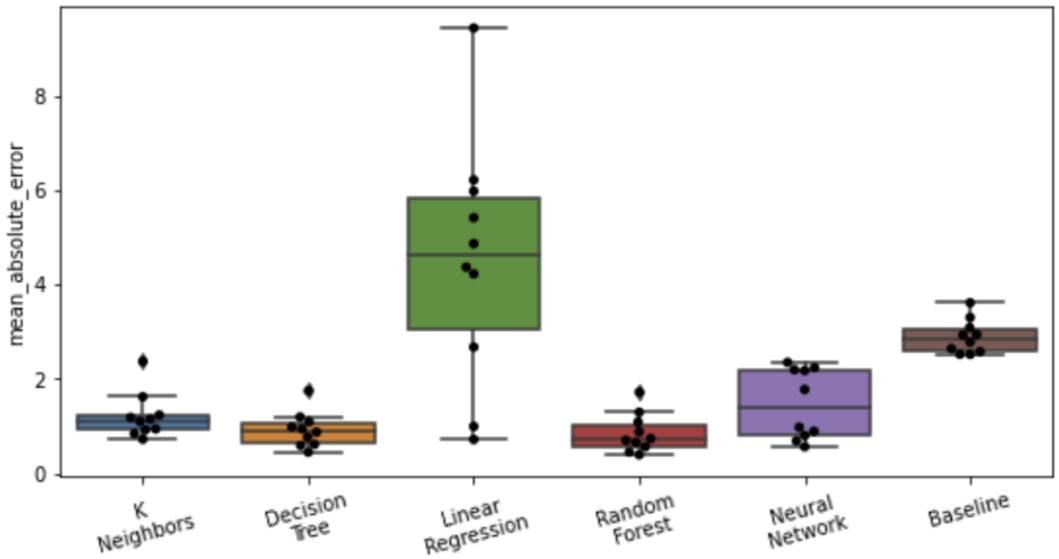
Regressor	Explained Variance Score (EVS)	Max Error	Mean Absolute Error (MAE)	Mean Squared Error (MSE)	r2 score
Random Forest	0.90 ± 0.09	2.38 ± 0.76	0.85 ± 0.39	1.52 ± 1.47	0.81 ± 0.22
Decision Tree	0.87 ± 0.10	2.73 ± 0.90	0.93 ± 0.35	1.85 ± 1.46	0.78 ± 0.21
Linear Regressor	0.87 ± 0.05	6.38 ± 2.93	4.49 ± 2.46	27.30 ± 25.00	-1.89 ± 2.08
Neural Network	0.88 ± 0.06	3.27 ± 1.02	1.47 ± 0.70	3.37 ± 2.50	0.58 ± 0.36
K-Nearest Neighbors	0.75 ± 0.17	4.09 ± 1.25	1.21 ± 0.45	2.96 ± 2.68	0.64 ± 0.36
Baseline	0.00 ± 0.00	5.02 ± 0.73	2.90 ± 0.34	10.42 ± 2.43	-0.11 ± 0.10


explained_variance_score
1.0
0.8
0.6
0.4
0.2
0.0
K Neighbors
Decision Tree
Linear Regression
Random Forest
Neural Network
Baseline

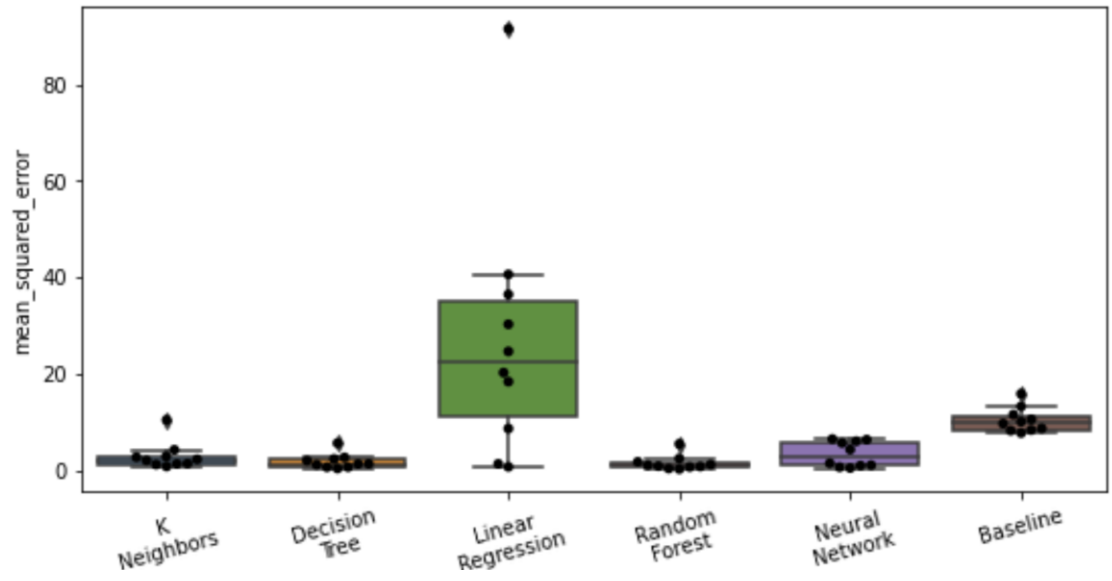
(a)


max_error
12
10
8
6
4
2
K Neighbors
Decision Tree
Linear Regression
Random Forest
Neural Network
Baseline

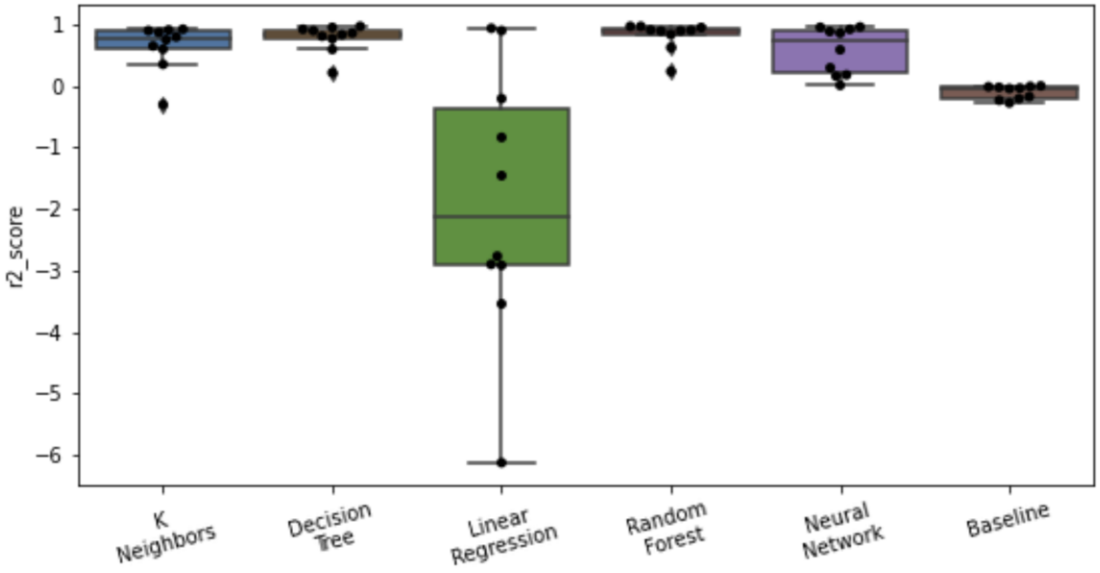
(b)


mean_absolute_error
8
6
4
2
0
K Neighbors
Decision Tree
Linear Regression
Random Forest
Neural Network
Baseline

(c)



(d)



(e)

Figure 61: Associated box plots for the leave-one-participant-out cross-validation of all the regressors used for predicting the fatigue level. a) EVS, b) Max error, c) Mean absolute error, d) Mean squared error, e) r2 score.

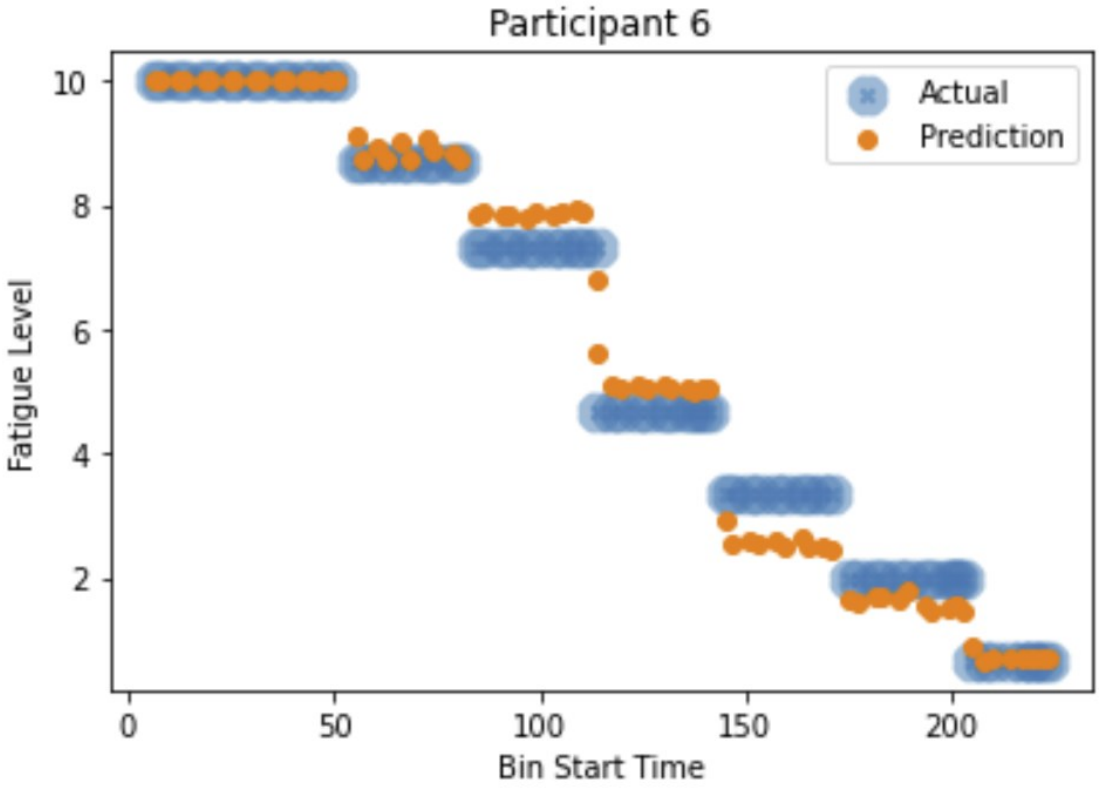


Figure 62: Comparison of the predicted fatigue level and the actual data set for participant 6 using the random forest regressor and the model from leave-one-participant-out cross-validation. (Fatigue level of 10 refers to least fatigued)

Further investigation into the performance of the RF regressor was conducted. Figure 62 illustrates the performance of the RF regressor on the dataset from participant 6. The model successfully predicts the fatigue level in a way that the levels are clear. It underestimates the fatigue level prior to the fatigue threshold of 4, and therefore, the error is within the patient's endurance. Moreover, the model overestimates the fatigue level after the threshold which is a desirable behavior to prevent the patient from excessive exercise.

4.3.2.3 Fatigue Level Detection Methods Comparison

- In both cross-validation methods, the RF regressor performs better than the other regressors followed by the DT regressor.
- Due to a more limited distribution of the data in the testing sets, the LOPO-CV validation approach metrics imply less performance than the 5F-CV. However, the fatigue level recognition has been performed with low error values and a high correlation with the DT and RF regressors. The RF regressor has performed the best, followed closely by the decision tree regressor.
- When comparing the RF and DT regressors between 5F-CV and LOPO-CV:
    - The spread of the mean errors in the LOPO-CV is larger (by a factor of 4) than the 5F-CV; however, the max error is almost similar in both analyses. This is a good sign as it indicates both training sets performed similarly despite the small testing set in the LOPO-CV. The low r^2 score in the LOPO-CV could potentially be related to the smaller testing set.
    - The maximum error in both approaches is less than 3, indicating that the fatigue level is predicted incorrectly by 3 points in a 1-10 rating scheme. Although this is a large number and could potentially overestimate the fatigue level, implementation of some failsafe mechanisms (such as prediction monitoring) when implementing the model could prevent any undesired outcome.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

With the goal of detecting and quantifying lower limb muscle fatigue during rehabilitation and therapy sessions for stroke patients while using rehabilitation robots like ViGRR-lite, this research developed and investigated a procedure suitable for clinical adaptation. An exergame was adopted to study the fatigue behavior of four different muscles during consecutive squat motions on ten healthy subjects. A squat motion resembles the sit-to-stand motion while promoting the speed of reaching fatigue in healthy subjects. Fatigue detection was then investigated using different machine learning methods to consider other factors such as participant physical information, perceived fatigue, and game data in addition to muscle activity data. Two main steps were taken to achieve the goal.

Test Procedure Implementation and EMG Analysis Development

After completing the experimental trial, Rectus Femoris (RF) and Biceps Femoris (BF) on the thigh, Gastrocnemius Medialis (GM) and Tibialis Anterior (TA) on the shank were selected to collect muscle activity data by using EMG system. The EMG data were analyzed by binning the data in three ways: First, constant non-overlapping bins, second, constant overlapping bins, and

third, bins based on the phase of the squat. The phasing analysis yielded the best set of results consistent with the participants' comments. The following conclusions were achieved:

- No reliable trend was recognized in the data after performing the primary trial. There was at least a muscle for some of the participants that did not follow the expected decreasing trend for the Mean and Median Power Frequency (MNF and MDF).
- It seemed that gender has a strong impact on muscle activity and being fatigued. In this study, most of the female participants were getting fatigued in all their muscles, whereas the male participants were only fatigued in some of the muscles. Therefore, it was concluded that participant information (such as gender) must be included in the analysis.
- Some muscles also showed trends of decreased activity in the middle of the experiment; however, a slight recovery occurred potentially due to an unconscious change in the participants' motion.
- The lack of participants may be the main reason for the uncertainty in the data; however, the hidden trends can be identified by the machine learning process.
- The procedure that was developed for participant preparation and testing was examined during the primary trials and proved to be expandable for clinical applications while all possible steps were taken to prevent noisy EMG data.

Fatigue Detection and Quantification with Machine Learning

The previous investigation raised the importance of simultaneously considering all the available data (the game data, the participant information, and EMG data) in the fatigue detection process. Different machine learning algorithms have been trained to find the best classifier and regressor to reach the goals of this research. The onset of fatigue was detected based on the classification

analysis utilizing the labeled data (binary) as per the perceived fatigue level of 6 in a 1-to-10 scheme (10 being the most fatigued).

- The random forest classifier provided the most desirable metrics for fatigue onset detection.
    - The RF classifier included the best accuracy and lowest standard deviation in both 5F-CV and LOPO-CV among the rest of the classifiers with the performance of 0.94±0.02 and 0.85±0.06, respectively.
    - The false negative rate of this classifier was 4.40±3.83 in 5F-CV and 5.80±5.21 for LOPO-CV (compared to the true negative rate of 71.00±3.85 in 5F-CV, and 31.90±4.95 in LOPO-CV). This metric is essential to the safety of the patient since it represents instances that a fatigued patient was incorrectly detected to be non-fatigued. Implementation of failsafe mechanisms could minimize the associated risks.
    - The F1 score of the RF classifier was by far higher than the other classifiers in both cross-validations (0.94±0.02 and 0.81±0.16 in 5F-CV and LOPO-CV, respectively).
- The LOPO-CV resulted in metrics lower than the 5F-CV since the population of the testing set was higher in the latter. The 5F-CV was trained for a shorter amount of time; however, it may not have included a uniform distribution of the instances.
- Participant 7 metrics were an outlier in the LOPO-CV analyses. Although no specific reason was found, the entire LOPO-CV was repeated without participant 7.
    - Removing participant 7 from the analysis improved the accuracy and F1 score by 3% and 7%, respectively.

- The expected increase in the metrics, such as the F1 score, was not grand; another indication that decreasing the number of data, even when knowing the data is not great, does not help the analysis that is already lacking data.

For the fatigue level quantification, regression was utilized where the labels were based on the perceived fatigue data that the participants provided during the experiment.

- In fatigue level detection, the RF regressor was performing better than the other regressors in both 5F-CV and LOPO-CV.
  - The r^2 score of the RF regressor was 0.98±0.01 and 0.81±0.22 for 5F-CV and LOPO-CV, respectively. Similar to classification, the LOPO-CV analysis includes a smaller population for testing, and the 5F-CV may not capture a representative distribution of the data.
  - The maximum error associated with the RF regressor in 5F-CV and LOPO-CV are 1.93±0.77 and 2.38±0.76, respectively. This metric illustrates the maximum offset that the model could report from the actual data. The main issue occurs when the fatigue level prediction underestimates the actual fatigue. Further testing and data collection in addition to adopting a failsafe procedure, can reduce the risks.
- Closely after the RF regressor, the DT regressor represented impressive r^2 scores of 0.96±0.02 and 0.78±0.21 for the 5F-CV and LOPO-CV, respectively.
- The LOPO-CV metrics were calculated based on smaller test sets as opposed to the 5F-CV. The LOPO-CV model is more reliable at this point; however, as more data is added to the training and testing sets, the 5F-CV could also be used as it is more efficient to train.

Overall, the machine learning models for muscle fatigue detection and quantification were developed using the EMG data that had been collected from healthy participants who had undertaken an exergame focused on consecutive squat motions. Despite the inconclusive nature of the muscle activity data for detecting fatigue, the utilization of these data along with the game data and the participant information in a machine learning analysis provided means of detecting and quantifying muscle fatigue. This thesis elaborated on the use of the perceived fatigue on identification and quantification of muscle fatigue when multiple muscles are engaged in an activity.

5.2 Future Work

Further investigations are required to improve the reliability and functionality of fatigue detection. The following list includes some suggestions for future advancements upon this research:

- Increasing the number of participants and choosing from a more diverse demographic such as a broader range of age, weight, and height, as machine learning algorithms require plenty of data to train a reasonable and reliable model.
- Collect data from stroke patients while performing the experiment and analyzing the EMG data to have a better understanding of how their muscle activity data differs from healthy participants. Consequently, the algorithms can be retrained with the new data to have a realistic training set and get closer to the research goal.
- Add an adaptive controller to the game, which based on the changes in muscle fatigue, can modify the rest time duration of the game. Furthermore, it allows confirming the assumption that slowing down the game increases the playing time of the participants. For

this goal, there is a need to run the experiment two or three times for each participant and progressively modify the rest time in the game.

- Some participants found the game tiring after a while. Increasing the engagement of the game with some auditory effects or changing the background objects could be helpful to make the game more exciting for participants.
- Collecting data from the hip and back muscles can be helpful as some participants expressed tiredness in those areas during the exercise.
- Use the actual rehabilitation robot instead of the exergame and collect data to see how the selected models will perform on the collected data from the robot as the original goal of this research was to use this algorithm on a robot to help with the assistance rehabilitation robot controller.

References

[1] E. J. Benjamin *et al.*, "Heart Disease and Stroke Statistics—2018 Update: A Report From the American Heart Association," *Circulation*, vol. 137, no. 12, pp. E67–E492, Mar. 2018, doi: 10.1161/CIR.0000000000000558.

[2] D. Hebert *et al.*, "Canadian stroke best practice recommendations: Stroke rehabilitation practice guidelines, update 2015," *International Journal of Stroke*, vol. 11, no. 4, pp. 459–484, Jun. 2016, doi: 10.1177/1747493016643553.

[3] M. P. Barnes, B. H. Dobkin, and Julien. Bogousslavsky, "Recovery after stroke," p. 656, 2005.

[4] "The World Health Report 2003: Shaping the Future - R. Beaglehole, World Health Organization - Google Books." https://books.google.ca/books?hl=en&lr=&id=Vv-rOQZs_e0C&oi=fnd&pg=PR7&dq=The+World+Health+Report.+World+Health+Organization,+2003.&ots=2Bl2dRd6hc&sig=ASSn55SCCmwyJ_ygqR2b759FcE4&redir_esc=y#v=onepage&q=The%20World%20Health%20Report.%20World%20Health%20Organization%2C%202003.&f=false (accessed Nov. 20, 2021).

[5] Z. Warraich and J. A. Kleim, "Neural Plasticity: The Biological Substrate For Neurorehabilitation," *PM&R*, vol. 2, no. 12, pp. S208–S219, Dec. 2010, doi: 10.1016/J.PMRJ.2010.10.016.

[6] K. J. Chisholm, K. Klumper, A. Mullins, and M. Ahmadi, "A task oriented haptic gait rehabilitation robot," 2014, doi: 10.1016/j.mechatronics.2014.07.001.

[7] "View of Novel Concept of a Lower-limb Rehabilitation Robot Targeting Bed-bound Acute Stroke Patients."

https://proceedings.cmbes.ca/index.php/proceedings/article/view/885/875 (accessed Nov. 20, 2021).

[8] M. Gruet, J. Temesi, T. Rupp, P. Levy, G. Y. Millet, and S. Verges, "Stimulation of the motor cortex and corticospinal tract to assess human muscle fatigue," *Neuroscience*, vol. 231, pp. 384–399, Feb. 2013, doi: 10.1016/J.NEUROSCIENCE.2012.10.058.

[9] J. J. Wan, Z. Qin, P. Y. Wang, Y. Sun, and X. Liu, "Muscle fatigue: general understanding and treatment," *Experimental & Molecular Medicine*, vol. 49, no. 10, p. e384, Oct. 2017, doi: 10.1038/EMM.2017.194.

[10] K. Kroenke, D. R. Wood, A. D. Mangelsdorff, N. J. Meier, and J. B. Powell, "Chronic Fatigue in Primary Care: Prevalence, Patient Characteristics, and Outcome," *JAMA*, vol. 260, no. 7, pp. 929–934, Aug. 1988, doi: 10.1001/JAMA.1988.03410070057028.

[11] M. N. Silverman, C. M. Heim, U. M. Nater, A. H. Marques, and E. M. Sternberg, "Neuroendocrine and Immune Contributors to Fatigue," *PM&R*, vol. 2, no. 5, pp. 338–346, May 2010, doi: 10.1016/J.PMRJ.2010.04.008.

[12] K. B. Norheim, G. Jonsson, and R. Omdal, "Biological mechanisms of chronic fatigue," *Rheumatology*, vol. 50, no. 6, pp. 1009–1018, Jun. 2011, doi: 10.1093/RHEUMATOLOGY/KEQ454.

[13] S. C. Gandevia, "Spinal and supraspinal factors in human muscle fatigue," *Physiological Reviews*, vol. 81, no. 4, pp. 1725–1789, 2001, doi: 10.1152/PHYSREV.2001.81.4.1725/ASSET/IMAGES/LARGE/9J0410162026.JPEG.

[14] D. G. Allen and H. Westerblad, "Role of phosphate and calcium stores in muscle fatigue," *The Journal of Physiology*, vol. 536, no. 3, pp. 657–665, Nov. 2001, doi: 10.1111/J.1469-7793.2001.T01-1-00657.X.

[15] R. M. Enoka and J. Duchateau, “Muscle fatigue: what, why and how it influences muscle function,” *The Journal of Physiology*, vol. 586, no. 1, pp. 11–23, Jan. 2008, doi: 10.1113/JPHYSIOL.2007.139477.

[16] E. Z. Ross, N. Middleton, R. Shave, K. George, and A. Nowicky, “Corticomotor excitability contributes to neuromuscular fatigue following marathon running in man,” *Experimental Physiology*, vol. 92, no. 2, pp. 417–426, Mar. 2007, doi: 10.1113/EXPPHYSIOL.2006.035972.

[17] B. Bigland-Ritchie and J. J. Woods, “Changes in muscle contractile properties and neural control during human muscular fatigue,” *Muscle & Nerve*, vol. 7, no. 9, pp. 691–699, Dec. 1984, doi: 10.1002/MUS.880070902.

[18] B. Bigland Ritchie, D. A. Jones, G. P. Hosking, and R. H. T. Edwards, “Central and Peripheral Fatigue in Sustained Maximum Voluntary Contractions of Human Quadriceps Muscle,” *Clinical Science and Molecular Medicine*, vol. 54, no. 6, pp. 609–614, Jun. 1978, doi: 10.1042/CS0540609.

[19] S. C. Gandevia, “Spinal and supraspinal factors in human muscle fatigue,” *Physiological Reviews*, vol. 81, no. 4, pp. 1725–1789, 2001, doi: 10.1152/PHYSREV.2001.81.4.1725/ASSET/IMAGES/LARGE/9J0410162026.JPEG.

[20] N. K. Vøllestad, “Measurement of human muscle fatigue,” *Journal of Neuroscience Methods*, vol. 74, no. 2, pp. 219–227, Jun. 1997, doi: 10.1016/S0165-0270(97)02251-6.

[21] C. K. edito Pierce, Patricia A. author.; Enoka, Roger M. editor.; Gandevia, Simon C. editor.; McComas, Alan J. editor.; Stuart, Douglas G. editor.; Thomas, *Fatigue Neural and Muscular Mechanisms*.

[22] A. M. Mengshoel, N. Vøllestad, and O. T. Førre, “Pain and fatigue induced by exercise in fibromyalgia patients with fibromyalgia”, Accessed: Nov. 14, 2021. [Online]. Available: https://www.researchgate.net/publication/15611917

[23] “Electromyography (EMG) - Mayo Clinic.” https://www.mayoclinic.org/tests-procedures/emg/about/pac-20393913 (accessed Nov. 11, 2021).

[24] “Electromyography (EMG) | Johns Hopkins Medicine.” https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/electromyography-emg (accessed Nov. 11, 2021).

[25] H. KINETICS GoRdon RoBeRtSon, Gr. caldwell, J. haMill, G. KaMen, and S. whittleSey, “Research Methods in Biomechanics Second edition,” 2014, Accessed: Nov. 11, 2021. [Online]. Available: www.HumanKinetics.com

[26] “DELSYS ® Bagnoli TM EMG System User’s Guide”.

[27] J. R. Daube and D. I. Rubin, “Needle electromyography,” *Muscle & Nerve*, vol. 39, no. 2, pp. 244–270, Feb. 2009, doi: 10.1002/MUS.21180.

[28] D. I. Rubin, “Needle electromyography: Basic concepts,” *Handbook of clinical neurology*, vol. 160, pp. 243–256, Jan. 2019, doi: 10.1016/B978-0-444-64032-1.00016-3.

[29] J. E. Desmedt, “Computer-aided electromyography and expert systems,” p. 320, 1989, Accessed: Nov. 13, 2021. [Online]. Available: https://books.google.com/books/about/Computer_aided_Electromyography_and_Expe.html?id=SRhsAAAAMAAJ

[30] A. Alkan and M. Günay, “Identification of EMG signals using discriminant analysis and SVM classifier,” *Expert Systems with Applications*, vol. 39, no. 1, pp. 44–47, Jan. 2012, doi: 10.1016/J.ESWA.2011.06.043.

[31] M. Garcia, T. V.-R. andaluza de medicina del deporte, and undefined 2011, "Surface electromyography: Why, when and how to use it," *redalyc.org*, vol. 4, no. 1, pp. 17–28, 2011, Accessed: Nov. 13, 2021. [Online]. Available: https://www.redalyc.org/pdf/3233/323327665004.pdf

[32] R. H. Chowdhury, M. B. I. Reaz, M. A. bin Mohd Ali, A. A. A. Bakar, K. Chellappan, and T. G. Chang, "Surface Electromyography Signal Processing and Classification Techniques," *Sensors (Basel, Switzerland)*, vol. 13, no. 9, p. 12431, Sep. 2013, doi: 10.3390/S130912431.

[33] G. de Luca, "Fundamental Concepts in EMG Signal Acquisition," 2003.

[34] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7420–7431, Jun. 2012, doi: 10.1016/J.ESWA.2012.01.102.

[35] A. Phinyomark, S. Thongpanja, H. Hu, P. Phukpattaranont, and C. Limsakul, "The Usefulness of Mean and Median Frequencies in Electromyography Analysis," *Computational Intelligence in Electromyography Analysis - A Perspective on Current Applications and Future Challenges*, Oct. 2012, doi: 10.5772/50639.

[36] C. M. Smith *et al.*, "Effects of the innervation zone on the time and frequency domain parameters of the surface electromyographic signal," *Journal of Electromyography and Kinesiology*, vol. 25, no. 4, pp. 565–570, Aug. 2015, doi: 10.1016/J.JELEKIN.2015.04.014.

[37] C. J. de Luca and P. Contessa, "Biomechanical benefits of the onion-skin motor unit control scheme," *Journal of Biomechanics*, vol. 48, no. 2, pp. 195–203, Jan. 2015, doi: 10.1016/J.JBIOMECH.2014.12.003.

[38] C. M. Smith *et al.*, "Combining regression and mean comparisons to identify the time course of changes in neuromuscular responses during the process of fatigue," *Physiological Measurement*, vol. 37, no. 11, pp. 1993–2002, Oct. 2016, doi: 10.1088/0967-3334/37/11/1993.

[39] L. A. C. Kallenberg, E. Schulte, C. Disselhorst-Klug, and H. J. Hermens, "Myoelectric manifestations of fatigue at low contraction levels in subjects with and without chronic pain," *Journal of Electromyography and Kinesiology*, vol. 17, no. 3, pp. 264–274, Jun. 2007, doi: 10.1016/J.JELEKIN.2006.04.004.

[40] J. H. T. Viitasalo and P. v. Komi, "Signal characteristics of EMG during fatigue," *European Journal of Applied Physiology and Occupational Physiology 1977 37:2*, vol. 37, no. 2, pp. 111–121, Jun. 1977, doi: 10.1007/BF00421697.

[41] N. Rahnama, A. Lees, and T. Reilly, "Electromyography of selected lower-limb muscles fatigued by exercise at the intensity of soccer match-play q," vol. 16, pp. 257–263, 2006, doi: 10.1016/j.jelekin.2005.07.011.

[42] M. Elshafei, D. E. Costa, and E. Shihab, "On the Impact of Biceps Muscle Fatigue in Human Activity Recognition," *Sensors 2021, Vol. 21, Page 1070*, vol. 21, no. 4, p. 1070, Feb. 2021, doi: 10.3390/S21041070.

[43] K. M. Chang, S. H. Liu, and X. H. Wu, "A Wireless sEMG Recording System and Its Application to Muscle Fatigue Detection," *Sensors 2012, Vol. 12, Pages 489-499*, vol. 12, no. 1, pp. 489–499, Jan. 2012, doi: 10.3390/S120100489.

[44] "Development of an EMG-based exergaming system for isometric muscle training and its effectiveness to enhance motivation, performance and muscle strength | Elsevier Enhanced Reader."

https://reader.elsevier.com/reader/sd/pii/S1071581918306748?token=8AD8A1D482B6561C9D21DA35D1992C43C72582988494A8D147CF992C66AEFEEB464629024EE611CFFF20FB0562E07104&originRegion=us-east-1&originCreation=20211216003049 (accessed Dec. 14, 2021).

[45] P. Soltani, P. Figueiredo, R. J. Fernandes, and J. P. Vilas-Boas, "Muscle activation behavior in a swimming exergame: Differences by experience and gaming velocity," *Physiology & Behavior*, vol. 181, pp. 23–28, Nov. 2017, doi: 10.1016/J.PHYSBEH.2017.09.001.

[46] J. Zhang, T. E. Lockhart, and R. Soangra, "Classifying Lower Extremity Muscle Fatigue During Walking Using Machine Learning and Inertial Sensors", doi: 10.1007/s10439-013-0917-0.

[47] M. Papakostas, H. Lab, V. Kanal, M. Abujelala, and K. Tsiakas, "Physical Fatigue Detection through EMG wearables and Subjective User Reports - A Machine Learning Approach Towards Adaptive Rehabilitation," vol. 19, doi: 10.1145/3316782.3322772.

[48] "1. The Machine Learning Landscape | Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition." https://learning.oreilly.com/library/view/hands-on-machine-learning/9781492032632/ch01.html#idm45022180582760 (accessed Aug. 29, 2021).

[49] T. M. (Tom M. Mitchell, "Machine Learning (McGraw-Hill International Editions Computer Science Series): Tom M. Mitchell: 9780071154673," p. 414, 1997.

[50] X.-D. Zhang, "Machine Learning," *A Matrix Algebra Approach to Artificial Intelligence*, pp. 223–440, 2020, doi: 10.1007/978-981-15-2770-8_6.

[51] I. el Naqa and M. J. Murphy, “What Is Machine Learning?,” *Machine Learning in Radiation Oncology*, pp. 3–11, 2015, doi: 10.1007/978-3-319-18305-3_1.

[52] “What is Supervised Learning?” https://searchenterpriseai.techtarget.com/definition/supervised-learning (accessed Aug. 29, 2021).

[53] “What is Supervised Learning? | IBM.” https://www.ibm.com/cloud/learn/supervised-learning (accessed Aug. 29, 2021).

[54] P. Cunningham, M. Cord, and S. J. Delany, “Supervised Learning,” *Cognitive Technologies*, pp. 21–49, 2008, doi: 10.1007/978-3-540-75171-7_2.

[55] A. Burkov, “The hundred-page machine learning book,” p. 141.

[56] “Emerging Artificial Intelligence Applications in Computer Engineering: Real ... - Google Books.” https://books.google.ca/books?hl=en&lr=&id=vLiTXDHr_sYC&oi=fnd&pg=PA3&dq=what+is+classification+in+machine+learning&ots=CZnrxv2Kip&sig=XoOdSQh1HChc9HDNVzHi_Ud8-vE#v=onepage&q=what%20is%20classification%20in%20machine%20learning&f=false (accessed Aug. 30, 2021).

[57] B. Heung, H. C. Ho, J. Zhang, A. Knudby, C. E. Bulmer, and M. G. Schmidt, “An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping,” *Geoderma*, vol. 265, pp. 62–77, Mar. 2016, doi: 10.1016/J.GEODERMA.2015.11.014.

[58] “Foundations of Machine Learning, second edition - Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar - Google Books.”

https://books.google.ca/books?hl=en&lr=&id=dWB9DwAAQBAJ&oi=fnd&pg=PR5&dq=what+is+machine+learning&ots=AyqPXTw-l3&sig=Gcnaq4dHTqe_txlBNlZZUpX50DI#v=onepage&q=what%20is%20machine%20learning&f=false (accessed Aug. 31, 2021).

[59] "What is Supervised Learning? | IBM." https://www.ibm.com/cloud/learn/supervised-learning (accessed Aug. 31, 2021).

[60] Trevor. Hastie, Robert. Tibshirani, and J. H. (Jerome H.) Friedman, "The elements of statistical learning : data mining, inference, and prediction : with 200 full-color illustrations," 2001.

[61] L. Wasserman, "All of statistics : a concise course in statistical inference," 2004.

[62] Trevor. Hastie, Robert. Tibshirani, and J. H. (Jerome H.) Friedman, "The elements of statistical learning : data mining, inference, and prediction : with 200 full-color illustrations," 2001.

[63] T. Pranckevičius and V. Marcinkevičius, "Comparison of Naïve Bayes, Random Forest, Decision Tree, Support Vector Machines, and Logistic Regression Classifiers for Text Reviews Classification," *Baltic J. Modern Computing*, vol. 5, no. 2, pp. 221–232, 2017, doi: 10.22364/bjmc.2017.5.2.05.

[64] "Emerging Artificial Intelligence Applications in Computer Engineering: Real ... - Google Books." https://books.google.ca/books?hl=en&lr=&id=vLiTXDHr_sYC&oi=fnd&pg=PA3&dq=what+is+classification+in+machine+learning&ots=CZnrxv2Kip&sig=XoOdSQh1HChc9HDNVzHi_Ud8-

vE#v=onepage&q=what%20is%20classification%20in%20machine%20learning&f=false (accessed Oct. 12, 2021).

[65] "Foundations of Machine Learning, second edition - Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar - Google Books." https://books.google.ca/books?hl=en&lr=&id=dWB9DwAAQBAJ&oi=fnd&pg=PR5&dq=what+is+machine+learning&ots=AyqPXTw-l3&sig=Gcnaq4dHTqe_txlBNlZZUpX50DI#v=onepage&q=LDA&f=false (accessed Oct. 12, 2021).

[66] P. Cunningham, M. Cord, and S. J. Delany, "Supervised Learning," *Cognitive Technologies*, pp. 21–49, 2008, doi: 10.1007/978-3-540-75171-7_2.

[67] "What is Supervised Learning? | IBM." https://www.ibm.com/cloud/learn/supervised-learning (accessed Oct. 12, 2021).

[68] "Designing Your Neural Networks. A Step by Step Walkthrough | by Lavanya Shukla | Towards Data Science." https://towardsdatascience.com/designing-your-neural-networks-a5e4617027ed (accessed Oct. 17, 2021).

[69] P. Cunningham, M. Cord, and S. J. Delany, "Supervised Learning," *Cognitive Technologies*, pp. 21–49, 2008, doi: 10.1007/978-3-540-75171-7_2.

[70] "Emerging Artificial Intelligence Applications in Computer Engineering: Real ... - Google Books." https://books.google.ca/books?hl=en&lr=&id=vLiTXDHr_sYC&oi=fnd&pg=PA3&dq=what+is+classification+in+machine+learning&ots=CZnrxv2Kip&sig=XoOdSQh1HChc9HDNVzHi_Ud8-

vE#v=onepage&q=what%20is%20classification%20in%20machine%20learning&f=false (accessed Oct. 17, 2021).

[71] S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: a methodology review," *Journal of Biomedical Informatics*, vol. 35, no. 5–6, pp. 352–359, Oct. 2002, doi: 10.1016/S1532-0464(03)00034-0.

[72] S. Weisberg, "Applied linear regression," p. 310, 2005, Accessed: Nov. 22, 2021. [Online]. Available: https://books.google.com/books/about/Applied_Linear_Regression.html?id=xd0tNdFOOjcC

[73] Trevor. Hastie, Robert. Tibshirani, and J. H. (Jerome H.) Friedman, "The elements of statistical learning : data mining, inference, and prediction : with 200 full-color illustrations," 2001.

[74] A. H. Al-Timemy, R. N. Khushaba, G. Bugmann, and J. Escudero, "Improving the Performance Against Force Variation of EMG Controlled Multifunctional Upper-Limb Prostheses for Transradial Amputees," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 6, pp. 650–661, Jun. 2016, doi: 10.1109/TNSRE.2015.2445634.

[75] V. Sze, Y. H. Chen, J. Einer, A. Suleiman, and Z. Zhang, "Hardware for machine learning: Challenges and opportunities," *Proceedings of the Custom Integrated Circuits Conference*, vol. 2017-April, Jul. 2017, doi: 10.1109/CICC.2017.7993626.

[76] "E X P E R T O P I N I O N 8," 2009, Accessed: Sep. 17, 2021. [Online]. Available: www.computer.org/intelligent

[77] T. Wuest, D. Weimer, C. Irgens, and K.-D. Thoben, “Machine learning in manufacturing: advantages, challenges, and applications,” *http://mc.manuscriptcentral.com/tpmr*, vol. 4, no. 1, pp. 23–45, Jun. 2016, doi: 10.1080/21693277.2016.1192517.

[78] A. H.- ECAI and undefined 1990, “General Limitations on Machine Learning.,” *mathtree.ru*, Accessed: Sep. 17, 2021. [Online]. Available: http://mathtree.ru/FileTransfer/form248618521_tmp.pdf

[79] A. Kumar, M. Boehm, and J. Yang, “Data management in machine learning: Challenges, techniques, and systems,” *Proceedings of the ACM SIGMOD International Conference on Management of Data*, vol. Part F127746, pp. 1717–1722, May 2017, doi: 10.1145/3035918.3054775.

[80] D. T. Pham and A. A. Afify, “Machine-learning techniques and their applications in manufacturing:,” *http://dx.doi.org/10.1243/095440505X32274*, vol. 219, no. 5, pp. 395–412, Aug. 2016, doi: 10.1243/095440505X32274.

[81] T. Wuest, D. Weimer, C. Irgens, and K.-D. Thoben, “Machine learning in manufacturing: advantages, challenges, and applications,” *http://mc.manuscriptcentral.com/tpmr*, vol. 4, no. 1, pp. 23–45, Jun. 2016, doi: 10.1080/21693277.2016.1192517.

[82] T. Wuest, D. Weimer, C. Irgens, and K.-D. Thoben, “Machine learning in manufacturing: advantages, challenges, and applications,” *http://mc.manuscriptcentral.com/tpmr*, vol. 4, no. 1, pp. 23–45, Jun. 2016, doi: 10.1080/21693277.2016.1192517.

[83] T. Wuest, D. Weimer, C. Irgens, and K.-D. Thoben, “Machine learning in manufacturing: advantages, challenges, and applications,” *http://mc.manuscriptcentral.com/tpmr*, vol. 4, no. 1, pp. 23–45, Jun. 2016, doi: 10.1080/21693277.2016.1192517.

[84] “Missing Data: Analysis and Design - John W. Graham - Google Books.” https://books.google.ca/books?hl=en&lr=&id=AO6k9XBaLpIC&oi=fnd&pg=PT19&ots=i1CMM4qYOH&sig=f8JS9oJgnQ-jm9WsXDiFucxZwds&redir_esc=y#v=onepage&q&f=false (accessed Sep. 17, 2021).

[85] “How, When, and Why Should You Normalize / Standardize / Rescale Your Data? – Towards AI — The Best of Tech, Science, and Engineering.” https://towardsai.net/p/data-science/how-when-and-why-should-you-normalize-standardize-rescale-your-data-3f083def38ff (accessed Sep. 26, 2021).

[86] “Data Transformation: Standardization vs Normalization - KDnuggets.” https://www.kdnuggets.com/2020/04/data-transformation-standardization-normalization.html (accessed Sep. 26, 2021).

[87] “Hyperparameter Optimization & Tuning for Machine Learning (ML) - DataCamp.” https://www.datacamp.com/community/tutorials/parameter-optimization-machine-learning-models (accessed Sep. 27, 2021).

[88] “Hyperparameter Optimization & Tuning for Machine Learning (ML) - DataCamp.” https://www.datacamp.com/community/tutorials/parameter-optimization-machine-learning-models (accessed Sep. 27, 2021).

[89] “A Practical Guide To Hyperparameter Optimization.” https://nanonets.com/blog/hyperparameter-optimization/ (accessed Sep. 27, 2021).

[90] SimonRichard, “Supervised analysis when the number of candidate features (p) greatly exceeds the number of cases (n),” *ACM SIGKDD Explorations Newsletter*, vol. 5, no. 2, pp. 31–36, Dec. 2003, doi: 10.1145/980972.980978.

[91] D. Berrar, “Cross-Validation Call for Papers for Machine Learning journal: Machine Learning for Soccer View project Cross-validation”, doi: 10.1016/B978-0-12-809633-8.20349-X.

[92] “A Gentle Introduction to k-fold Cross-Validation.” https://machinelearningmastery.com/k-fold-cross-validation/ (accessed Sep. 27, 2021).

[93] “What is Cross Validation in Machine learning? Types of Cross Validation.” https://www.mygreatlearning.com/blog/cross-validation/ (accessed Sep. 27, 2021).

[94] “Accuracy, Precision, Recall or F1? | by Koo Ping Shung | Towards Data Science.” https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9 (accessed Sep. 19, 2021).

[95] “Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures - Exsilio Blog.” https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/ (accessed Sep. 19, 2021).

[96] “3.3. Metrics and scoring: quantifying the quality of predictions — scikit-learn 1.0.1 documentation.” https://scikit-learn.org/stable/modules/model_evaluation.html#regression-metrics (accessed Nov. 22, 2021).

[97] “Understanding Confusion Matrix | by Sarang Narkhede | Towards Data Science.” https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62 (accessed Oct. 03, 2021).

[98] A. Tharwat, “Classification assessment methods,” *Applied Computing and Informatics*, vol. 17, no. 1, pp. 168–192, 2018, doi: 10.1016/J.ACI.2018.08.003.

[99] T. Saito and M. Rehmsmeier, “The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets,” *PLoS ONE*, vol. 10, no. 3, Mar. 2015, doi: 10.1371/JOURNAL.PONE.0118432.

[100] “Understanding AUC - ROC Curve | by Sarang Narkhede | Towards Data Science.” https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5 (accessed Oct. 04, 2021).

[101] “Precision-recall curves – what are they and how are they used?” https://acutecaretesting.org/en/articles/precision-recall-curves-what-are-they-and-how-are-they-used (accessed Oct. 10, 2021).

[102] “DELSYS ® Bagnoli TM EMG System User’s Guide”.

[103] “Exercise intensity: How to measure it - Mayo Clinic.” https://www.mayoclinic.org/healthy-lifestyle/fitness/in-depth/exercise-intensity/art-20046887 (accessed Oct. 20, 2021).

[104] “Fitbit Inspire & Inspire HR | Health & Fitness Trackers.” https://www.fitbit.com/global/dk/products/trackers/inspire (accessed Dec. 18, 2021).

[105] “NI USB-6216 from National Instruments | SelectScience.” https://www.selectscience.net/products/ni-usb-6216/?prodID=85884 (accessed Nov. 30, 2021).

[106] “Bowflex® How-To | Squats for Beginners - YouTube.” https://www.youtube.com/watch?v=aclHkVaku9U&ab_channel=Bowflex (accessed Oct. 23, 2021).

[107] “Sensor Locations.” http://seniam.org/sensor_location.htm (accessed Oct. 23, 2021).

[108] "File:Muscles front and back.svg - Wikimedia Commons." https://commons.wikimedia.org/wiki/File:Muscles_front_and_back.svg (accessed Dec. 19, 2021).

[109] M. Halaki and K. Ginn, "Normalization of EMG Signals: To Normalize or Not to Normalize and What to Normalize to?," *Computational Intelligence in Electromyography Analysis - A Perspective on Current Applications and Future Challenges*, Oct. 2012, doi: 10.5772/49957.

[110] "IEEE Xplore Full-Text PDF:" https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7323310 (accessed Jan. 20, 2022).

[111] C. J. de Luca, L. Donald Gilmore, M. Kuznetsov, and S. H. Roy, "Filtering the surface EMG signal: Movement artifact and baseline noise contamination," *Journal of Biomechanics*, vol. 43, no. 8, pp. 1573–1579, May 2010, doi: 10.1016/J.JBIOMECH.2010.01.027.

[112] "Squats Female Exercise Guide Silhouettes. Squatting Woman Illustration an Athletic Young Woman in Sportswear Top, Leggings, and Sneakers Does the Weight Loss Workout. #412234932 - Larastock." https://larastock.com/deposit-photo-412234932/?url=search%2Fsquatting%2F (accessed Oct. 27, 2021).

[113] "VMO Muscle & Knee Rehabilitation - Sportsinjuryclinic.net." https://www.sportsinjuryclinic.net/rehabilitation-exercises/knee-hamstring-thigh-exercises/vmo-rehab (accessed Oct. 29, 2021).

[114] O. Dupuy, W. Douzi, D. Theurot, L. Bosquet, and B. Dugué, "An evidence-based approach for choosing post-exercise recovery techniques to reduce markers of muscle

damage, Soreness, fatigue, and inflammation: A systematic review with meta-analysis," *Frontiers in Physiology*, vol. 9, no. APR, p. 403, Apr. 2018, doi: 10.3389/FPHYS.2018.00403/BIBTEX.

[115] N. S. Artzi *et al.*, "Prediction of gestational diabetes based on nationwide electronic health records," *Nature Medicine*, vol. 26, no. 1, pp. 71–76, Jan. 2020, doi: 10.1038/s41591-019-0724-8.

[116] J. Goecks, V. Jalili, L. M. Heiser, and J. W. Gray, "How Machine Learning Will Transform Biomedicine," *Cell*, vol. 181, no. 1, pp. 92–101, Apr. 2020, doi: 10.1016/j.cell.2020.03.022.

[117] J. Chan, T. Rea, S. Gollakota, and J. E. Sunshine, "Contactless cardiac arrest detection using smart devices," *npj Digital Medicine*, vol. 2, no. 1, p. 52, Dec. 2019, doi: 10.1038/s41746-019-0128-7.

[118] J. H. T. Viitasalo and P. v Komi, "Applied Physiology Signal Characteristics of EMG during Fatigue," vol. 121, pp. 111–121, 1977.

[119] M. Bilodeau, S. Schindler-Ivens, D. M. Williams, R. Chandran, and S. S. Sharma, "EMG frequency content changes with increasing force and during fatigue in the quadriceps femoris muscle of men and women," *Journal of Electromyography and Kinesiology*, vol. 13, no. 1, pp. 83–92, Feb. 2003, doi: 10.1016/S1050-6411(02)00050-0.

[120] "Life Insurance And BMI | How Body Mass Index Affects Insurance Rates." https://www.thelifeinsuranceblog.com/life-insurance-and-bmi/ (accessed Aug. 24, 2021).

[121] A. F. Roche *et al.*, "Relationships Between the Body Mass Index and Body Composition," *Obesity Research*, vol. 4, no. 1, pp. 35–44, Jan. 1996, doi: 10.1002/J.1550-8528.1996.TB00510.X.

[122] J. J. Reilly, “Diagnostic accuracy of the BMI for age in paediatrics,” *International Journal of Obesity 2006 30:4*, vol. 30, no. 4, pp. 595–597, Mar. 2006, doi: 10.1038/sj.ijo.0803301.

[123] M. McDonagh, M. White, C. D.- Gerontology, and undefined 1984, “Different effects of ageing on the mechanical properties of human arm and leg muscles,” *karger.com*, Accessed: Aug. 24, 2021. [Online]. Available: https://www.karger.com/Article/Abstract/212606

[124] B. MG, M. BH, B. DA, M. JE, and B. RA, “Isometric muscle force production as a function of age in healthy 20- to 74-yr-old men.,” *Medicine and Science in Sports and Exercise*, vol. 23, no. 11, pp. 1302–1310, Nov. 1991, Accessed: Aug. 24, 2021. [Online]. Available: https://europepmc.org/article/med/1766348

[125] R. Merletti, D. Farina, M. Gazzoni, and M. P. Schieroni, “Effect of age on muscle functions investigated with surface electromyography,” *Muscle & Nerve*, vol. 25, no. 1, pp. 65–76, Jan. 2002, doi: 10.1002/MUS.10014.

[126] L. Larsson, G. Grimby, and J. Karlsson, “Muscle strength and speed of movement in relation to age and muscle morphology”.

[127] “scikit-learn: machine learning in Python — scikit-learn 1.0.1 documentation.” https://scikit-learn.org/stable/ (accessed Dec. 04, 2021).

[128] “View of Novel Concept of a Lower-limb Rehabilitation Robot Targeting Bed-bound Acute Stroke Patients.” https://proceedings.cmbes.ca/index.php/proceedings/article/view/885/875 (accessed Oct. 20, 2021).

Appendices

Appendix A : Perceived Fatigue Form

This is the perceived fatigue form attached here:

Time (min:sec)	Fatigue Score 1-10	Time (min:sec)	Fatigue Score 1-10
0:00		7:30	
0:30		8:00	
1:00		8:30	
1:30		9:00	
2:00		9:30	
2:30		10:00	
3:00		10:30	
3:30		11:00	
4:00		11:30	
4:30		12:00	
5:00		12:30	
5:30		13:00	
6:00		13:30	
6:30		14:00	
7:00		14:30	

Subjective Muscle Fatigue Scores

Subject Number _____

Date _____

Start Time _____

Researcher _____

Appendix B : Email Invitation

Here is the email invitation which is discussed on section 3.5.

Email Invitation

Hello,

We are Mahdokht Golmohammadishouraki and Nick Berezny, Master's and PhD students in the Department of Mechanical Engineering at Carleton University. We are working on a research project under the supervision of Prof. Mojtaba Ahmadi.

We are writing to you today to invite you to participate in a study entitled "Experiments with Haptic Gait Rehabilitation Robotic Platforms." This study aims to research muscle activity while using a gait rehabilitation robot. Due to COVID-19 situation, we perform a modified set of experiments that are done remotely without compromising safety. We will supervise the experiment using teleconferencing software (Microsoft team) to assist with the setup and run and monitor the entire test. Your data will be stored and protected by Microsoft Team in North America, (and for Canadians the data centers are in Quebec City and Toronto) , but may be disclosed via a court order or data breach. We drop off the package to you and collect it two days after the experiment is concluded. The box and the contents will be sanitized by the researchers after receiving and before sending it to the next participant. We suggest you sanitize the box and the contents before and after experimenting. The equipment and the box will be left in the open air for at least four days to ensure a safe transition from one participant to the next by the researchers. During the experiment, you will wear surface EMG sensors, which will be placed on your leg muscles. The EMG sensor needs skin preparation, including cleaning and shaving small areas where the electrodes will be placed. There is no risk of shock because the electrodes are not active. After that, you will be asked to do some squats to play a simple game based on your strength to reach a moderate level of fatigue. We will use a Fitbit to monitor your heart rate to make sure the squats are completed comfortably and safely. The results of this experiment will help us improve our system and give further insight into the field of stroke rehabilitation. You can withdraw from the study at any time and for any reason up until five days after the trial. If you choose to withdraw, all the information you have provided will be destroyed. After the withdrawal deadline, the data will be anonymized when any information relating to the participants will be destroyed. All research data, including audio-recordings and any notes, will be password-protected.

The ethics protocol for this project was reviewed by the Carleton University Research Ethics Board, which provided clearance to carry out the research. (Clearance expires on: July 31, 2020) If you have any ethical concerns with the study, please contact the Carleton University Research Ethics Board-B (by phone at 613-520-2600 ext. 4085 or via email at ethics@carleton.ca).

If you would like to participate in this research project or have any questions, please contact us, mahdokhtgolmohammadi@carleton.ca and Nicholas.berezny@carleton.ca.

Sincerely,

Mahdokht Golmohammadishouraki

Nick Berezny

Appendix C : Questionnaire

As Discussed in section 3.6, these are the questionnaires which had been used for the experiment.

C.1 Pre-Questionnaire

Subject ID: Date:

Pre-interview questionnaire

1. Have you done moderate exercise in the past month?

2. Do you have any conditions or health issues that might affect your ability to perform moderate exercises? Such as heart conditions, arthritis, history of stroke, or other relevant illnesses.

C.2 Post-Questionnaire

Subject ID: Date:

Post-Questionnaire (Experiment #1)

(1) Did the game help you not notice being tired (and continue to do more squats)? Please explain:

(2) Which part of your leg got tired first? In which muscle do you feel more tired or fatigued?

(3) Any other comments on the game?

Appendix D : Consent Form

The attached is the Consent Form which the participant should sign before starting the experiment to make sure they know exactly how the experiment is going to be and what safety and risk are they may face.

Carleton UNIVERSITY
Canada's Capital University

Research Consent Form

Name and Contact Information of Researchers:

Researcher:
Nicholas Berezny
Department of Mechanical and Aerospace Engineering
Carleton University
1125 Colonel By Drive
Ottawa ON
K1S 5B6
Tel: 613 314 0480
Email: Nicholas.berezny@carleton.ca

Researcher:
Mahdokht Golmohammadishouraki
Department of Mechanical and Aerospace Engineering
Carleton University
1125 Colonel By Drive
Ottawa ON
K1S 5B6
Tel: 613 869 4837
Email: Mahdokhtgolmohammadi@carleton.ca

Supervisor:
Professor Mojtaba Ahmadi
Department/of Mechanical and Aerospace Engineering
Carleton University
E-mail Address: mahmadi@mae.carleton.ca
Telephone: 613-520-2600 x4057

Project Title:

Experiment with Haptic Gait Rehabilitation Robotic Platform

Carleton University Project Clearance

Clearance #: 10090 11-1631 Date of Clearance: July 12, 2019

Carleton UNIVERSITY
Canada's Capital University

Invitation:

You are invited to take part in a research project because you are in good health and capable of doing leg exercises. The information in this form is intended to help you understand the experiments so that you can decide whether to agree or not in participating. Your participation in this study is voluntary, and a decision not to participate will not be used against you in any way. As you read this form, and decide whether to participate, please ask all the questions you might have, take whatever time you need, and consult with others as you wish.

What is the purpose of the study?

Stroke rehabilitation seeks to restore motor control to patients in order to increase their independence when they leave the hospital. This is typically carried out by physio- and occupational therapists manually manipulating the patient's limbs through a series of exercises without knowing the muscle's activity. Rehabilitation robots can be used as tools by therapists to provide more exercise at a higher intensity, which has been linked to more significant clinical outcomes, and by using an EMG sensor, there would be useful data for researchers and therapists to monitor the patient's progress by recording their muscles' activity. This study evaluates a new lower-limb robotic technology which uses visualizations, games, and force feedback and progress monitoring strategies to improve the user experience, accelerate recovery, and provide useful data for therapists and researchers.

During COVID-19 quarantine and physical distancing period, we would not do any experiments that involved any human-to-human physical contact. However, we can perform a modified set of experiments that are done remotely without compromising safety. This will enable the collection of basic data that we need to continue the research until restrictions are lifted.

What will I be asked to do?

If you agree to take part in the study, we will ask you to:

Participate in an experiment of the robotic system with a simple simulated gaming activity that is run by the user performing basic waist up-down motions as an input (moderated squats) with EMG(Electromyography). For this, we have developed a vision-based body motion system that is very light weight and easy to set up. We will participate using teleconferencing software (Microsoft Teams) to be able to assist with the setup and run and monitor the entire experiment. Your data will be stored and protected by Microsoft Team in North America, (and for Canadians the data centers are in Quebec City and Toronto) , but may be disclosed via a court order or data breach.

We drop off the package to you and collect it two days after the experiment is concluded. The box and the contents will be sanitized by the researchers after receiving and before sending it to the next participant.

Carleton UNIVERSITY
Canada's Capital University

We suggest you sanitize the box and the contents before and after experimenting.The equipment and the box will be left in the open air for at least four days to ensure a safe transition from one participant to the next by the researchers. You will attach the sensor to your muscles, and use waist vertical movements to play the simple game. The data will be collected from the EMG and body motion sensors, as well as the game performance data.

The experiment consists of an exergame, or a virtual video game which encourages you to do squats. A webcam and marker is used to track the up and down motion of you as you squat, which is used to control the virtual character in the game. You will wear passive EMG electrodes, which are used to get this measure of muscle fatigue. During the game, you will be periodically asked to rate their perceived muscle fatigue on a numerical scale. The game will be played over short intervals of 5-15 minutes, and you can stop at any time, or can request a break if you are feeling overwhelmed or over-exerted. As mentioned, a heart rate monitor will display your heart rate at all times, and the experiment will be paused if it exceeds recommended rates for moderate exercise.

There will be questionnaires, as well as a brief interview after the experiment related to your experience. There are 9 steps to the experiment, for a total estimated time of 1hr 30 minutes. This will be done twice, separated by approximately 1 month.

1. Connect via Video Chat + Pre-questionnaire (10 min)
2. Game setup - orienting the webcam, attaching the markers, running software (20 min)
3. Game testing and practice (5 min)
4. EMG Electrode setup - locating correct muscle locations, attaching electrodes, testing (20 min)
5. Play the game (for 5-15min)
6. Removing the electrodes/shut down (10 min)
7. Post-questionnaire (10 min)

Risks and Inconveniences:

Since the EMG sensors are not active, there is no risk of shock during the experiment. As a result of skin preparation for the EMG, you may experience irritation.

For more safety, we will add a heart rate sensor to monitor participants' heart rate. For each participant we will calculate their maximum heart rate according to Mayo Clinic protocol (https://www.mayoclinic.org/healthy-lifestyle/fitness/in-depth/exercise-intensity/art-20046887). The Maximum heart rate is calculated by subtracting the participant age from 220. Once we know their maximum heart rate, we can calculate the desired target heart rate zone - the level at which the heart is being exercised and conditioned but not overworked. The American Heart Association recommends a

Carleton UNIVERSITY
Canada's Capital University

target heart rate of 50% to about 70% of your maximum heart rate for moderate exercise. If the pulse reaches that beats per minute, we will pause the experiment to get back to the resting rate.

Note that the box and the contents will be sanitized by the researchers after receiving and before sending it to the next participant. The equipment and the box will be left in the open air for at least four days to ensure a very safe transition from one participant to the next by the researchers.

Possible Benefits

You may not receive any direct benefit from your participation in this study. However, your participation may allow researchers to better understand the merits and demerits of the current rehabilitation robotic system so that improvements can be made in future rehabilitation robotic technologies, which will in turn affect the efficiency of the current stroke rehabilitation practice.

Compensation/Incentives

You will not be paid or compensated for your participation in this study.

No waiver of your rights

By signing this form, you are not waiving any rights or releasing the researchers from any liability.

Withdrawing from the study

If you withdraw your consent during the course of the study, all information collected from you before your withdrawal will be discarded, unless you request that it be removed from the study data.

After the study, you may request that your data be removed from the study and deleted by notice given to the Principal Investigator (named above) within 5 days after your completion. Your data will be anonymized (any information that can match the data with the person will be destroyed) after the withdrawal period is over. The data will be used permanently by the researchers for further investigation and publication purposes.

Confidentiality

We will treat your personal information as confidential, although absolute privacy cannot be guaranteed. No information that discloses your identity will be released or published without your specific consent.

Carleton UNIVERSITY
Canada's Capital University

Research records may be accessed by the Carleton University Research Ethics Board in order to ensure continuing ethics compliance.

You will be assigned a code so that your identity will not be directly associated with the data you have provided. All data, including coded information, will be kept in a password-protected file on a secure computer. Any information identifying your data will be destroyed after the withdrawal period is reached.

We will password protect any research data that we store or transfer.

Data Retention

After the study is completed, your de-identified data will be retained for future research use.

New information during the study

In the event that any changes could affect your decision to continue participating in this study, you will be promptly informed.

Ethics review

This project was reviewed and cleared by the Carleton University Research Ethics Board [B]. If you have any ethical concerns with the study, please contact Carleton University Research Ethics Board-B (by phone: 613-520-2600 ext. 4085 or by email: ethics@carleton.ca). For all other questions about the study, please contact the researcher."

Statement of consent – print and sign name

I voluntarily agree to participate in this study. ___Yes ___No

I agree to be contacted for follow up research ___Yes ___No

_____ _____

Signature of participant Date

Research team member who interacted with the subject

Carleton
UNIVERSITY
Canada's Capital University

I have explained the study to the participants and answered any and all of their questions. The participant appeared to understand and agree. I provided a copy of the consent form to the participant for their reference.

_____ _____

Signature of researcher Date

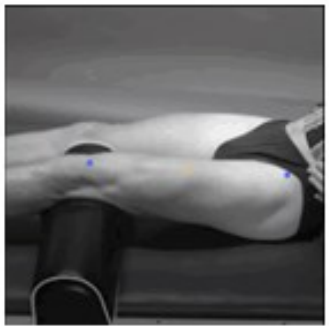
Appendix E : Sensor Placement

Muscle location for EMG sensors

1. Quadriceps Femoris (rectus femoris):

Location: The electrodes need to be placed at 50% on the line from the anterior spina iliaca superior to the superior part of the patella.

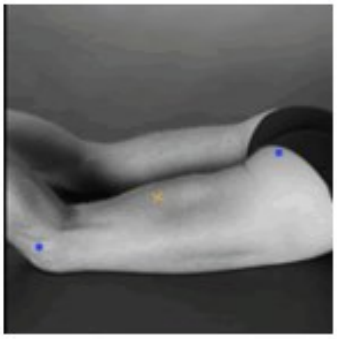
Orientation: In the direction of the line from the anterior spina iliaca superior to the superior part of the patella.



2. Biceps femoris:

Location: The electrodes need to be placed at 50% on the line between the ischial tuberosity and the lateral epicondyle of the tibia.

Orientation: In the direction of the line between the ischial tuberosity and the lateral epicondyle of the tibia.



3. Tibialis anterior:

Location: The electrodes need to be placed at 1/3 on the line between the tip of the fibula and the tip of the medial malleolus.

Orientation: In the direction of the line between the tip of the fibula and the tip of the medial malleolus.

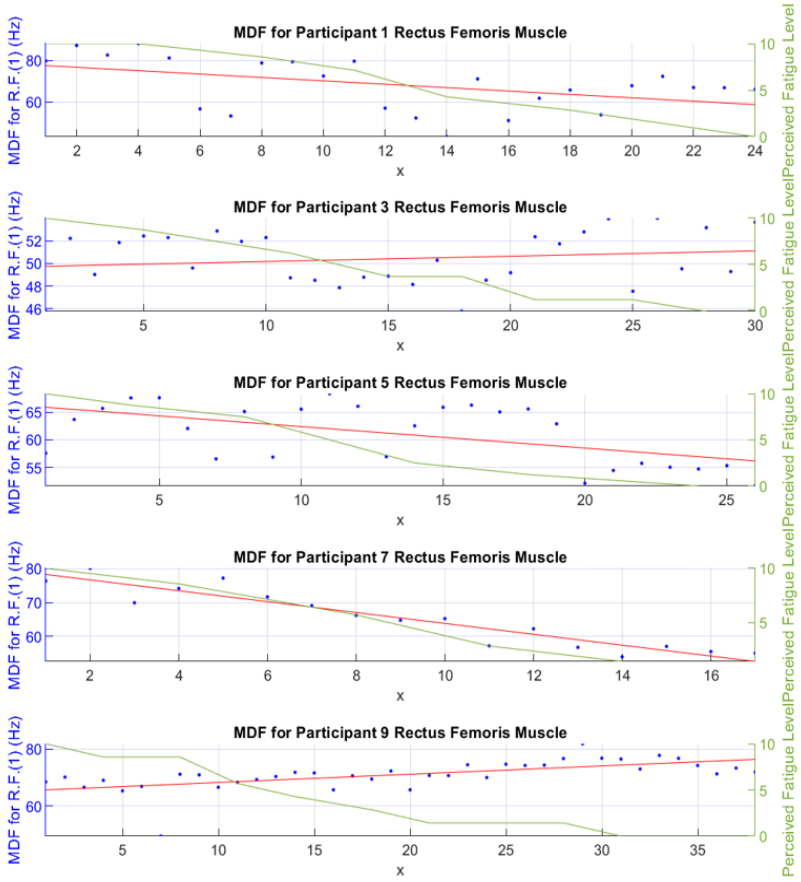
4. Gastrocnemius Medialis :

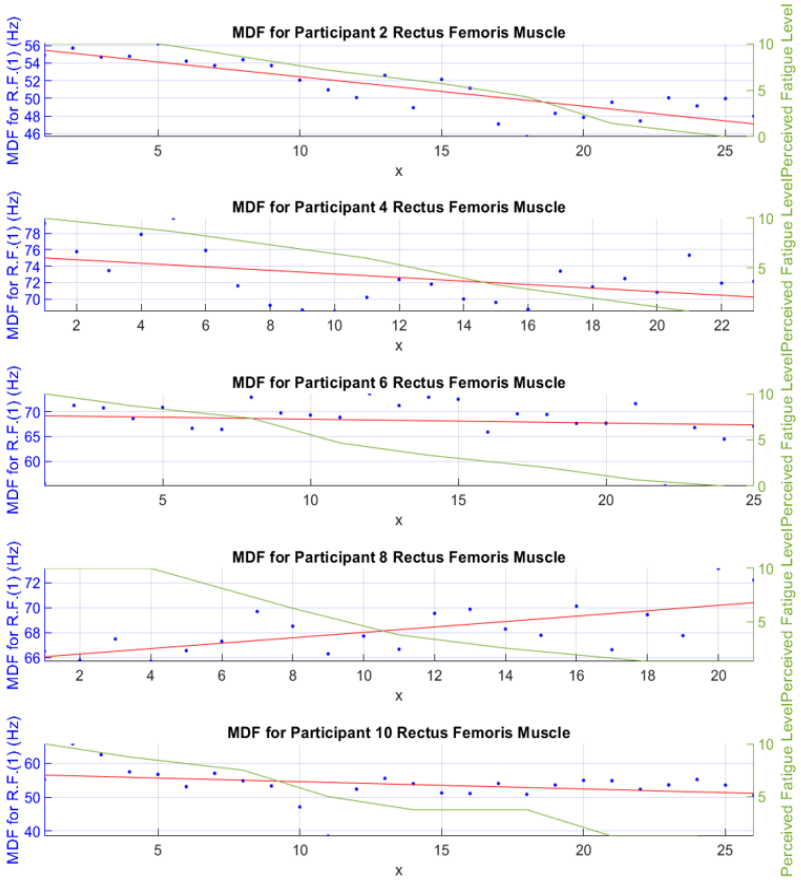
Location: Electrodes need to be placed on the most prominent bulge of the muscle.

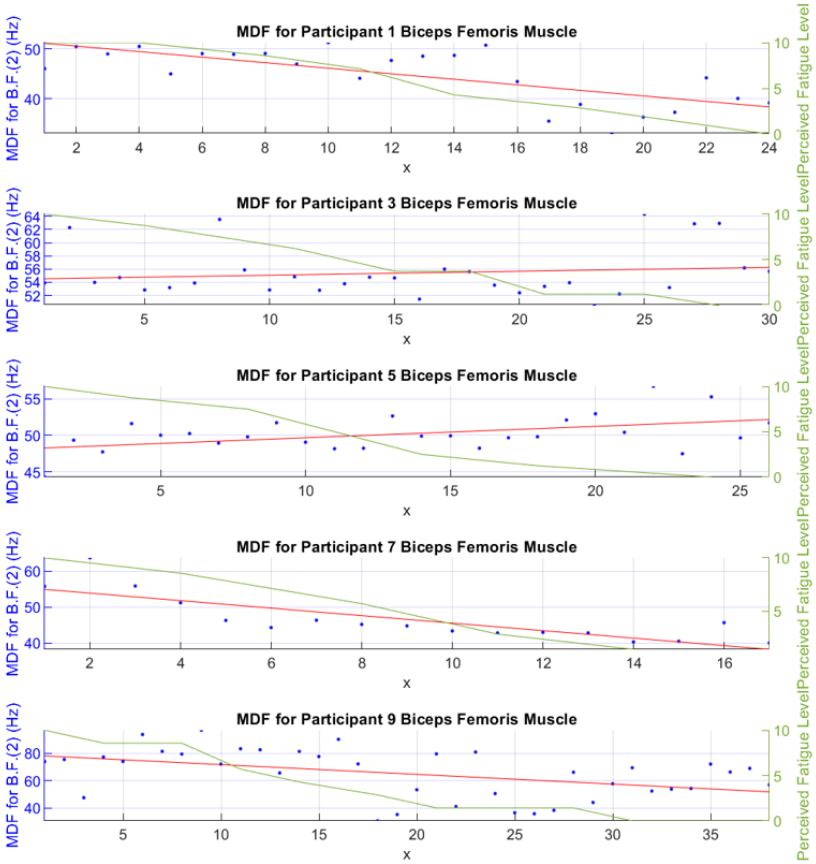
Orientation: In the direction of the leg.

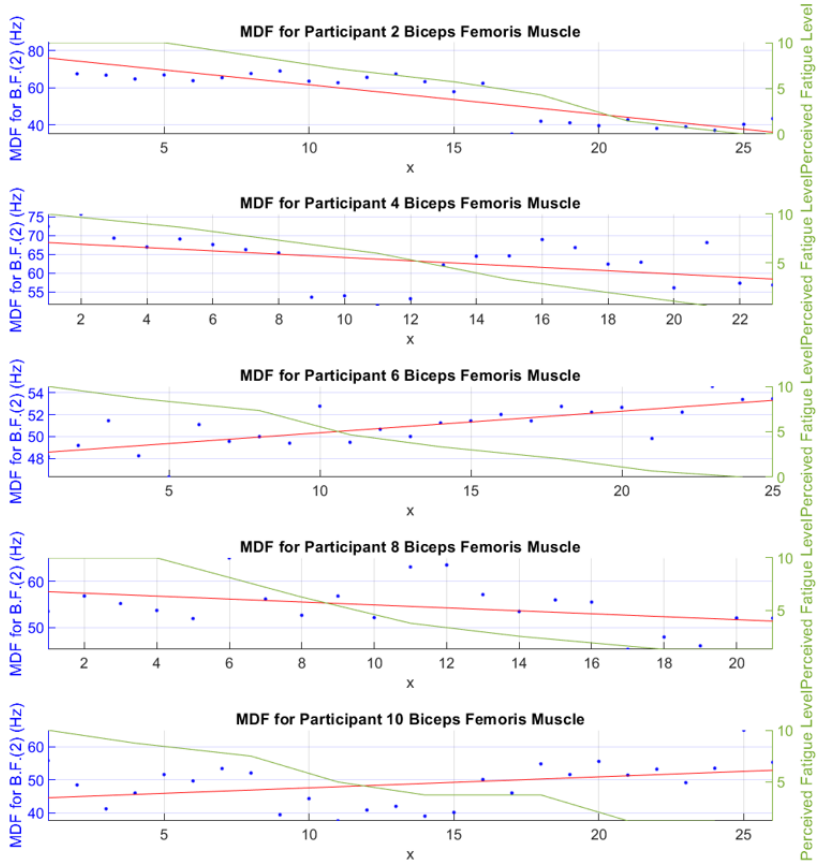
Appendix F : Primary Trial Analysis Results:

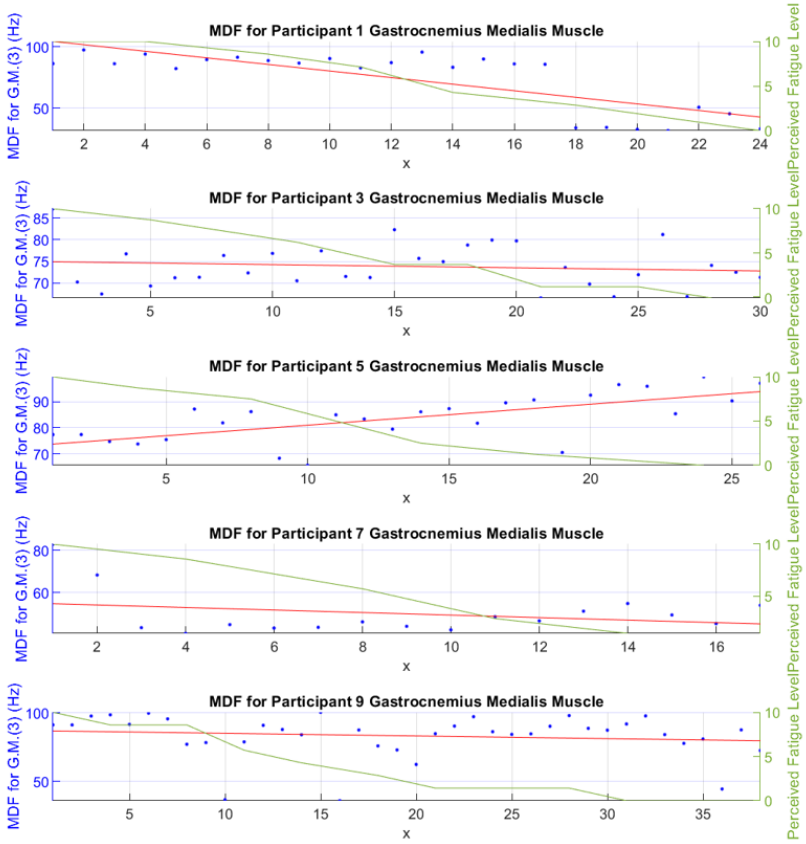
F.1 Discrete Window Analysis

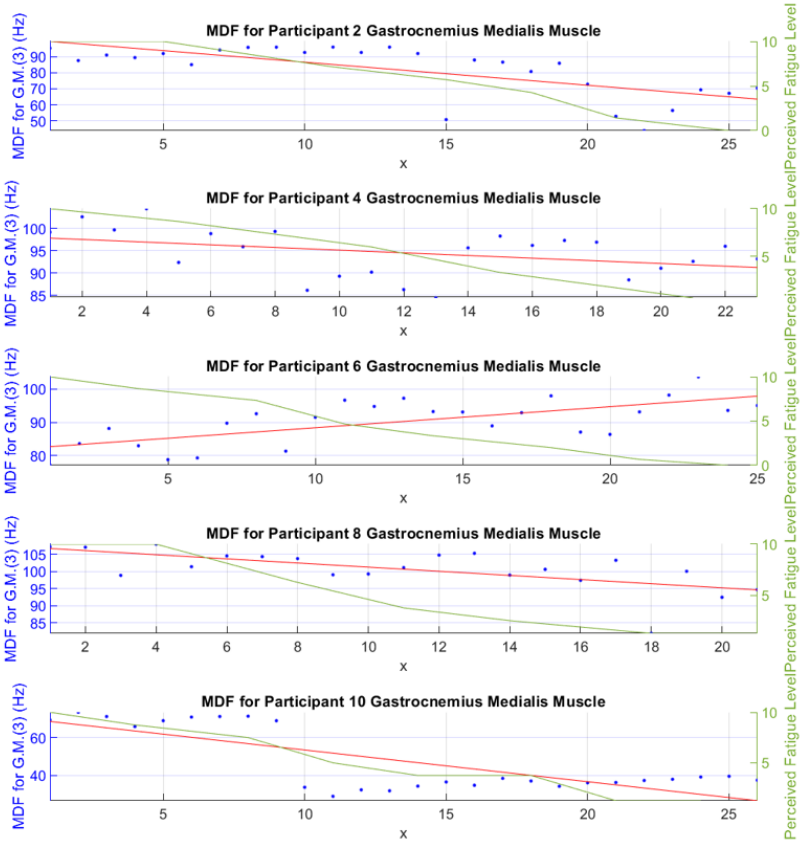


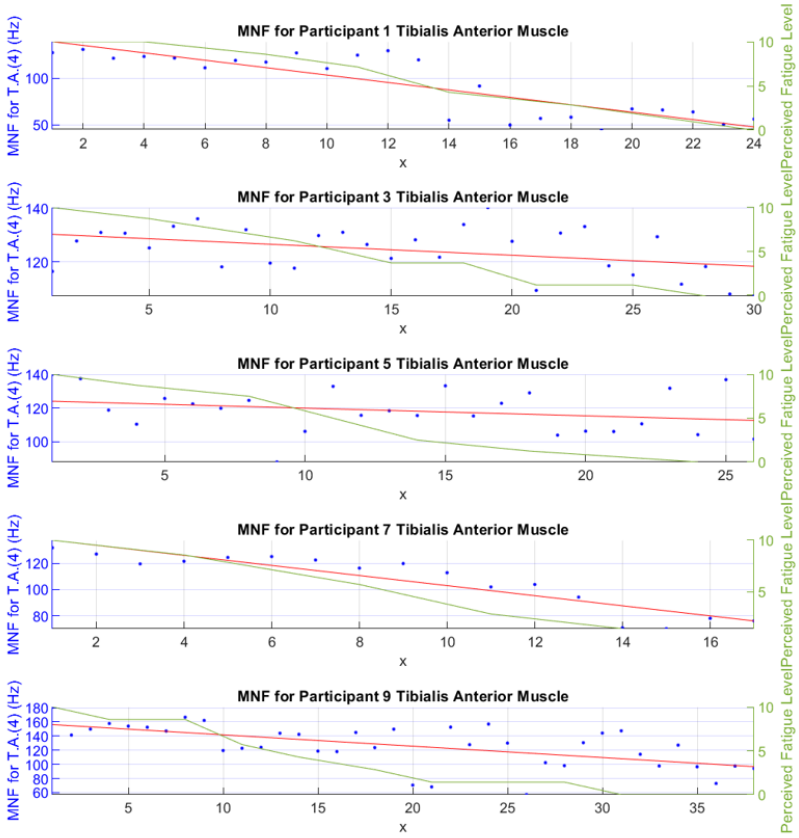


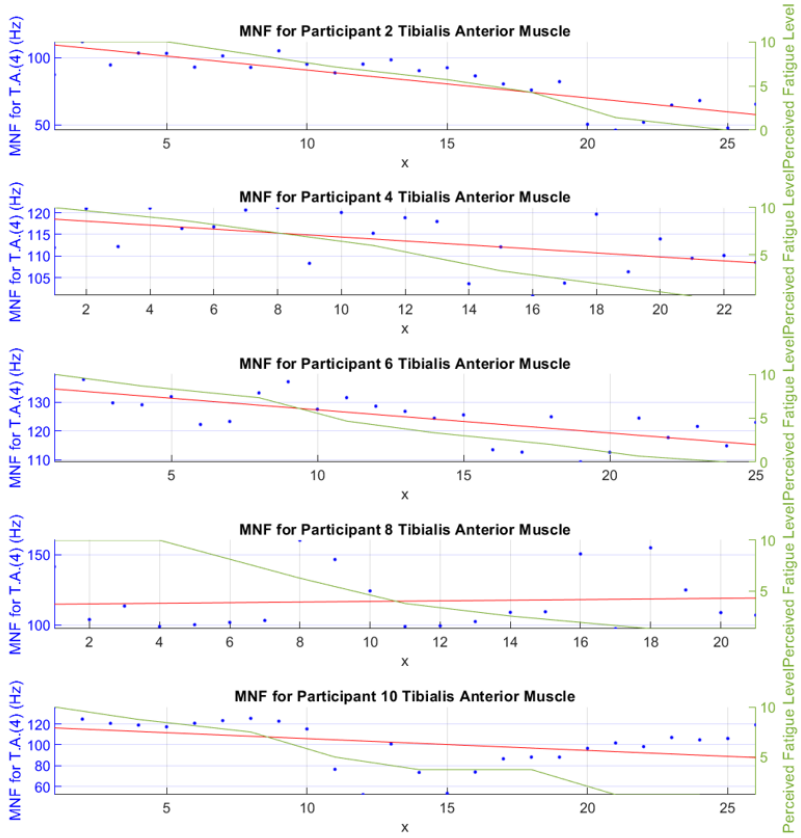

MDF for Participant 1 Biceps Femoris Muscle
MDF for Participant 3 Biceps Femoris Muscle
MDF for Participant 5 Biceps Femoris Muscle
MDF for Participant 7 Biceps Femoris Muscle
MDF for Participant 9 Biceps Femoris Muscle
MDF for B.F.(2) (Hz)
Perceived Fatigue Level
x

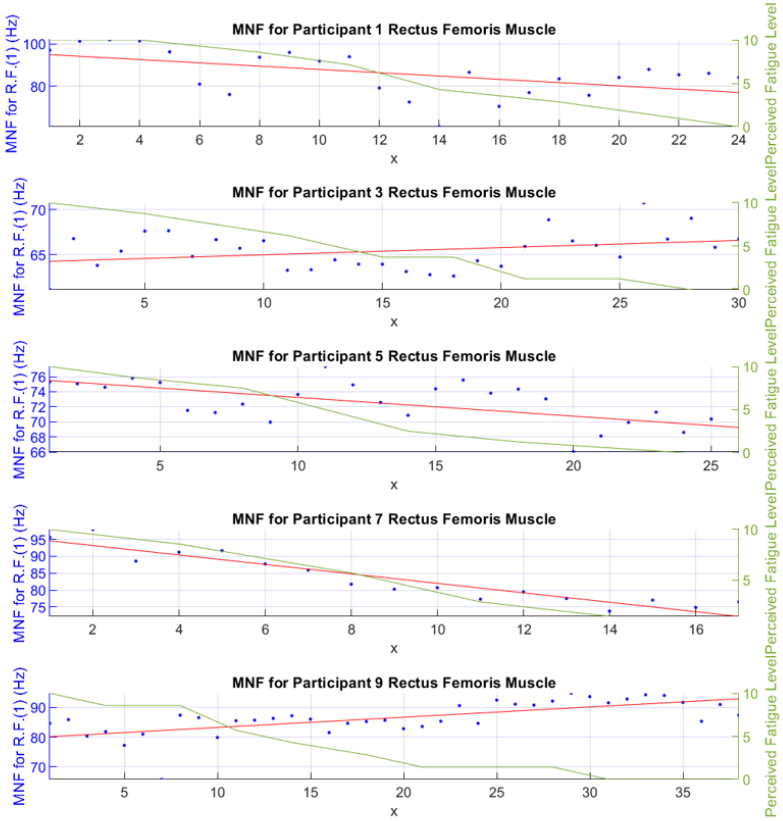

MDF for Participant 2 Biceps Femoris Muscle
MDF for Participant 4 Biceps Femoris Muscle
MDF for Participant 6 Biceps Femoris Muscle
MDF for Participant 8 Biceps Femoris Muscle
MDF for Participant 10 Biceps Femoris Muscle
MDF for B.F.(2) (Hz)
Perceived Fatigue Level
x

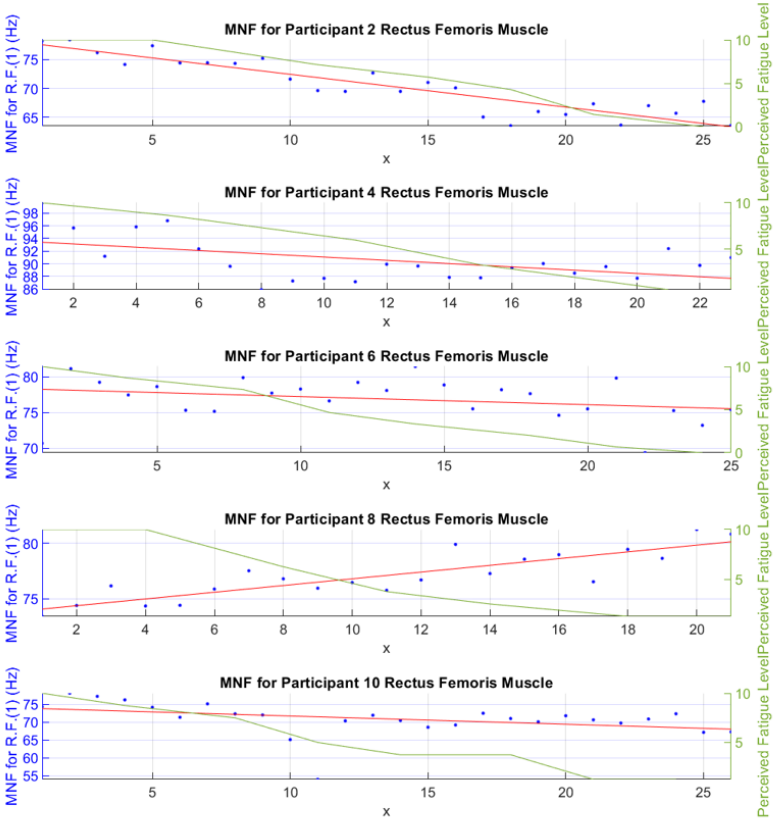

MDF for Participant 1 Gastrocnemius Medialis Muscle
MDF for Participant 3 Gastrocnemius Medialis Muscle
MDF for Participant 5 Gastrocnemius Medialis Muscle
MDF for Participant 7 Gastrocnemius Medialis Muscle
MDF for Participant 9 Gastrocnemius Medialis Muscle
MDF for G.M.(3) (Hz)
Perceived Fatigue Level
x

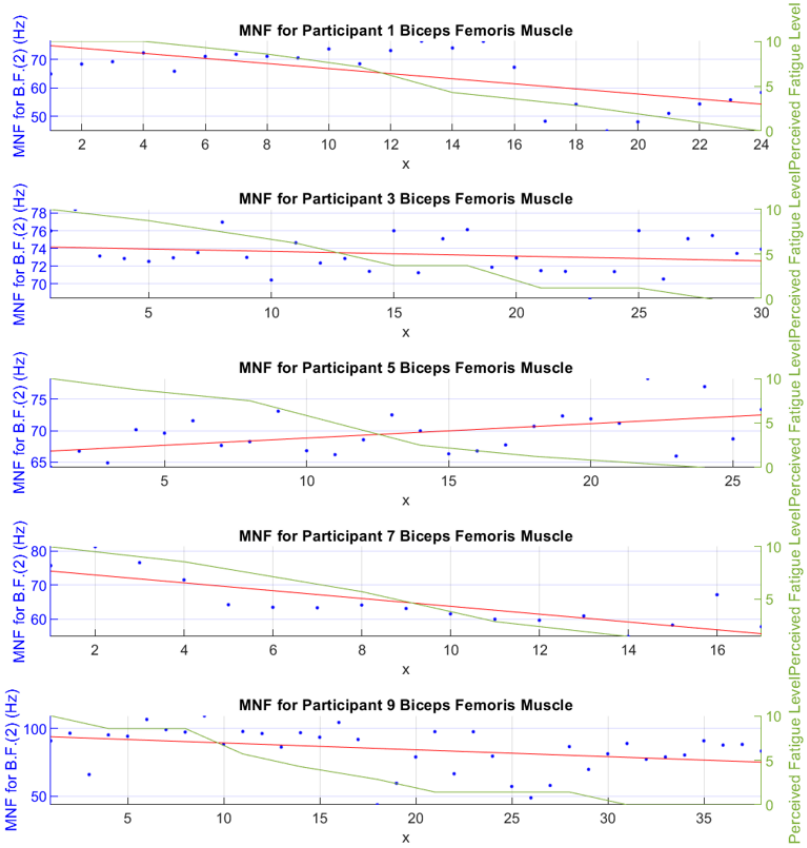

MDF for Participant 2 Gastrocnemius Medialis Muscle
MDF for Participant 4 Gastrocnemius Medialis Muscle
MDF for Participant 6 Gastrocnemius Medialis Muscle
MDF for Participant 8 Gastrocnemius Medialis Muscle
MDF for Participant 10 Gastrocnemius Medialis Muscle
MDF for G.M.(3) (Hz)
Perceived Fatigue Level
x

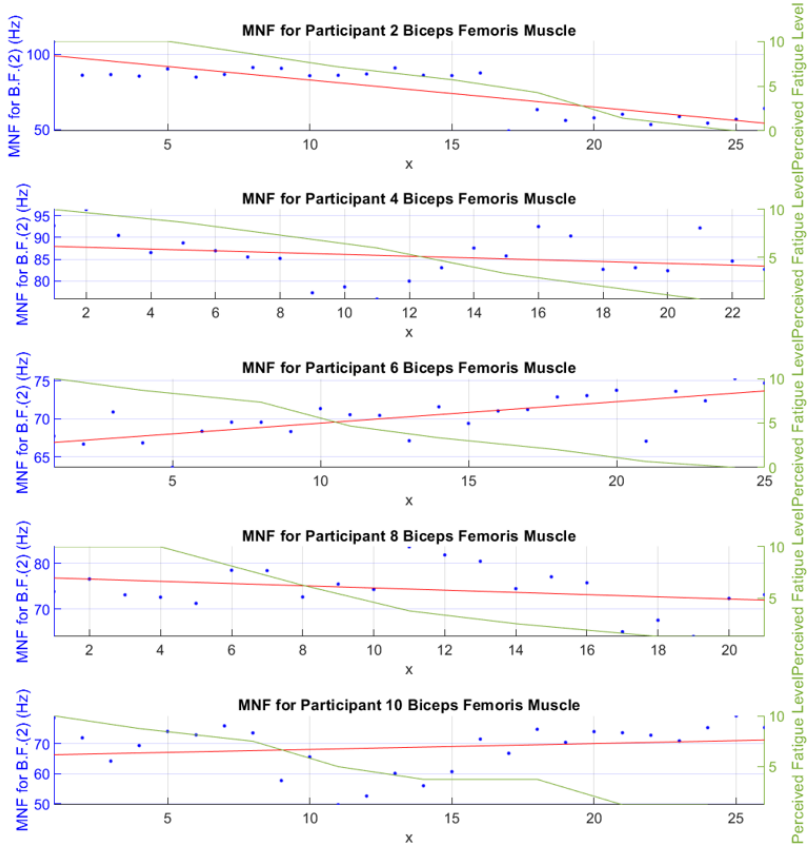

MNF for Participant 1 Tibialis Anterior Muscle
MNF for Participant 3 Tibialis Anterior Muscle
MNF for Participant 5 Tibialis Anterior Muscle
MNF for Participant 7 Tibialis Anterior Muscle
MNF for Participant 9 Tibialis Anterior Muscle
MNF for T.A.(4) (Hz)
Perceived Fatigue LevelPerceived Fatigue LevelPerceived Fatigue LevelPerceived Fatigue Level
x

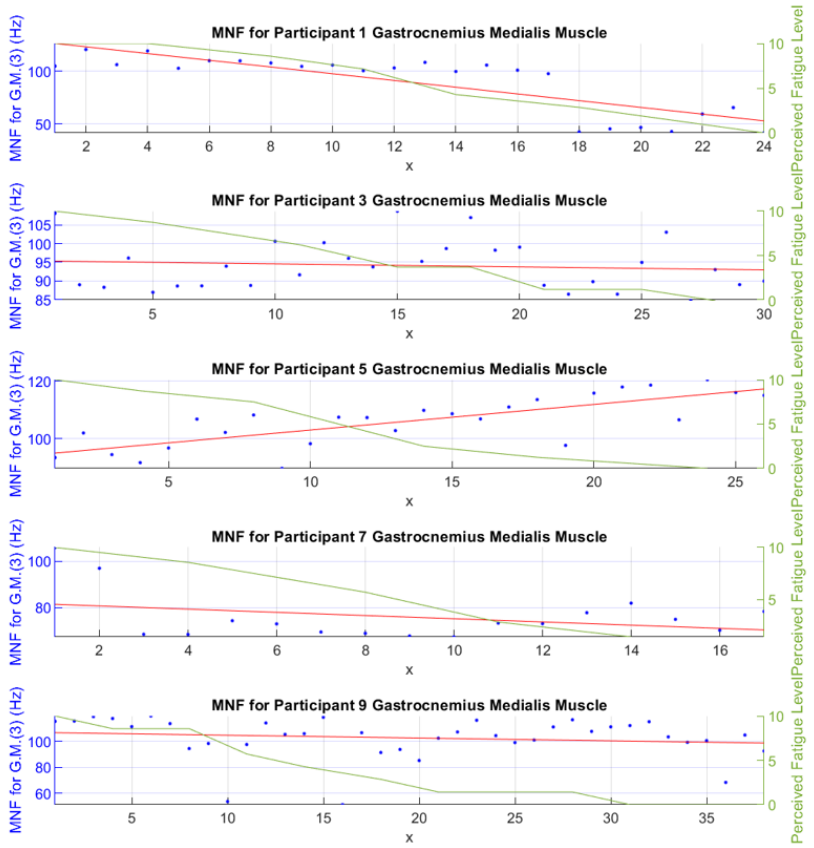

MNF for Participant 2 Tibialis Anterior Muscle
MNF for Participant 4 Tibialis Anterior Muscle
MNF for Participant 6 Tibialis Anterior Muscle
MNF for Participant 8 Tibialis Anterior Muscle
MNF for Participant 10 Tibialis Anterior Muscle
MNF for T.A.(4) (Hz)
Perceived Fatigue LevelPerceived Fatigue LevelPerceived Fatigue LevelPerceived Fatigue Level
x

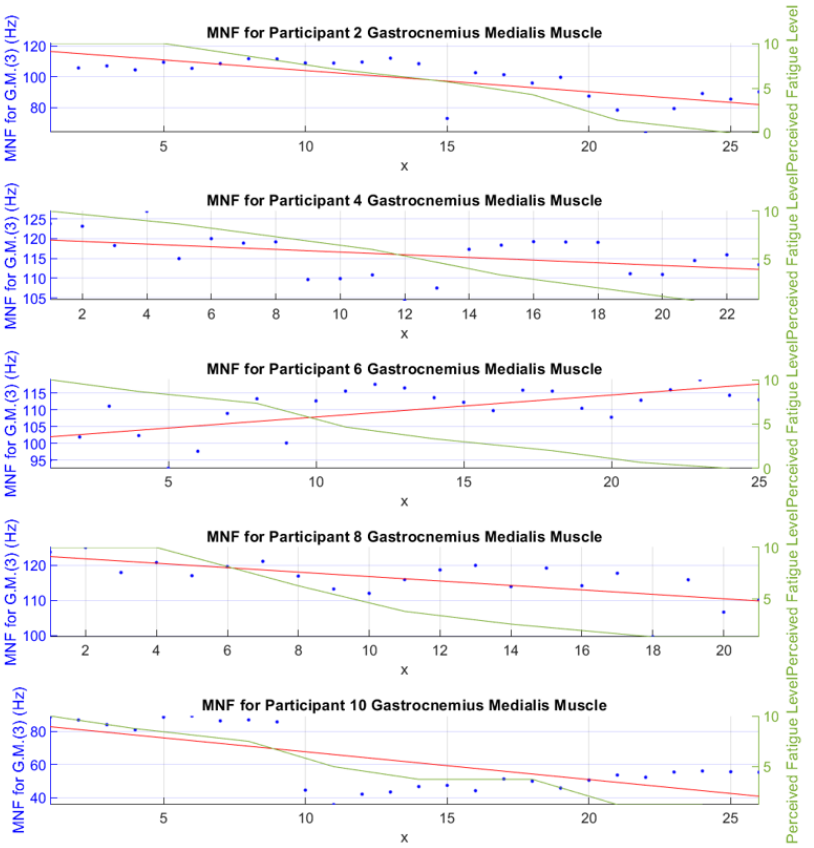

MNF for Participant 1 Rectus Femoris Muscle
MNF for Participant 3 Rectus Femoris Muscle
MNF for Participant 5 Rectus Femoris Muscle
MNF for Participant 7 Rectus Femoris Muscle
MNF for Participant 9 Rectus Femoris Muscle
MNF for R.F.(1) (Hz)
Perceived Fatigue Level
x


MNF for Participant 2 Rectus Femoris Muscle
MNF for Participant 4 Rectus Femoris Muscle
MNF for Participant 6 Rectus Femoris Muscle
MNF for Participant 8 Rectus Femoris Muscle
MNF for Participant 10 Rectus Femoris Muscle
MNF for R.F.(1) (Hz)
Perceived Fatigue Level
x

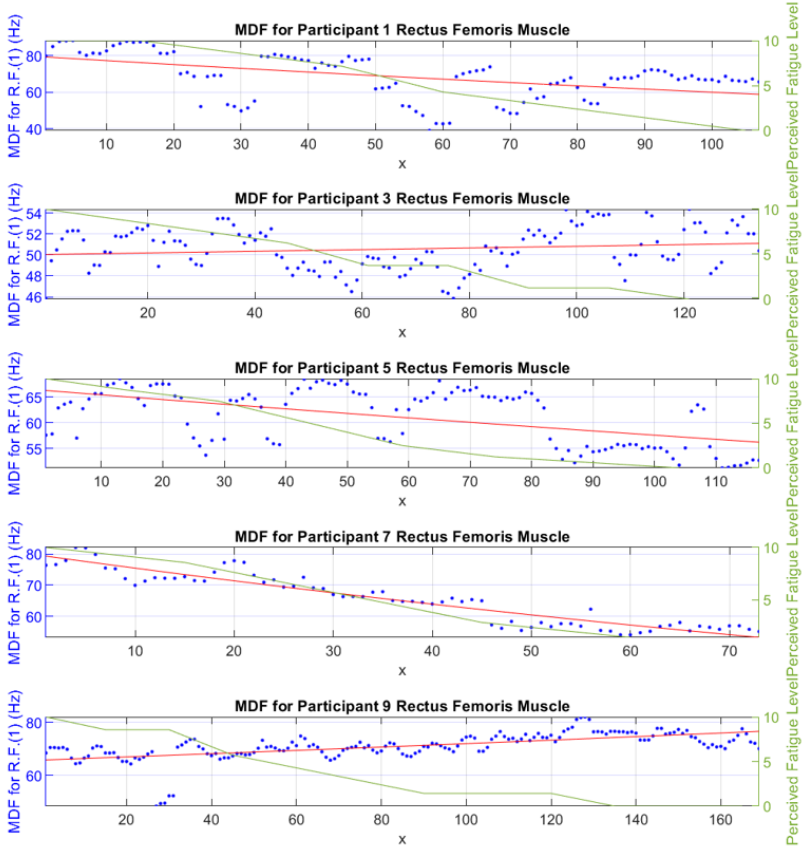

MNF for Participant 1 Biceps Femoris Muscle
MNF for Participant 3 Biceps Femoris Muscle
MNF for Participant 5 Biceps Femoris Muscle
MNF for Participant 7 Biceps Femoris Muscle
MNF for Participant 9 Biceps Femoris Muscle
MNF for B.F.(2) (Hz)
Perceived Fatigue Level
x

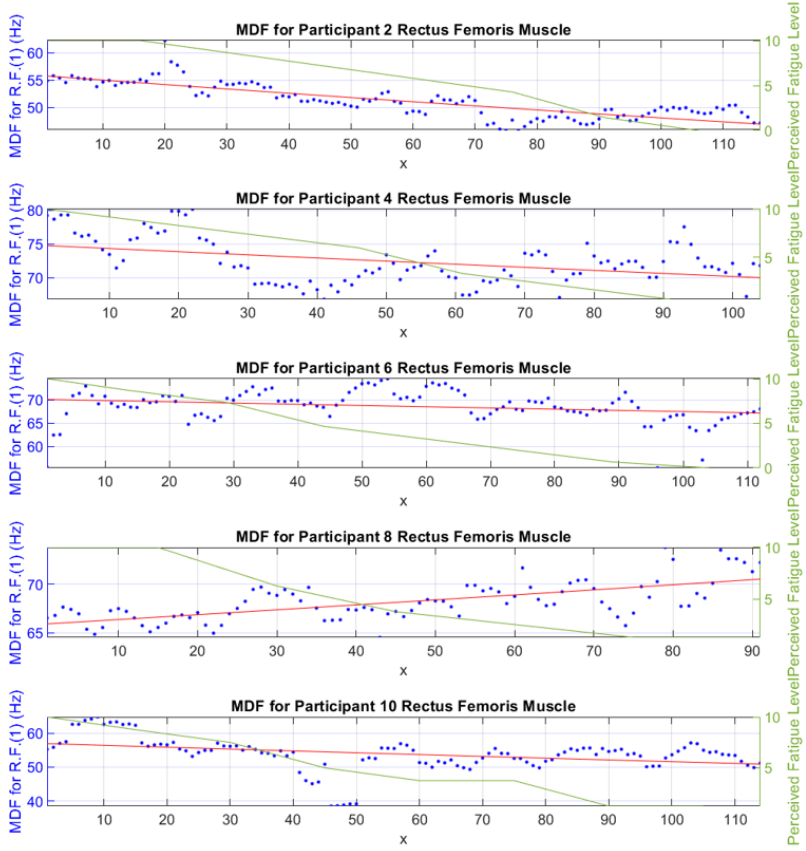

MNF for Participant 2 Biceps Femoris Muscle
MNF for Participant 4 Biceps Femoris Muscle
MNF for Participant 6 Biceps Femoris Muscle
MNF for Participant 8 Biceps Femoris Muscle
MNF for Participant 10 Biceps Femoris Muscle
MNF for B.F.(2) (Hz)
Perceived Fatigue Level
x

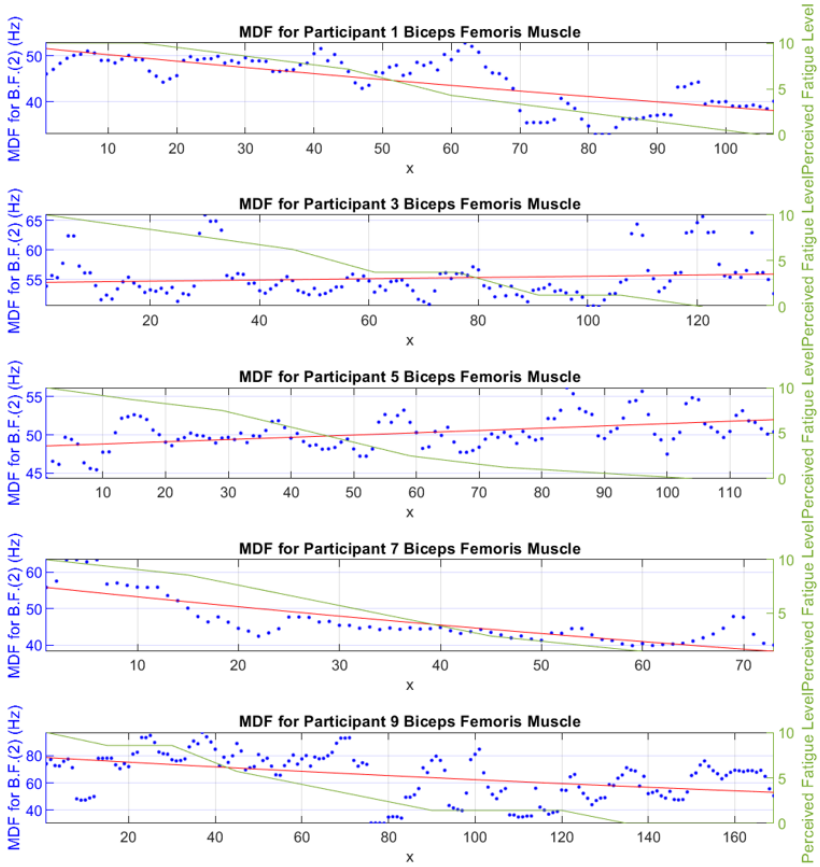

MNF for Participant 1 Gastrocnemius Medialis Muscle
MNF for Participant 3 Gastrocnemius Medialis Muscle
MNF for Participant 5 Gastrocnemius Medialis Muscle
MNF for Participant 7 Gastrocnemius Medialis Muscle
MNF for Participant 9 Gastrocnemius Medialis Muscle
MNF for G.M.(3) (Hz)
Perceived Fatigue LevelPerceived Fatigue LevelPerceived Fatigue LevelPerceived Fatigue Level
x

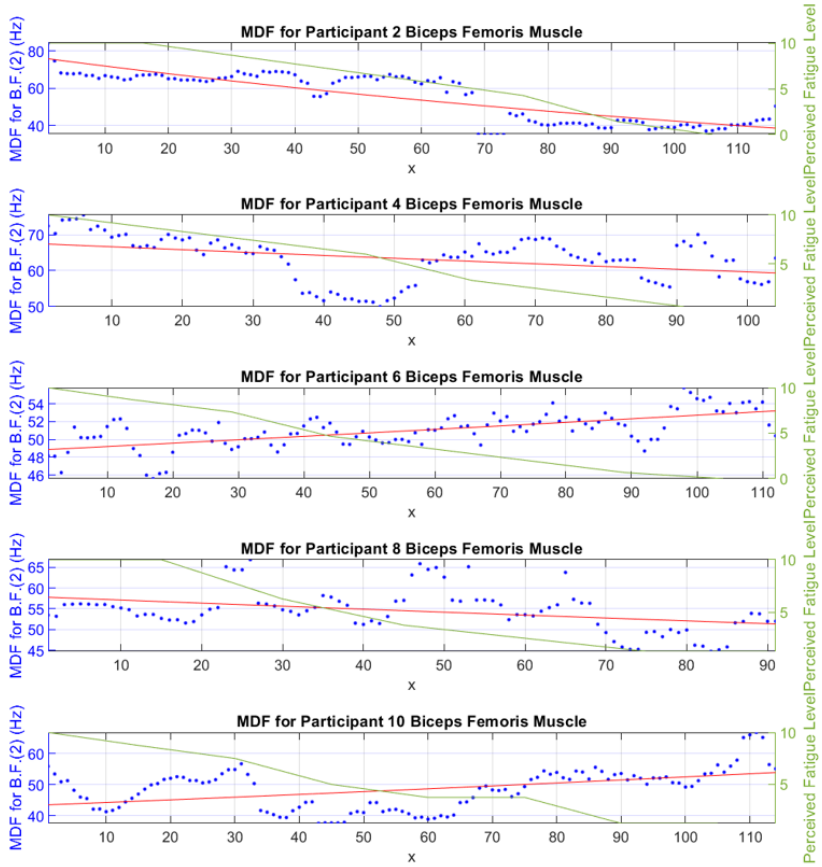

MNF for Participant 2 Gastrocnemius Medialis Muscle
MNF for Participant 4 Gastrocnemius Medialis Muscle
MNF for Participant 6 Gastrocnemius Medialis Muscle
MNF for Participant 8 Gastrocnemius Medialis Muscle
MNF for Participant 10 Gastrocnemius Medialis Muscle
MNF for G.M.(3) (Hz)
Perceived Fatigue LevelPerceived Fatigue LevelPerceived Fatigue LevelPerceived Fatigue Level
x

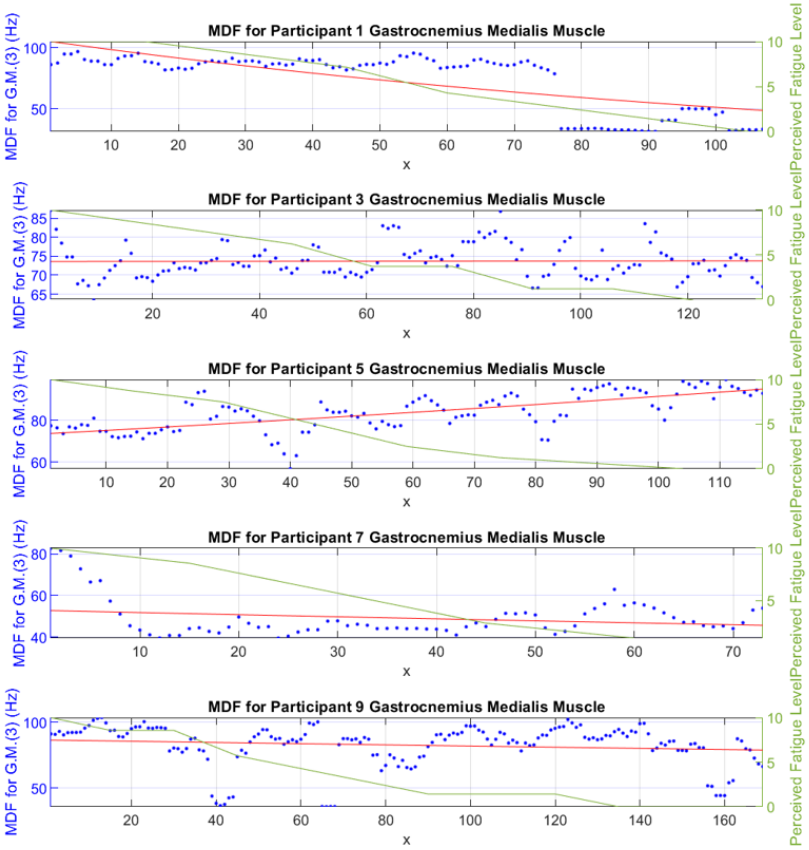
F.2 Moving Window Analysis

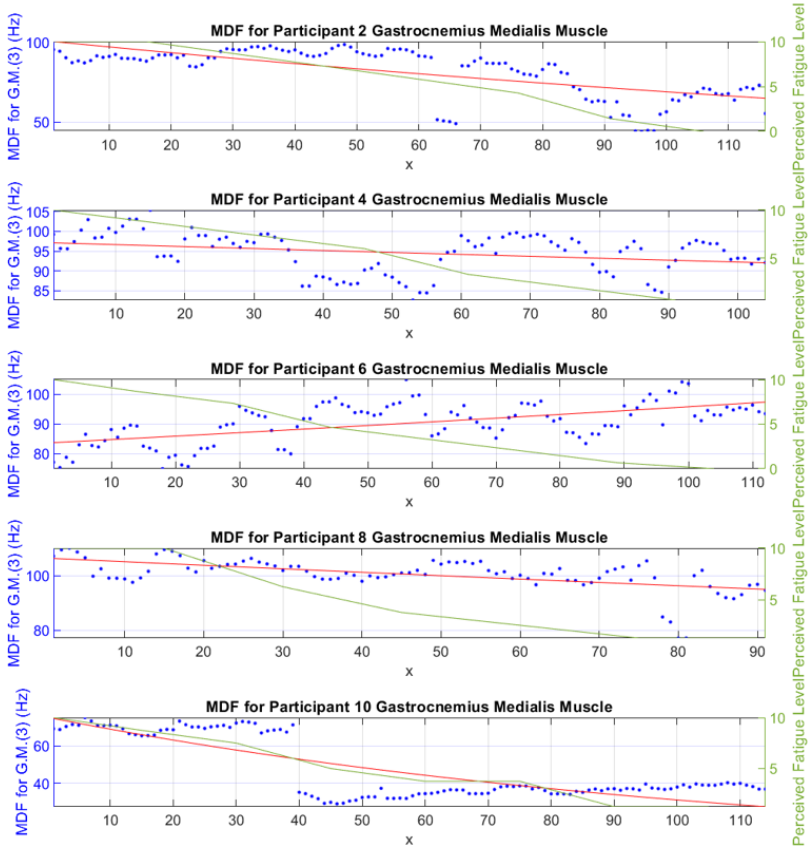


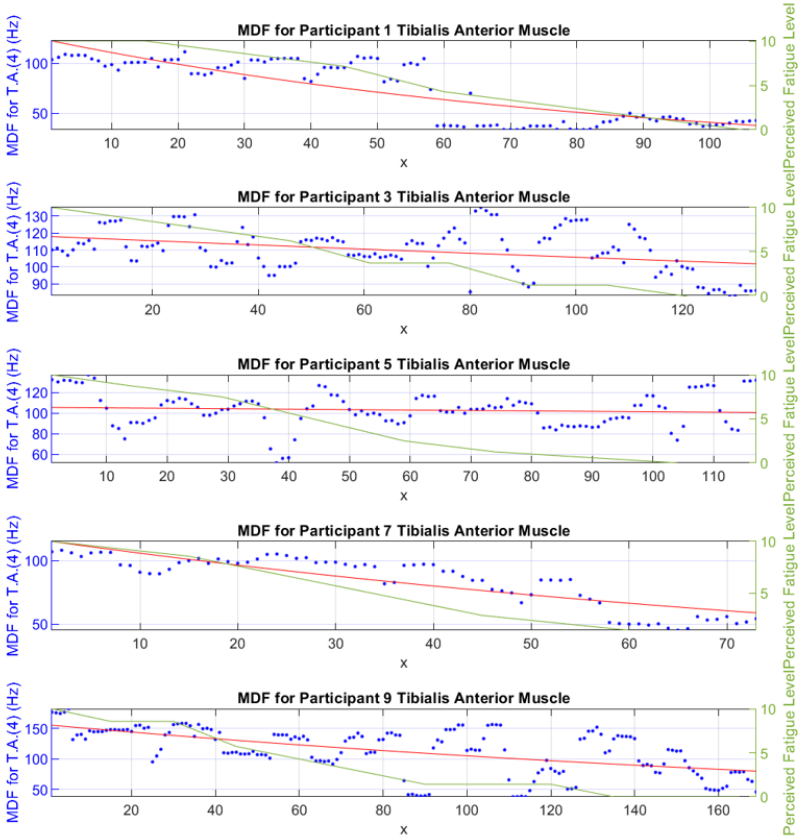


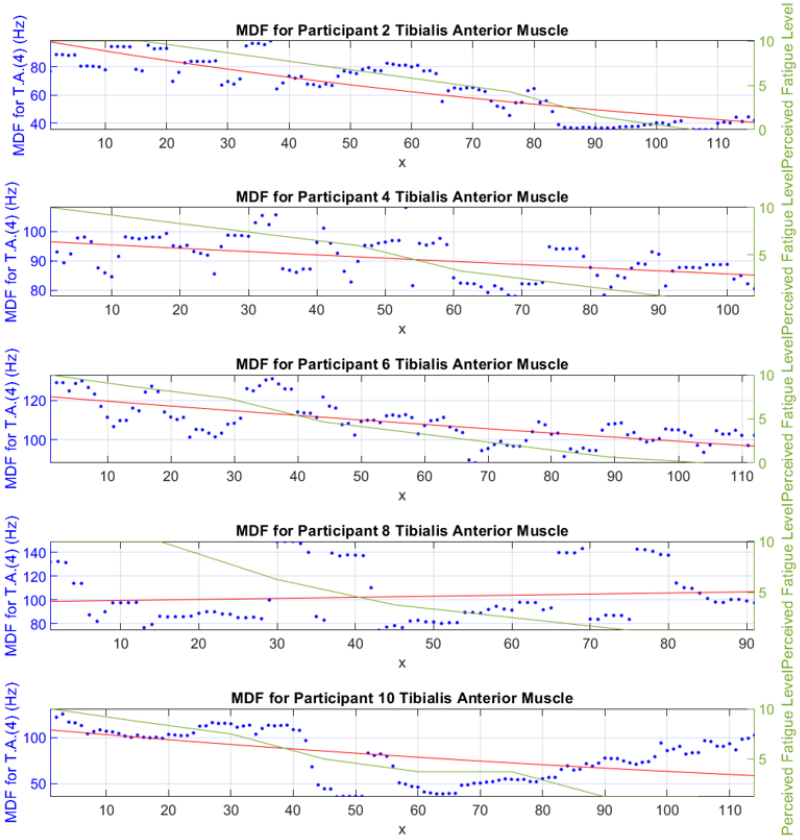

MDF for Participant 1 Biceps Femoris Muscle
MDF for Participant 3 Biceps Femoris Muscle
MDF for Participant 5 Biceps Femoris Muscle
MDF for Participant 7 Biceps Femoris Muscle
MDF for Participant 9 Biceps Femoris Muscle
MDF for B.F.(2) (Hz)
Perceived Fatigue Level
x

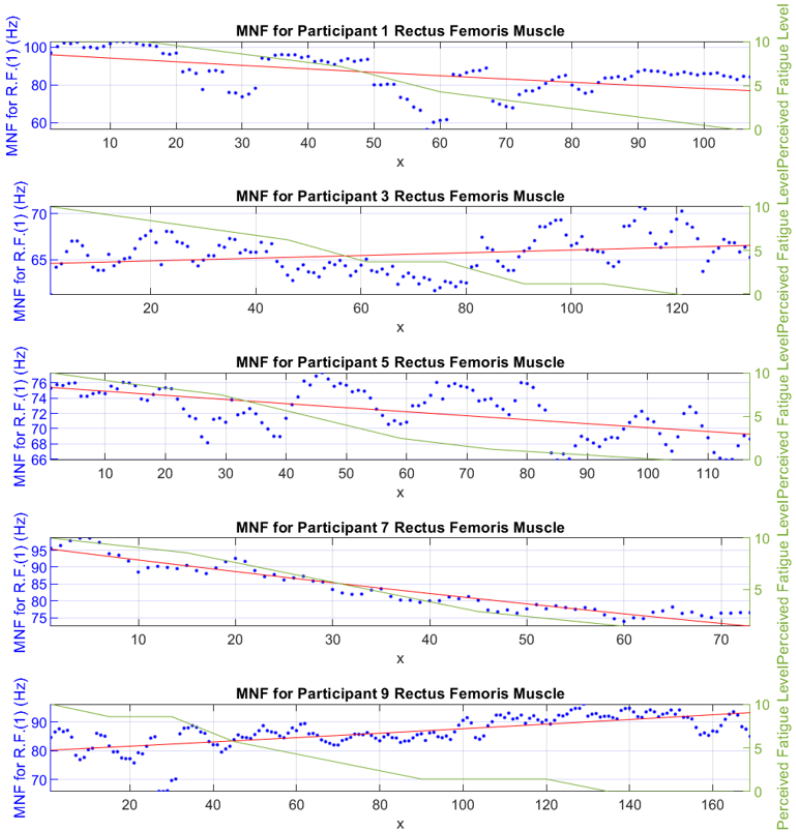

MDF for Participant 2 Biceps Femoris Muscle
MDF for Participant 4 Biceps Femoris Muscle
MDF for Participant 6 Biceps Femoris Muscle
MDF for Participant 8 Biceps Femoris Muscle
MDF for Participant 10 Biceps Femoris Muscle
MDF for B.F.(2) (Hz)
Perceived Fatigue Level
x

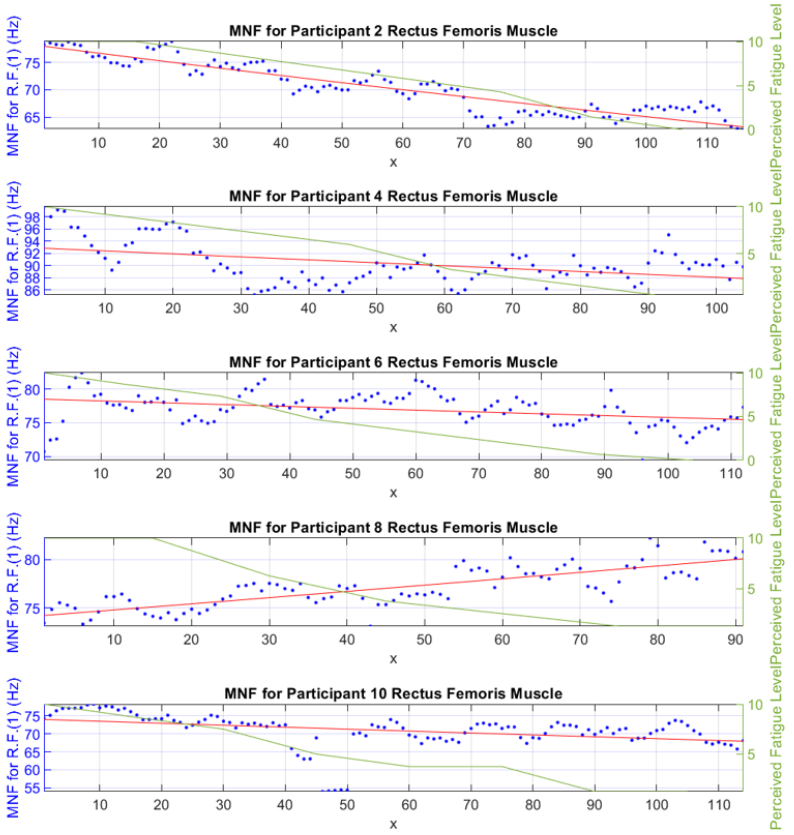

MDF for Participant 1 Gastrocnemius Medialis Muscle
MDF for Participant 3 Gastrocnemius Medialis Muscle
MDF for Participant 5 Gastrocnemius Medialis Muscle
MDF for Participant 7 Gastrocnemius Medialis Muscle
MDF for Participant 9 Gastrocnemius Medialis Muscle
MDF for G.M.(3) (Hz)
Perceived Fatigue Level
x

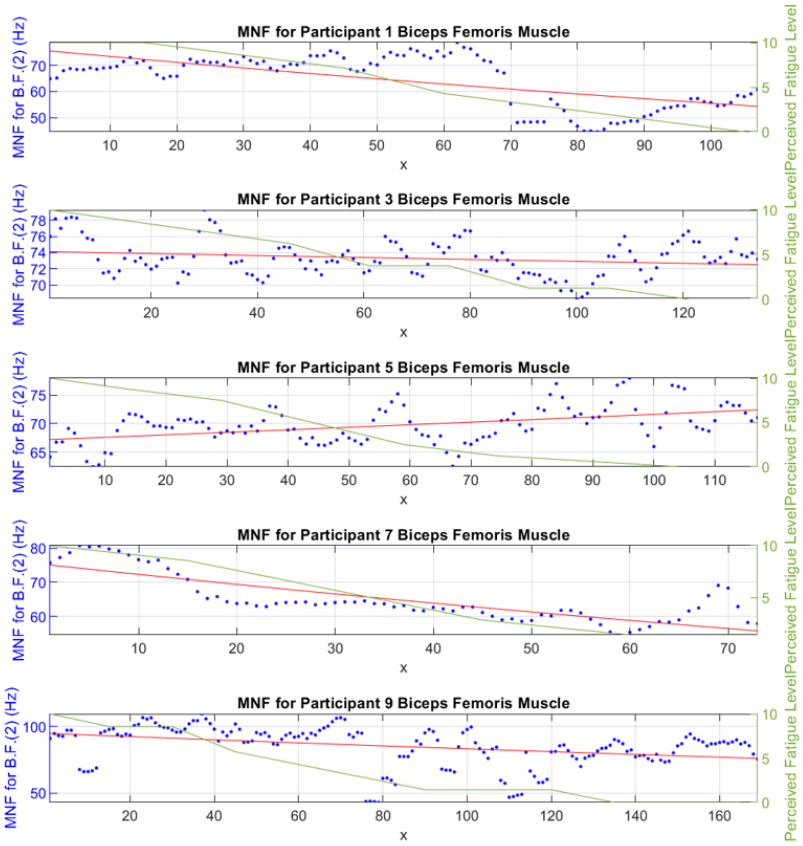

MDF for Participant 2 Gastrocnemius Medialis Muscle
MDF for Participant 4 Gastrocnemius Medialis Muscle
MDF for Participant 6 Gastrocnemius Medialis Muscle
MDF for Participant 8 Gastrocnemius Medialis Muscle
MDF for Participant 10 Gastrocnemius Medialis Muscle
MDF for G.M.(3) (Hz)
Perceived Fatigue Level
x

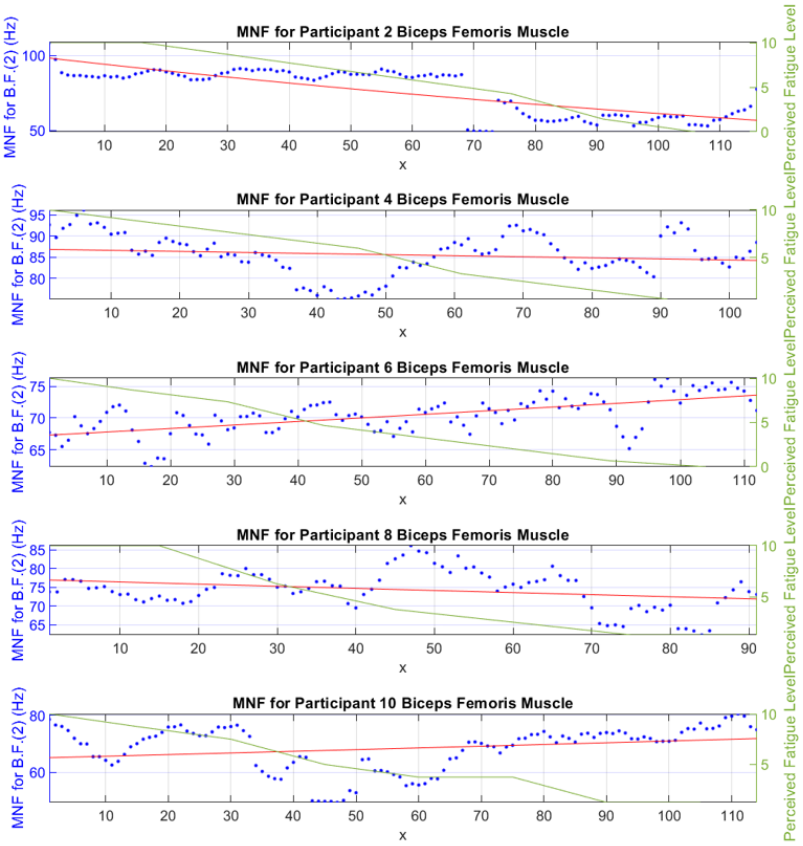

MDF for Participant 1 Tibialis Anterior Muscle
MDF for Participant 3 Tibialis Anterior Muscle
MDF for Participant 5 Tibialis Anterior Muscle
MDF for Participant 7 Tibialis Anterior Muscle
MDF for Participant 9 Tibialis Anterior Muscle
MDF for T.A.(4) (Hz)
Perceived Fatigue LevelPerceived Fatigue LevelPerceived Fatigue LevelPerceived Fatigue Level
x

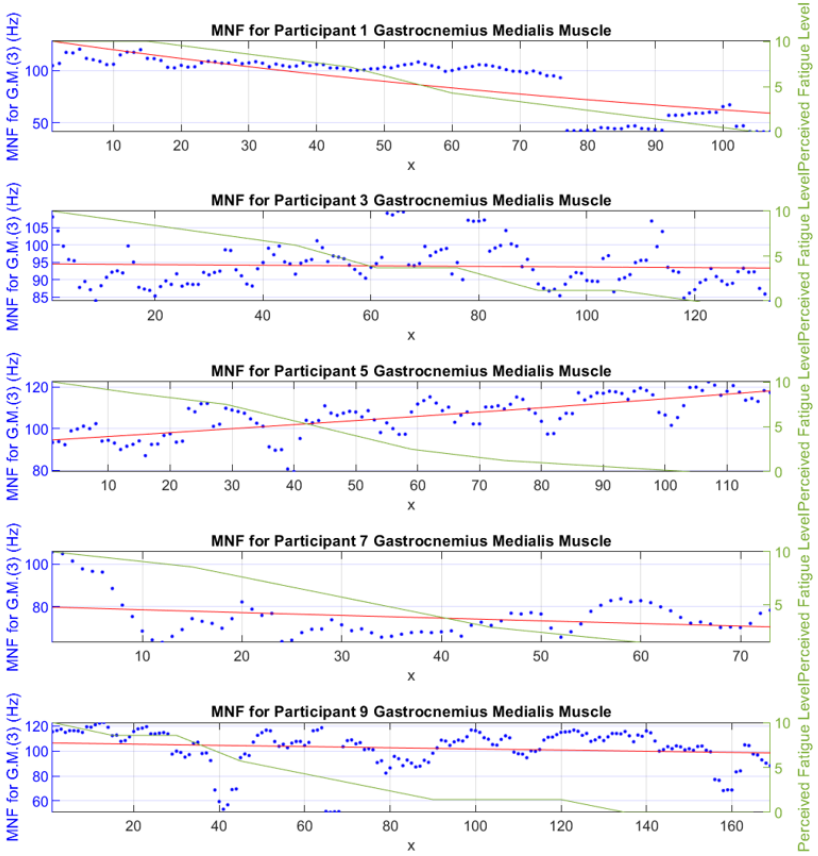

MDF for Participant 2 Tibialis Anterior Muscle
MDF for Participant 4 Tibialis Anterior Muscle
MDF for Participant 6 Tibialis Anterior Muscle
MDF for Participant 8 Tibialis Anterior Muscle
MDF for Participant 10 Tibialis Anterior Muscle
MDF for T.A.(4) (Hz)
Perceived Fatigue LevelPerceived Fatigue LevelPerceived Fatigue LevelPerceived Fatigue Level
x

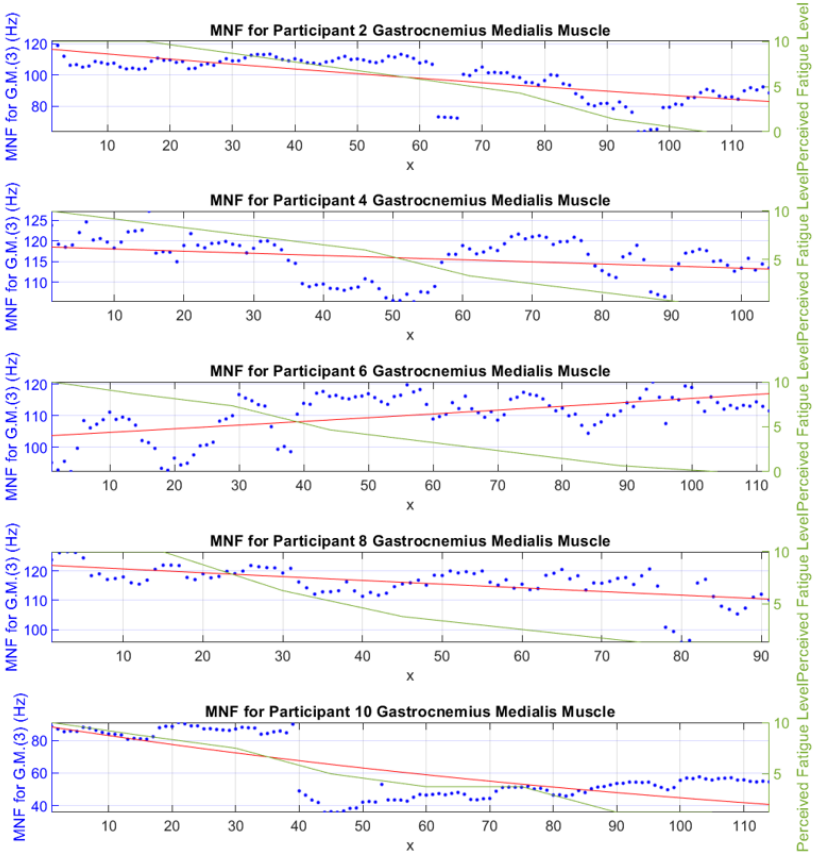

MNF for Participant 1 Rectus Femoris Muscle
MNF for Participant 3 Rectus Femoris Muscle
MNF for Participant 5 Rectus Femoris Muscle
MNF for Participant 7 Rectus Femoris Muscle
MNF for Participant 9 Rectus Femoris Muscle
MNF for R.F.(1) (Hz)
Perceived Fatigue Level
x


MNF for Participant 2 Rectus Femoris Muscle
MNF for Participant 4 Rectus Femoris Muscle
MNF for Participant 6 Rectus Femoris Muscle
MNF for Participant 8 Rectus Femoris Muscle
MNF for Participant 10 Rectus Femoris Muscle
MNF for R.F.(1) (Hz)
Perceived Fatigue Level
x


MNF for Participant 1 Biceps Femoris Muscle
MNF for Participant 3 Biceps Femoris Muscle
MNF for Participant 5 Biceps Femoris Muscle
MNF for Participant 7 Biceps Femoris Muscle
MNF for Participant 9 Biceps Femoris Muscle
MNF for B.F.(2) (Hz)
Perceived Fatigue Level
x


MNF for Participant 2 Biceps Femoris Muscle
MNF for Participant 4 Biceps Femoris Muscle
MNF for Participant 6 Biceps Femoris Muscle
MNF for Participant 8 Biceps Femoris Muscle
MNF for Participant 10 Biceps Femoris Muscle
MNF for B.F.(2) (Hz)
Perceived Fatigue Level
x


MNF for Participant 1 Gastrocnemius Medialis Muscle
MNF for Participant 3 Gastrocnemius Medialis Muscle
MNF for Participant 5 Gastrocnemius Medialis Muscle
MNF for Participant 7 Gastrocnemius Medialis Muscle
MNF for Participant 9 Gastrocnemius Medialis Muscle
MNF for G.M.(3) (Hz)
Perceived Fatigue Level
x


MNF for Participant 2 Gastrocnemius Medialis Muscle
MNF for Participant 4 Gastrocnemius Medialis Muscle
MNF for Participant 6 Gastrocnemius Medialis Muscle
MNF for Participant 8 Gastrocnemius Medialis Muscle
MNF for Participant 10 Gastrocnemius Medialis Muscle
MNF for G.M.(3) (Hz)
Perceived Fatigue Level
x