Stine Olsen

Exploring possibilities on predicting subjective evaluation of athlete's feelings after workout sessions using deep learning.

NTNU
Norwegian University of
Science and Technology

Stine Olsen

Exploring possibilities on predicting subjective evaluation of athlete's feelings after workout sessions using deep learning.

Master's thesis in Electronic Systems Design
Supervisor: Ilangko Balasingham
Co-supervisor: Kimmo Kansanen, Naimahmed Nesaragi, Fredrik Mentzoni
February 2024

Norwegian University of Science and Technology

NTNU

Norwegian University of
Science and Technology

NTNU – Trondheim
Norwegian University of
Science and Technology

Exploring possibilities on predicting subjective evaluation of athlete's feeling after workout sessions using deep learning

Stine Olsen

January 2024

MASTER THESIS

Department of Electrical Engineering

Norwegian University of Science and Technology

Supervisor: Ilangko Balasingham

Co-Supervisor: Kimmo Kansanen

Co-Supervisor: Naimahmed Nesaragi

External Supervisor: Olympiatoppen, Fredrik Mentzoni

Preface

This master's thesis completes a Master of Technology at the Norwegian University of Science and Technology (NTNU) under the Department of Electrical Engineering in the autumn of 2023. The project was proposed by Olympiatoppen (OLT) and was completed in collaboration with supervisors Ilangko Balasingham and Kimmo Kansanen at NTNU. The work presented in this thesis continues the work done in the specialization project [1], which was completed in the spring of 2023.

Parts of the second chapter are updated and extended versions of the work done in the specialization project [1]. The authors implement the software using Python libraries, including numpy, pandas, scipy, pyhrv, hrvanalysis, and more. An anonymous rower recorded the data set during the summer of 2023.

The code implemented can be shared by sending mail to olsen.stine@outlook.com or viewed at the GitHub repository.

This thesis's main contribution is exploring possibility on prediction subjective score of athlete after workout using deep learning. Additionally, exploring prediction using transfer learning from SWELL-KW datasetand training data provided by OLT.

I had little prior knowledge and experience about this topic, but I have learned a lot during this project, particularly regarding various machine learning models and time series data related to how the heart behaves and the signal it sends. This thesis hopefully reflects the amount of knowledge gained over the past year.

Trondheim, February 2024
Stine Olsen

ii

Acknowledgments

Abstract

Wearables, such as training watches, have become increasingly popular in recent years. These devices can provide valuable information about our health and fitness by measuring parameters such as Heart Rate Variability (HRV).

The goal of this master's thesis was to investigate methods and the possibility of predicting an athlete's subjective assessment after a workout session. The motivation behind this was to prevent fatigue and ensure athletes the best possible performance in competitions. To determine the subjective predictions, two athletes logged all their training sessions and provided their subjective scores for each session. With a dataset of 290 sessions, which contains subjective scores and HRV, various machine learning models were trained.

Through the use of transfer learning, we show that predicting subjective scores is possible given a more balanced and larger data set, but with the data provided in this thesis, it was not feasible to achieve the desired accuracy.

As this thesis proves the feasibility of predicting subjective scores given a more balanced dataset, Olympiatoppen can expand with more athletes and sessions to achieve a better model that can be used in their work.

Sammendrag

Wearables har blitt stadig mer populært de siste årene. Denne typen enherter, som for eksempel treningsklokker kan gi verdifull informasjon om vår egen helse og treningsvaner ved å måle parametere som hjertefrekvensvariabilitet (HRV).

Målet med denne masteroppgaven var å undersøke metoder og muligheten for å forutsi en idrettsutøvers subjektive vurdering etter en treningsøkt. Motivasjonen bak dette var å forhindre utmattelse og sikre utøverne best mulig prestasjoner i konkurranser. For å bestemme de subjektive karakterene, loggførte to idrettsutøvere alle treningsøktene sine og ga deres subjektive karakterer for hver økt. Med et datasett på 290 økter, som inneholder subjektive karakterer og HRV, ble ulike maskinlæringsmodeller trent.

Gjennom bruk av «overføringslæring» viser vi at prediksjon av subjektive karakterer er mulig gitt et mer balansert og større datasett. Med dataene gitt i denne oppgaven var det ikke mulig å oppnå ønsket nøyaktighet.

Ettersom denne oppgaven beviser muligheten for å forutsi subjektive karakterer gitt et mer balansert datasett, kan Olympiatoppen utvide med flere idrettsutøvere og økter for å oppnå en bedre modell som kan brukes i deres arbeid.

Contents

List of Figures

List of Tables

Acronyms

ANS Autonomic Nervous System

CNN Convolutional Neural Network

CNS Central Nervous System

DFA Detrended Fluctuation Analysis

ECG Electrocardiogram

HRV Heart Rate Variability

IBI Interbeat Interval

LSTM Long-Short-Term Memory

ML Machine Learning

NN Neural Network

NTNU Norwegian University of Science and Technology

OLT Olympiatoppen

RBF Radial Basis Function

RHR Resting Heart Rate

RNN Recurent Neural Network

RPE Rate of Perceived Extertion

SVM Support Vector Machine

Swell-KW Swell Knowledge Work

Chapter 1

Introduction

The number of articles relating to Heart Rate Variability has increased drastically during the last few years and is a topic with increasing interest. And even though it has become more popular in recent years, it is not a new concept. Chinese physician Wang Shu-He (180—270 AD) was the first to observe that highly variable heart rate was a sign of good health, and stated that "'If the heart beats as regularly as the knocking of a woodpecker or the falling of rain drops from a roof then the patient will be dead within 4 days". This was before measurement instruments used today were invented, and the physicians had to listen to the heart without any instruments to help. Today, this process is easier with the help of instruments such as stethoscopes or Electrocardiogram (ECG)s, in addition to the many types of wearables that exist in today's market. Wang Shu-He discovered [2] a great indicator of cardiac health, which physicians are still investigating centuries after his discovery.

Smartwatches have become widely popular in recent years; in 2016, 102 million devices were shipped worldwide, compared to 2022, where 173 million were shipped [3], [4]. The increase in demand is connected to its convenience, as now you can easily connect the watch to your smartphone and other smart devices. The wrist wearable includes various sensors to measure heart rate, temperature, and movements. Optical heart rate sensors and movement-related sensors such as accelerometers, GPS, gyroscopes, and compasses are the most common sensors in smartwatches. Together, these can be used to learn more about health and for fitness tracking [5], [6]. Smartwatches allowed users to monitor their physical activity, track steps, and measure heart rate, in addition to functionalities such as communication, notifications, mobile apps, and more [6].

1.1 Motivation

This work aims to explore the possibility of predicting an athlete's subjective feeling after a workout based on Heart Rate Variability (HRV) data from a wearable. We want to look further into parameters and features that could help prevent fatigue for an athlete and could be used to help an athlete perform as well as possible at competitions. All this is connected to the athlete's readiness to train.

In sports science, athlete performance optimization, the ability to accurately gauge an athlete's readiness to train has emerged as a critical factor. Methods used until now have been effective but are often limited in their precision. This work will also look into a parameter that has become more and more common called Detrended Fluctuation Analysis (DFA) Alpha 1 and its potential to be used as a reliable indicator of an athlete's readiness to train.

1.2 Problem Description

This work explores the parameter Detrended Fluctuation Analysis (DFA) Alpha 1 and its possibility of helping determine an athlete's readiness to train. The second objective is to train different Neural Networks using the provided data sets and compare their performance. We wish to be able to predict the subjective feelings of an athlete after training. The performance of the methods will be evaluated on the possibility of being used in a future application for Olympiatoppen.

1.3 Approach

The approach involves calculating DFA Alpha 1 and analyzing its correlation to subjective and objective parameters, focusing on the subjective parameters: feeling and Rate of Perceived Extertion (RPE). Continued in this thesis the subjective feeling will be termed subjective score. We also aim to predict the subjective score, using different machine learning techniques. These techniques will be compared regarding their ability to predict the correct subjective score. The performance of the neural networks and used data sets will be compared to evaluate the effectiveness of each technique.

1.4 Outline

The introduction of the general problem and the objectives of the work is done in chapter 1. Chapter 2 provides relevant background information on the heart and how it behaves during

exercise, ways of measuring the heart's activity, and different data processing techniques. Lastly, the chapter describes various principles and algorithms regarding machine learning and artificial intelligence.

Chapter 3 presents a description of the implementations used in this work and a description of the datasets used. Additionally, the chapter 3 describes the features used in transfer learning and a brief overview of the evaluation criteria for the models.

Following this, the results of each model are presented in chapter 4. These results are discussed in chapter 5 and recommendations for future work.

Chapter 2

Background

The objective of this chapter is to present the necessary background information about the heart and how it is affected by exercise. Parameters used for measuring an athlete's fitness and exercise will be presented, in addition to a brief introduction to Detrended Fluctuation Analysis (DFA) and time series data. Finally, this chapter will present principles for prepossessing data and explain different types of machine learning, which will be discussed later in the thesis.

2.1 Anatomy of the heart

The heart is an organ in the human body that is responsible for pumping blood throughout the body. The heart is located in the front of the chest, slightly left of the center, behind the breastbone. It consists of four chambers, two on the top and two on the bottom. The two chambers on top are called the left and right atrium. The right atrium receives deoxygenated blood from the systemic veins, and the left receives oxygenated blood from the pulmonary veins. The two lower chambers are the left and right ventricles. These are thick-walled as they pump blood out of the heart forcefully. The heart chambers and flow are illustrated in Figure 2.1 [7].

The heart can be divided into two pumps, the left and the right, each consisting of one atrium and one ventricle. Both pumps need to work in rhythm [7].

Blood flows from the right atrium to the right ventricle and is then pumped to the lungs to receive oxygen. From the lungs, the blood flows to the left atrium and then to the left ventricle. From there, it is pumped into the systemic circulation [8], [9].
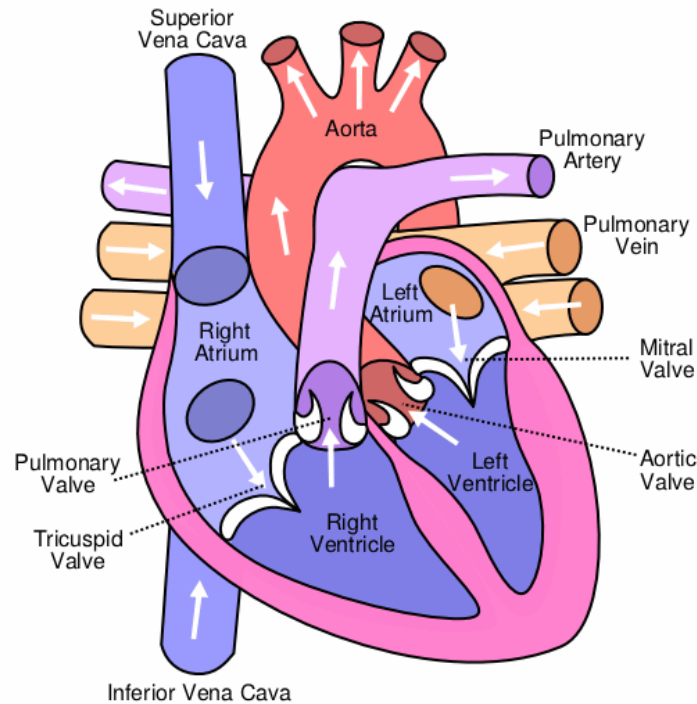
Figure 2.1: Diagram of the heart, visualizing the heart chambers. Figure retrieved from Wikimedia user Wapcaplet, used under the CC BY-SA 3.0 license

The heart cycle consists of two phases: diastole and systole. Systole is the heart's contraction, and diastole is the relaxation of the heart. The heart cycle has three main phases [10].

In the first phase, both atria and ventricles are in diastole; the chambers are relaxed and filled with blood. In the second phase, the atria are in systole. The atria contracts, and blood is pushed into the ventricles. Finally, in the last phase, the ventricle is in systole. The ventricles contract, the right ventricle pushes blood out of the heart through the pulmonary artery, and the left ventricle pushes oxygenated blood to the body through the aorta [10].

2.2   The Nervous system

The nervous system is divided into two large systems, the Central Nervous System (CNS) and the Autonomic Nervous System (ANS). CNS is made up of the brain and spinal cord. The ANS is made up of the nerves that branch off the spinal cord and go through all parts of your body.

The ANS is divided into two parts: the parasympathetic and sympathetic nervous systems. The parasympathetic nervous system is responsible for restoring the body to normal by repairing, resting, and adjusting the body. This is done by slowing down heart rate, decreasing blood pressure, and dilating blood vessels [11].

The other system, the sympathetic nervous system, is responsible for the opposite of the parasympathetic. The sympathetic system is responsible for increasing heart rate and blood pressure, diverting blood flow to muscles. Additionally, the sympathetic system releases sugar into the bloodstream to fuel the muscles efficiently **paharia_what_2010** , [12].

We can say that the sympathetic nervous system is responsible for the body's "fight or flight" response, and the parasympathetic nervous system is responsible for resting and repairing.

Both systems must work together to maintain the internal balance, especially regarding physical, psychological, and environmental stress. The parasympathetic system has the largest effect on HRV, as it helps us recover from exertion, for example, exercise [12].
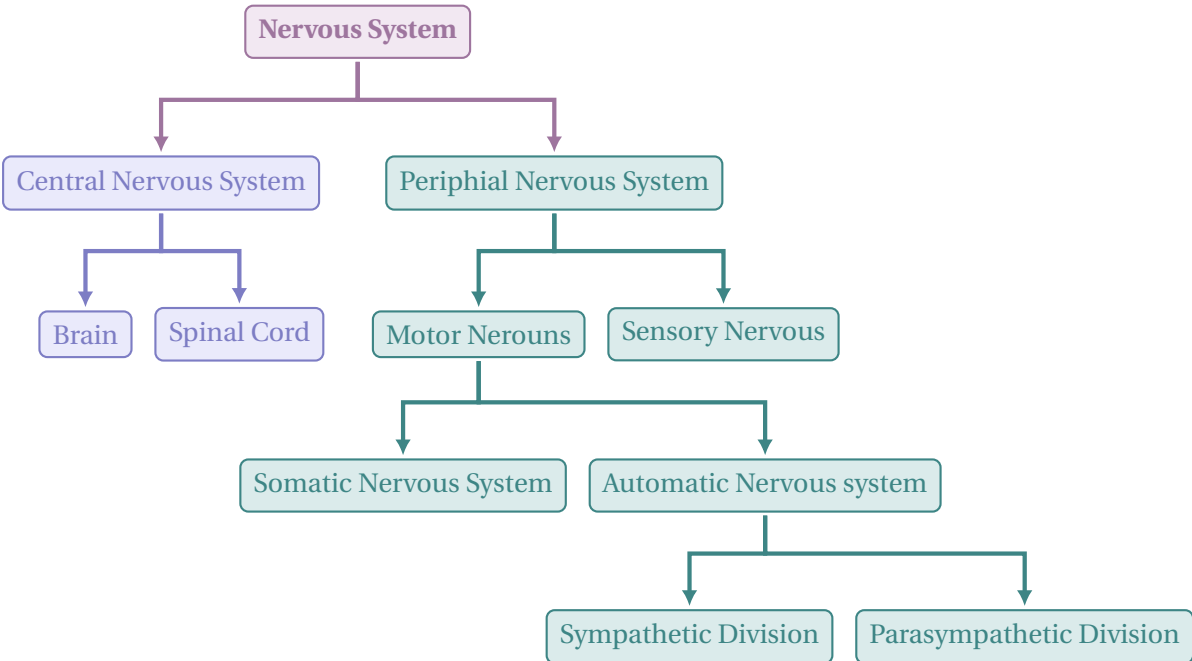
2.3   Electrochardiography

An electrocardiogram (ECG) is a method of recording the electrical signals in the heart. It's a non-invasive test used to quickly monitor heart health and detect problems relating to the heart. While commonly used in a medical sense, more personal devices, such as smart-watches, have begun to provide ECG measurements.

ECG measurements give waveforms that can describe the workings of the heart. The ECG waveform is typically divided into the P wave, the QRS complex, and the T wave [13]. The P wave is the first wave that appears on the ECG, and it represents the electrical activity that

occurs as the atria contract to pump blood into the ventricles. QRS complex is the largest and most prominent wave on the ECG, and it represents the electrical activity that occurs as the ventricles contract to pump blood out of the heart. The QRS complex is typically composed of three waves: the Q wave, the R wave, and the S wave [14]. T wave is the final wave on the ECG, and it represents the electrical activity that occurs as the ventricles relax and prepare to fill with blood again [15].
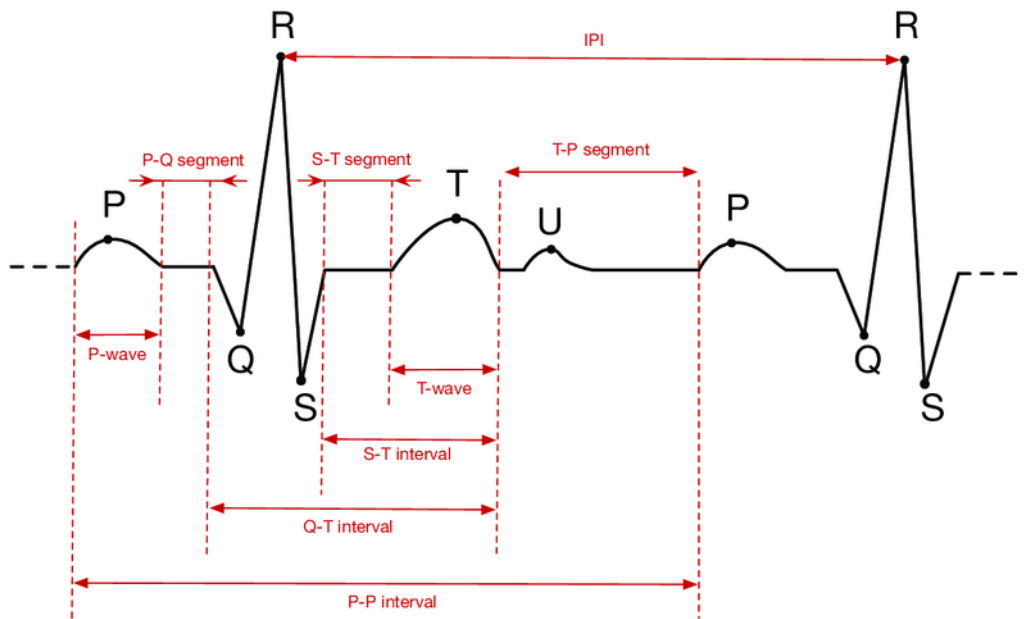
Figure 2.2: with peaks P, Q, R, S, T, U. Retrived from [16], used with CC BY-SA 4.0 license .

2.3.1 Single-Lead ECG

A single-lead ECG (1-L ECG) uses two electrodes, which can, for instance, be placed on the chest or arm to detect ECG signals [17]. One advantage of 1-L ECGs is their low cost and increased user comfort. Compared to a 12-lead system, a 1-L ECG requires fewer hardware components, and the patient experiences more comfort and mobility because there are fewer attachments to the body. In addition, with a wireless patch in a 1-L ECG system, patients can move and function normally without the hassle of wires.

However, a significant disadvantage of 1-L ECGs is the need for more comprehensive information in diagnosing heart diseases. Some heart conditions cannot be detected with a 1-L and require information from multiple angles of the heart, provided by a 12-lead system. Nevertheless, research indicates that advancements in microelectronics, biosensors, and optimized software algorithms can make single-lead systems medically relevant [17], [18].

2.3.2 Three-Lead ECG

A monitoring system with a simple three-electrode setup uses two electrodes for active monitoring and a third as a ground electrode. These electrodes can be configured in various ways to capture lead I, II, or III individually, with the signal acquisition being bipolar between the selected pair for the specific lead. The electrodes are placed in such a way as to minimize movement artifacts during continuous monitoring [19].

For a three-lead ECG, the three electrodes are placed as follows: one below the right collarbone, the other below the chest, and the last one below the left collarbone.
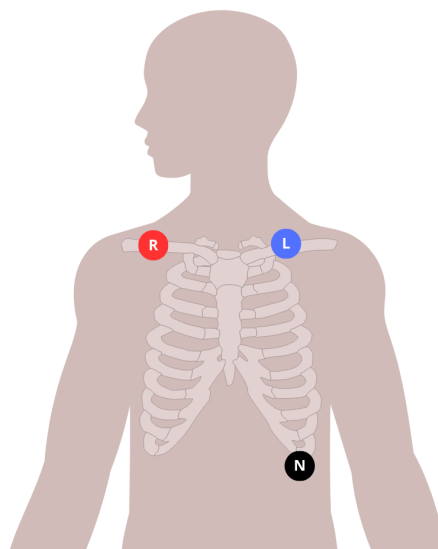
Figure 2.3: Lead placement for three-lead ECG

2.3.3 HRV

HRV is the subtle variation in a series of beat-to-beat measures of the heart.

HRV refers to the subtle variations in the intervals between consecutive heartbeats. It involves the variation in the RR interval, which is the distance between the peaks of the R wave in the QRS complex observed on an ECG, as seen in Fig. 2.2. This interval signifies the time between successive contractions of the ventricles [20].

Understanding the intricacies of the heart's electrical system is crucial. Nodes within the heart generate electrical impulses, prompting the ventricles to contract and propel blood throughout the body. Each contraction, represented by an R wave on the ECG, corresponds to a heartbeat. An individual's heart rate over time is not uniform, which means the intervals between beats show variability [21].

2.3.4 Interbeat Intervals (IBI)

HRV is a physiological measure of variations between consecutive heartbeats and can be measured from the Interbeat Interval (IBI). IBI can be derived from an ECG signal in different ways, most often from the beginning of the QRS complex of an ECG wave, and can be called QQ intervals, but IBI can also be derived from the peaks in the QRS complex and is then called RR intervals. In many cases, it can be difficult to locate the QQ intervals, so RR intervals are often used because of easily detectable reference points [22].

Another common type of IBI is the NN-intervals, which are the intervals between sinus beats, in other words, normal heartbeats. These NN-intervals are also detected using R-peaks of an ECG signal [22].

2.3.5 Relation between HRV and training

HRV is an efficient, inexpensive, and non-invasive tool to determine recovery status and an athlete's adaption to stress, both in exercise and other stresses in life.

In general, a higher HRV is considered positive, as it indicates a responsive nervous system and a lower HRV indicates that it reduces responsiveness. Similarly, a lower Resting Heart Rate (RHR) is considered a positive response to training, and a higher RHR is considered a sign of fatigue. A straightforward positive scenario involves an increased HRV and a decreased RHR, indicating effective coping with training and life stress [21].

When one of the metrics is considered normal and one is indicating a problem, additional context information is needed. An elevated HRV coupled with an increased RHR generally suggests fatigue or potential illness unless observed at the outset of a brief training phase. A reduced HRV and an elevated RHR often point to accumulated physical and mental fatigue unless an athlete is tapering, in which case it might signal improved readiness to perform. Tapering is a strategy to reduce training volume leading up to an event or competition [23].

A decrease in HRV and a decrease in RHR is generally a sign of prolonged low-intensity, high-volume training or large changes in life stress. It can be a sign of overtraining if it is not reversed with rest [23].

2.4 Detrended Fluctuation Analysis

Detrended Fluctuation Analysis Alpha 1 (DFA) Alpha 1 is non-linear analysis metric that quantifies the shorrt- and long-term correlation properties of heart rate variability. [24], [25].

DFA is calculated by dividing a data sequence X_t into consecutive time windows, where each window contains n samples. DFA is then estimated for each window using Eq. (2.1), where X_t is the cumulative sum of the signal and Y_t representing the piecewise strait line. Then, the RMS deviation from the trend (fluctuation) is computed.

$$f(n) = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (X_t - Y_t)^2} \tag{2.1}$$

A higher DFA Alpha 1 value indicates better health as it has stronger long-term correlations in the HRV signal. A lower DFA Alpha 1 value indicates weaker long-term correlations. A lower value is therefore associated with higher intensity training and higher heart rate. With increased exercise and stress, the heartbeats will become more chaotic. Additionally, a lower DFA Alpha 1 value can be observed in people with heart disease, hypertension, diabetes, and depression [1], [25].

2.5 Time Series Data

Time series data is a set of numeric observations recorded over time. It can be found in a variety of domains, such as finance, weather, and health. Time series data is often analyzed due to its uncertainty; we want to understand and predict its future behavior. Time series data is represented with an x-axis of time, anything from seconds to years [26].

When discussing time series data, we often look at its trends, which can be explained as the long-term changes in the mean.

HRV data is an example of time series data.

2.6 Classification and Machine Learning

Machine Learning (ML) is a field where algorithms that can predict by using data and finding hidden patterns. A ML model is trained and tested by dividing the dataset used for the model

into three parts: the training set, the validation set, and the test set. These three parts of dataset are utilized at different stages of the classification process.

For each ML model constructed, there are parameters that determine the significance of each feature in the prediction. These parameters are known as weights and are optimized using the training set, creating the best model for the given configuration. The process of generating these models is determined by the algorithm's hyperparameters and is fine-tuned using the validation set [27].

2.6.1 Support Vector Machines (SVM)

Support Vector Machine (SVM) are one of the most influential approaches to supervised learning. SVM works by finding the most optimal hyperplane to classify the data into various classes [28].

SVM is driven by the linear function $w^T x + b$. The function gives out a positive or negative class based on its location to the hyperplane [28].
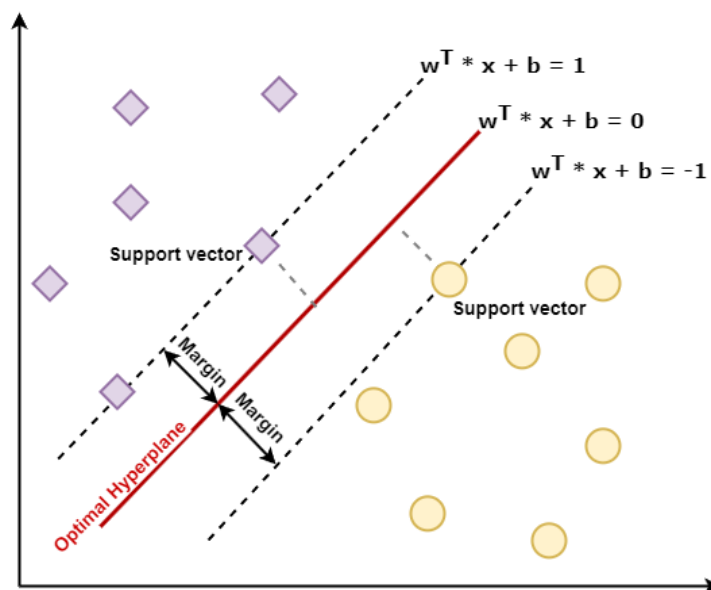
Figure 2.4: Support Vector Machine

SVM is often used in classification as the goal of the SVM is to find a hyperplane in N-dimensional space and use the hyperplane to classify data points. The size of N, which is the dimension we have in the hyperplane, depends on the number of features we have [28], [29].
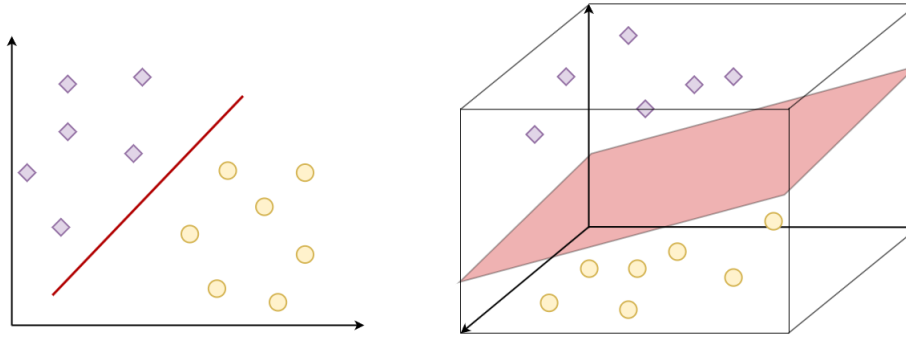
Figure 2.5: Hyperplane of N=2 and N=3

Hyperplanes serve as decision boundaries for classifying data points, separating them into distinct classes based on their placement relative to the hyperplane. The dimension of the hyperplane is dependent on the number of features involved. If we have a case of two features, the hyperplane will exist in 2-dimensional space and take the form of a line. With three input features, it exists in the three-dimensional space, and the hyperplane will be two-dimensional. as illustrated in Figure 2.5. Visualizing hyperplanes becomes increasingly challenging as the number of features surpasses three [30].

SVM can be divided into two parts: linear and non-linear. These can further be divided into the different kernels typically used, for instance, linear, polynomial, gaussian, and Radial Basis Function (RBF) [29].

Support vectors are the data points closest to the hyperplane and are used to determine the hyperplane characteristics, such as the position and orientation [30]. A larger margin is optimal, as the classes have greater boundaries.

2.6.2 Neural Networks

Neural Network (NN) is a category of ML algorithms inspired by the human brain. It consist of interconnected processing units, also called neurons. These neurons work together to learn from data [28].

Each neuron receives input from others, processes it by assigning weights, and generates an output transmitted to other neurons. Fig. 2.7 illustrates the process of a single neuron. NN have proven effective inn various tasks, including classification [31].

The learning process of NN employs backpropagation, where the error between predicted
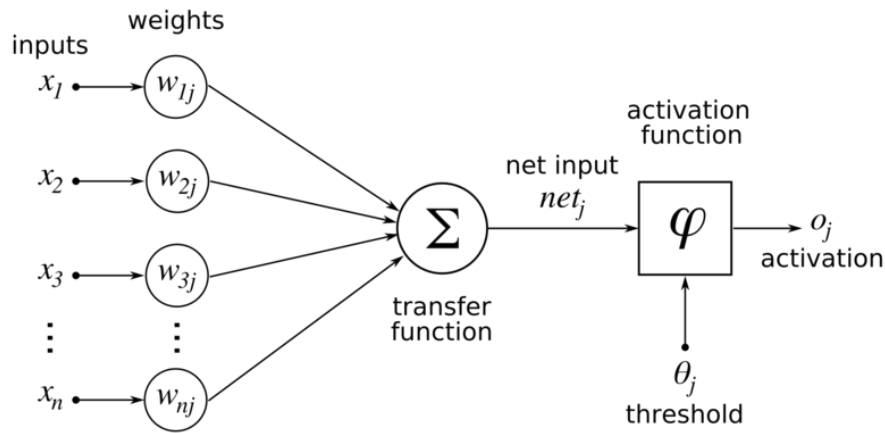
Figure 2.6: Illustration of a single neuron for ML. Illustration by Wikimedia user Chrislb, used under the CC BY-SA 3.0 license .

and actual outputs is propagated back through the network. Then it is used to adjust the neurons' weights [32]. Because of this, neural networks can learn complex relationships between input and output data and find applications across domains [33], [34].
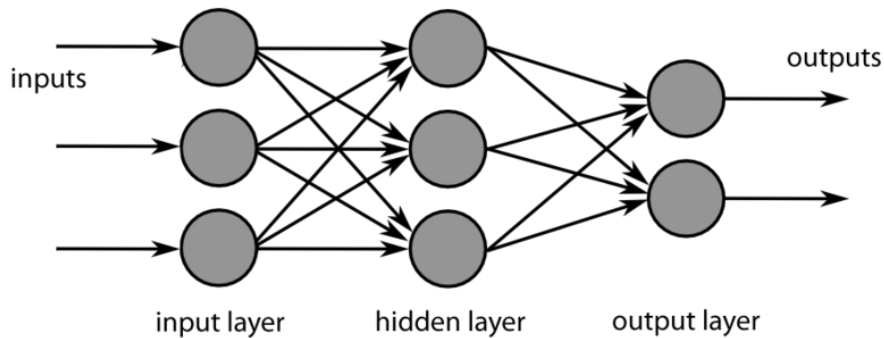
Figure 2.7: Illustration of a NN. Illustration by Wikimedia user Chrislb, used under the CC BY-SA 3.0 license .

Deep learning is a sub-field of ML characterized by using deep NN, which involves hidden layers. These hidden layers are capable of learning hierarchical representations of data. They can learn complex features directly from raw data without needing manual feature engineering or preprocessing, so they have proven to be particularly adept in signal classification tasks.

Additionally, these models can efficiently manage large datasets and high-dimensional inputs through shared weights and local connections. They undergo end-to-end training using gradient descent algorithms with backpropagation, which allows them to optimize model parameters to minimize classification errors [31].

2.6.3 Convolutional Neural Network

Convolutional Neural Network (CNN)s are a type of NNs commonly used for classification tasks. Usually, they are made using three layers, with the convolutional layer being the most crucial—hence its name.

The convolutional layer consists of a filter or a feature detector that is moved across the data while checking if that specific feature is present. The output is a series of values that depict where the feature is present in the data, also known as a feature map. The process can be done multiple times using filters with different properties, allowing for more complex features to be found as the sum of its parts [35].

Following the convolutional layer are the pooling layer and the fully connected layer. The pooling layer reduces the possibility for overly complex models by reducing the number of parameters, and makes the algorithms more robust. Finally, the full-connected layer produces the output, translating from the similarity of features to an estimated match of the category [35], [36].

2.6.4 Recurrent Neural Network

Recurent Neural Network (RNN) are a type of NN that uses knowledge from previous outputs to improve the following results [37].

output layer
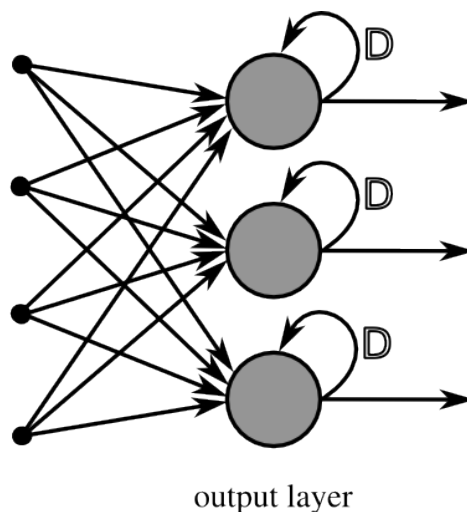
Figure 2.8: Illustration of an RNN by Wikimedia user Chrislb, used under the CC BY-SA 3.0 license .

RNNs is reliant on the previous nodes within the sequence, as visualized in Fig. 2.8. In a feed-forward network or multi-layer perceptron, each neuron executes its function with inputs

17

and passes the result to the subsequent node [37], [38].

In recurrent neural networks, a recurrent connection allows the output to loop back to the RNN neuron rather than solely progressing to the next node. Each node in the RNN model serves as a memory cell, sustaining the calculation and operation implementation [38].
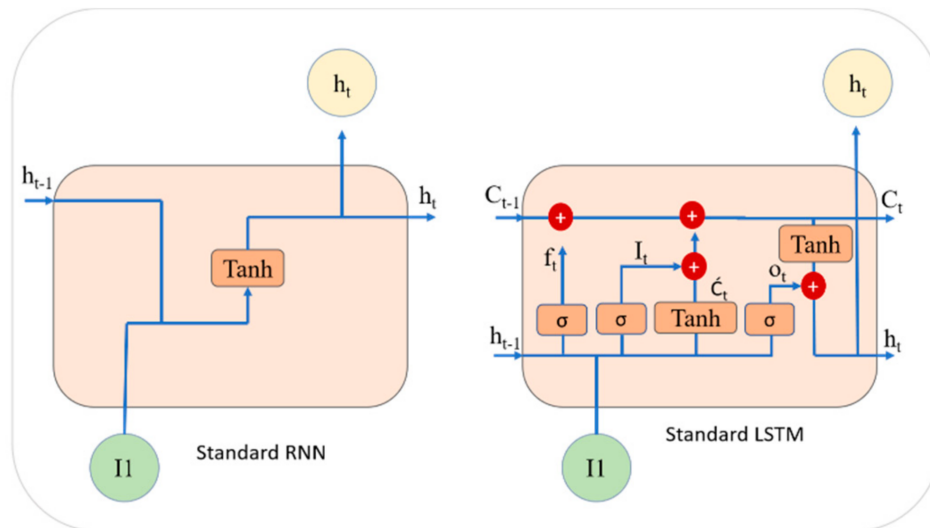
Figure 2.9: Standard architecture of RNN and LSTM models [39]. Used under Creative Commons Attribution (CC BY) license

In the event of an inaccurate forecast, the RNN self-learns with backpropagation, adjusting to achieve a more accurate prediction. As RNNs can recall past inputs, it's particularly effective in time series prediction, a concept known as Long Short-Term Memory (LSTM) [37].

2.6.5 Long Short Term Memory

Long-Short-Term Memory (LSTM) is a type of RNN designed to achieve more efficient outcomes by remembering information from previous nodes for an extended duration [37].

LSTM networks remove some of the challenges with traditional RNNs, such as vanishing gradients and long-term dependence issues. Gradient vanishing occurs when information is lost as connections recur over an extended period. LSTM solves this problem by disregarding unnecessary data in the network [37].

LSTM uses three gates as illustrated in Fig. 2.9. The LSTM has a forget, input, and output gate to regulate and update cell states. The forget gate decides which information in the cell state should be ignored, helping the LSTM remember essential inputs and produce accurate results, even with many inputs [40].

2.6.6 Transfer Learning

Transfer learning is a technique in machine learning that is used to train a task using a model trained on another task. This method is popular in deep learning. Transfer learning is especially popular in problems relating to image recognition and natural language processing. Given the vast computational and temporal resources needed to construct neural network models for these challenges, transfer learning offers significant advantages by using the knowledge gained from related problems [41]

A base network is initially trained on a base dataset and task in transfer learning. The acquired features are then repurposed or transferred to a second target network, which is trained on a different target dataset and task. This process is effective when the features are general to both the base and target tasks and will then improve the learning efficiency for the target task [42].
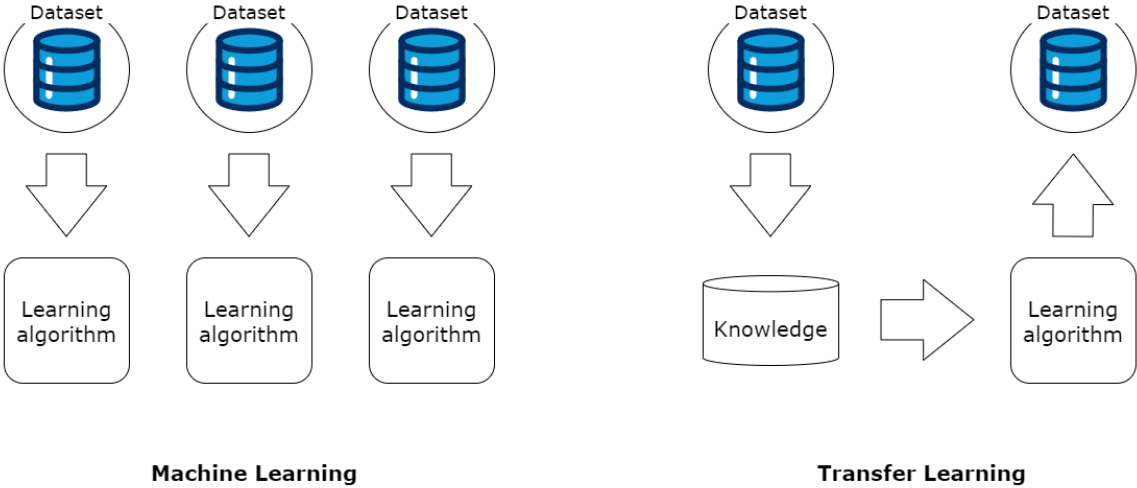
Figure 2.10: Traditional machine learning and transfer learning

Transfer learning in deep learning, specifically inductive transfer, benefits by narrowing the model bias by applying a model trained on a different yet related task. This approach proves valuable in optimizing the performance of models for new tasks or accelerating progress [43].

There are two common methods of applying transfer learning. The first method is to develop a model using a base dataset and then transfer the model to the target task. The reused model can employ all or some of the original model's parts [42].

The second method is to choose a pre-trained model from a set of available models trained

on large datasets. Then, reuse the model as a starting point for the target task. In the same way as the other method, parts or the whole model can be used. Lastly, you can adapt or fine-tune the model and redefine it based on the input and/or the target [42].

The decision to use transfer learning depends on factors like data availability, relatedness, and constraints. Transfer learning has several potential benefits, such as a higher initial accuracy, a better improvement rate during training, and a better convergence[42].

Chapter 3

Materials and Methods

This chapter will present all the steps and methods used. The steps are represented in Fig. 3.1, and will be presented in the same order. The first phase of the workflow involves collection of raw data, followed by the preprocessing of the raw signals to the desired format. Then, the feature extraction will be presented for the dataset provided by Olympiatoppen in addition to the Swell Knowledge Work (Swell-KW) dataset, which is used in transfer learning. Section 3.4.1 will present an in-depth presentation of the Swell-KW dataset before presenting the machine learning and classification used.
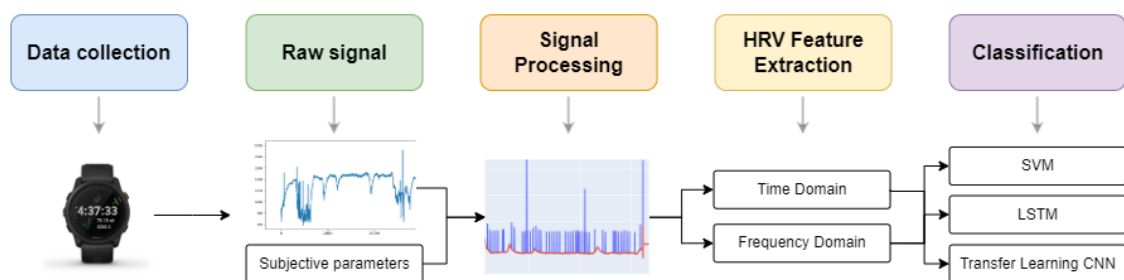
Figure 3.1: Overall workflow

3.1 Software Tools

The implementation of this project relied on *Python* and various packages. The interpolation and signal processing was implemented using the Signal library from *SciPy* [44]. *Numpy* [45] was mainly used for mathematical purposes. Feature extraction was computed using *pyhrv*[46] and *hrvanalysis*[47], both open-sourced toolboxes for HRV analysis. To present graphical data *Matplotlib*[48] was used.

3.2 Dataset

The dataset consists of two well-trained athlehes, recorded using a Forerunner 495 watch from Garmin.

The recordings took place in 2023, with Athlete 1 recording from May 15 to November 1st and Athlete 2 recording from January 25th to November 2nd. In all, there are 655 sessions, whereas 217 are without subjective logging.

After removing sessions without HRV data and subjective logging, we are left with 290 sessions. These sessions consist of various activities and intensities, as shown in Table 3.1.

Table 3.1: Overview of sessions recorded, number of sessions per activity, and average subjective measurements.

	Count of sessions	Average of RPE	Average of subjective score (1-5)
Candidate 1	**121**	**4.289**	**3.678**
Rowing	109	4.358	3.716
Run	6	3.500	3.500
Strength	6	3.833	3.167
Candidate 2	**169**	**4.130**	**3.379**
Bike	149	4.114	3.416
Run	19	4.211	3.105
Strength	1	5.000	3.000
Grand Total	**290**	**4.197**	**3.503**

The lengths of sessions also vary. Candidate 1's shortest session was 7 minutes, and the longest was 149 minutes. As for candidate 2, the shortest session was 2.5 minutes, and the longest was 269 minutes.

3.2.1 Protocol

The protocol consisted of two candidates, one at the Olympic level and one at a slightly lower level but still significantly better than the average person.

Olympiatoppen (OLT) has provided an intensity scale for their workouts. The intensity zones are correlated to the RPE CR10 scale (or RPE Borg, which will not used in this instance).

Table 3.2: Intensity zones defined by Olympiatoppen

Zone	RPE CR10	Description	% of max HR	Breathing
I-1	1-2	Very easy	~55% - 72%	Can talk effortlessly
I-2	2-3	Quite easy	~72% - 82%	Can say longer sentences relatively effortlessly
I-3	4-5	Pleasantly exhausting	~82% - 87%	Can say short sentences
I-4	6-7	Exhausting	~87% - 92%	Can say words or very short sentences
I-5	8-10	Very exhausting	~92% - 100%	Can only say one or two words while breathing heavily
I-6	9-10	Very very exhausting		Can only say short words while gasping for breath
I-7	10	Very very exhausting		Not applicable for controlling intensity
I-8	-	Maximal mobilization		Not applicable for controlling intensity

The candidates logged their subjective score after each session using TrainingPeaks, a training app from Peaksware, LLC. The app is created for endurance athletes. It is compatible with over 100 apps and devices, such as Apple Watch, Garmin, Polar, and Zwift. The app is a training diary and can be used to log all training, both subjective, but also adds the data for the sessions [49].

subjective score is indicated with a smiley face on how well you feel towards the questionnaire. These smiley faces are converted into values ranging between 1 and 9, where 1 is strong, and 9 is weak. These values were then converted into another scale for a more understandable range. The new scale goes from 1-5, where 1 is weak and 5 is strong. The subjective parameters were later exported to an Excel file (.xlsx).
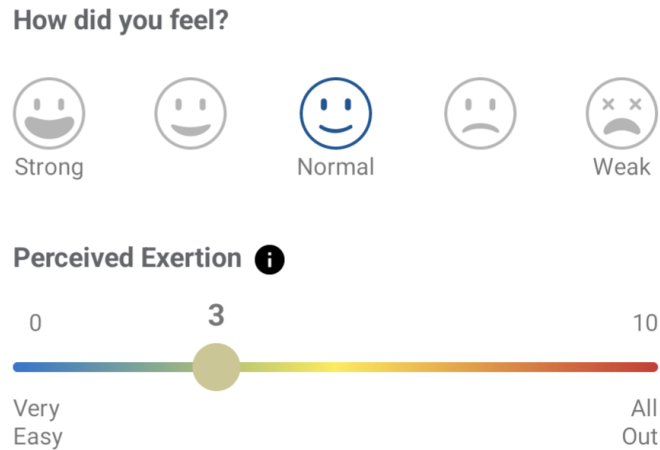
Figure 3.2: Display of subjective logging in TrainingPeaks application

On some days in the recording period, multiple sessions are recorded a day, some at different times and others because of a change in activity.

All data from the sessions was transferred using ANT+ technology. The files are in a format called *.FIT*, and can retrieved in Python using *fitparse*[50] library in Python.

3.2.2 Data Selection

The final data selection consists of 290 sessions, where 121 are from candidate 1 and 169 from candidate 2. The final data selection can be seen in Table 3.1 with the final number of sessions per activity, and its average subjective values.

The distribution of the number of sessions with each subjective score on the 1-5 scale is represented in Fig. 3.3. As we can see, most sessions are rated in the middle as a 3 or 4, where 3 is average, and 4 is a somewhat strong subjective score.

Count feeling (with HRV)	
1	3
2	3
3	41
4	57
5	17
	121
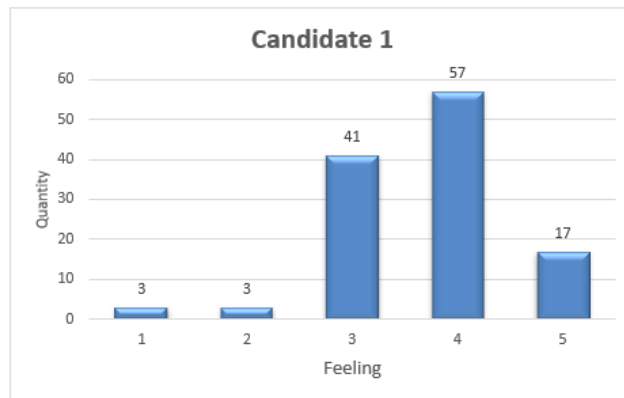
(a) Candidate 1

Count feeling (with HRV	
1	2
2	17
3	78
4	59
5	13
	169
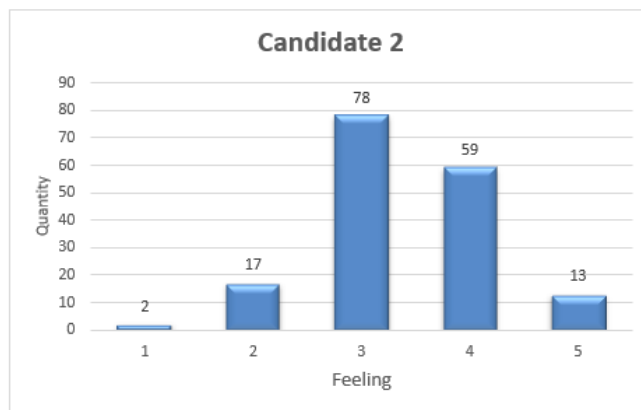
(b) Candidate 2

Figure 3.3: Histogram of how many files contain each subjective score

3.3 Preprocessing

3.3.1 Interpolation filtering

A simple interpolation filter was added to remove artifacts caused by using ANT+ transfer. As data is transferred using ANT+, we get artifacts in the data, which will look like spikes, as we can see in Fig. 3.4. These artifacts are the product of data packages that are missing in transfer, and the filter detects where the data has artifacts and if it is the result of one or two packages missing. Then, the filter interpolates based on the number of packages missing.

If one package is missing, the HRV data will spike double the value of the nearest values, and if two packages are missing, it will be three times more than the nearest values.

Figure 3.4 shows the difference between the original data and the final interpolated data. As seen in the original data, packages are missing, and therefore, we get spikes in the data. These spikes are then interpolated, giving us the data used further in this thesis.
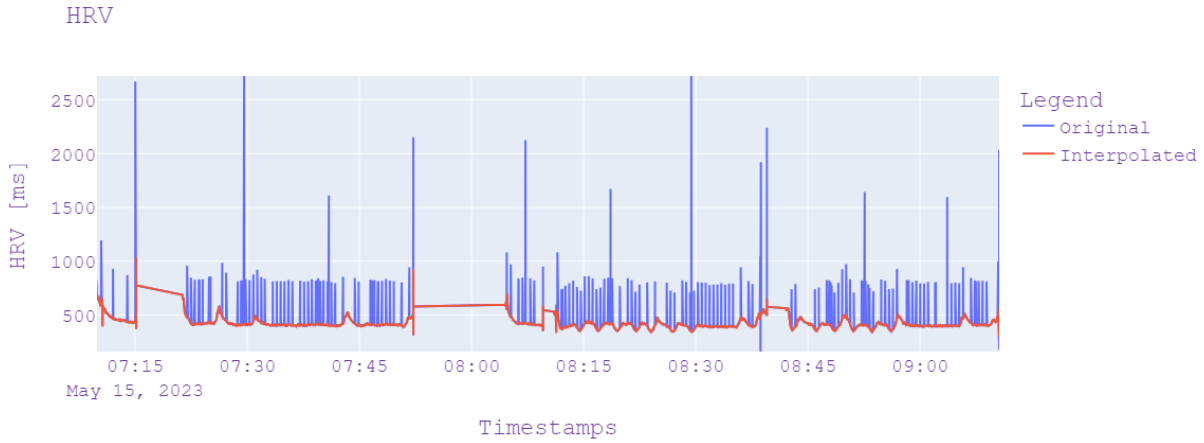
(a) Original
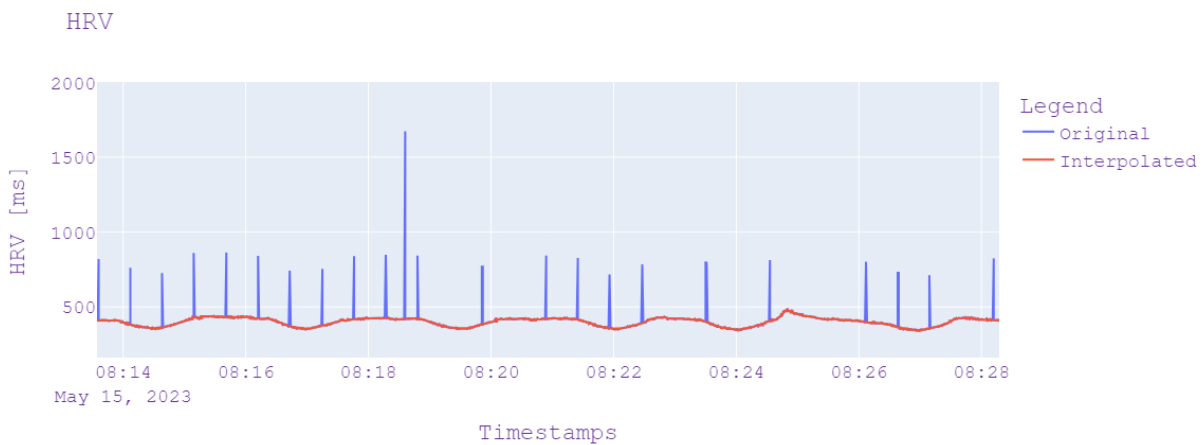
(b) Zoomed

Figure 3.4: Plot of interpolated data and original data.

3.4 Transfer learning with SWELL-KW

3.4.1 Swell Knowledge Work Dataset

Swell-KW is a dataset collected to research stress in workplace environments. Swell-KW is part of the SWELL project, SWELL being an acronym of Smart Reasoning Systems for Well-being at Work and at Home. [51]

The dataset was collected using 25 people who were performing tasks such as writing reports, making presentations, emailing, and information searching. Of the 25 participants, 8

were female and 17 male, with an average age of 25 (standard deviation 3.25). They were exposed to stress factors such as email interruptions and time pressure when performing these tasks. It collected data from various sensors and recorded computer logging, facial expressions, body postures, heart rate, and skin conductance. From this, we will only be using the heart rate data, in which we have looked at the entire process from raw data to preprocessed data, then finally to feature extraction. [52]

The participants in the experiment went through working conditions such as no stress, time pressure, and interruption. During no stress the participants are allowed to work on the tasks as long as they needed for a maximum of 45 minutes, they are not aware of the maximum duration of their tasks. In time pressure the time to finish a task was reduced drastically. The time was reduced to 2/3 of the time they used with no stress. Finally, in the interruption condition they received eight emails during their assigned tasks, some emails were relevant as they gave requested actions for the tasks, while other was irrelevant to their tasks. The results from this exercise will be used to estimate perceived stress. [52]

The dataset is available at: DANS: The SWELL Knowledge Work Dataset for Stress and User Modeling Research

The raw data is an ECG, which was recorded using a Mobi device with a recording frequency of 2048 Hz. The electrodes were placed with self-adhesive electrodes, one below the right collarbone, the other below the chest, and one below the left collarbone, as illustrated in Figure 3.5.

The ECG data was then preprocessed using the recording software Portilab2. Both raw and preprocessed ECG data were provided. Some preprocessing was programmed into the recording software Portilab2, giving a filtered signal with clear peaks in the QRS complex.

Then the heart rate and heart rate variability were calculated. To calculate this they used a peak detection algorithm for the filtered signal in Matlab and calculated the distance between R-R peaks. Heart rate variability was determined using RMSSD.

For the feature extraction, the features were calculated from the inter-beat intervals, in this case, RR-intervals, that were gathered from the ECG signal. The HRV indexes are computed on 5 minutes of RR data, using a sliding window. In other words, a new sample is added to the data, and the oldest is removed and then HRV is calculated again. This is done until the end of the entire signal.
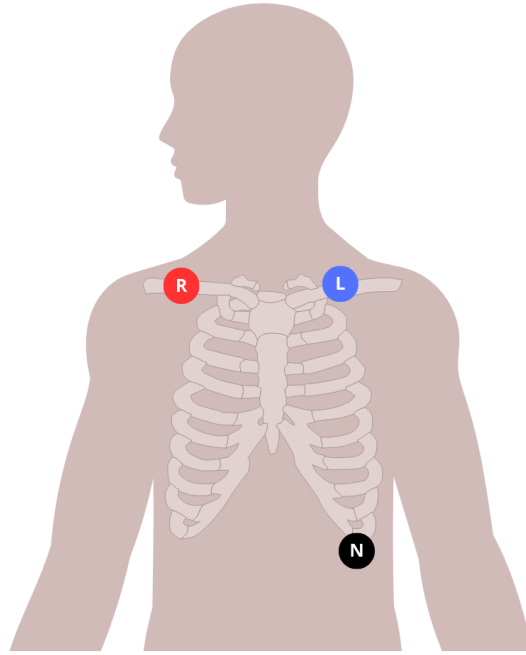
Figure 3.5: Lead placement for three-lead ECG

Table 3.3: Features in SWELL-KW

Feature	Short Description	Source
MEAN_RR	Mean of all RR intervals	
MEDIAN_RR	Median of all RR intervals	
SDRR	Standard deviation of all RR intervals	[53]
RMSSD	Square root of the mean of the sum of the squares of the difference between adjacent RR intervals	[53]
SDSD	Standard deviation of all interval of differences between adjacent RR intervals	
SDRR_RMSSD	Ratio of SDRR over RMSSD	
HR	Heart Rate (beats per minute)	
pNN25	% of adjacent RR intervals differing by more than 25 ms	
pNN50	% of adjacent RR intervals differing by more than 50 ms	
SD1	Poincaré plot descriptor of the short-term HRV	
SD2	Poincaré plot descriptor of the long-term HRV	
KURT	Kurtosis of all RR intervals	
SKEW	Skewness of all RR intervals	
MEAN REL RR	Mean of all relative RR intervals	[53]
Continued on next page		

Table 3.3 – continued from previous page

Feature	Short Description	Source
MEDIAN_REL_RR	Median of all relative RR intervals	[53]
SDRR_REL_RR	Standard deviation of all relative RR interval	[53]
RMSSD_REL_RR	Square root of the mean of the sum of the squares of the difference between adjacent relative RR intervals	[53]
SDSD_REL_RR	Standard deviation of all interval of differences between adjacent relative RR intervals	[53]
SDRR_RMSSD_REL	Ratio of SDRR_REL over RMSSD_REL	[53]
KURT_REL_RR	Kurtosis of all relative RR intervals	[53]
SKEW_REL_RR	Skewness of all relative RR intervals	[53]
VLF	Very low (0.003Hz - 0.04Hz) frequency band of the HRV power spectrum	[54]
VLF_PCT	Percentage of VLF to total	
LF	Low (0.04Hz - 0.15Hz) frequency band of the HRV power spectrum	[54]
LF_PCT	Percentage of LF to total	
LF_NU	LF in normalized unit	
HF	High (0.15Hz - 0.4Hz) frequency band of the HRV power spectrum	[54]
HF_PCT	Percentage of HF to total	
HF_NU	HF in normalized unit	
TP	Total HRV power spectrum	[54]
LF/HF	Ratio of LF to HF	[54]
HF/LF	Ratio of HF to LF	[54]
sampen	Sample entropy of the RR signal	
higuci	Higuchi Fractal Dimension	

3.4.2   Classification using Swell-KW

A model was trained using the Swell-KW dataset. The general model was provided by a repository on GitHub [55] and then fine-tuned locally. A pre-trained CNN model was trained using the Swell-KW datasets and the provided tree classes (interrupted, stress, and no stress) with an accuracy of 0.99.

In order to understand the model trained a bit more, we can look deeper into the importance

of each feature. The feature importance can be used to make feature reduction. We could cut the least important features and keep the features that improve the model. From the *scikit-learn (scklearn)* library in Python, we can use the predefined function `SelectBest`, which is a function that select features according to the highest scores.
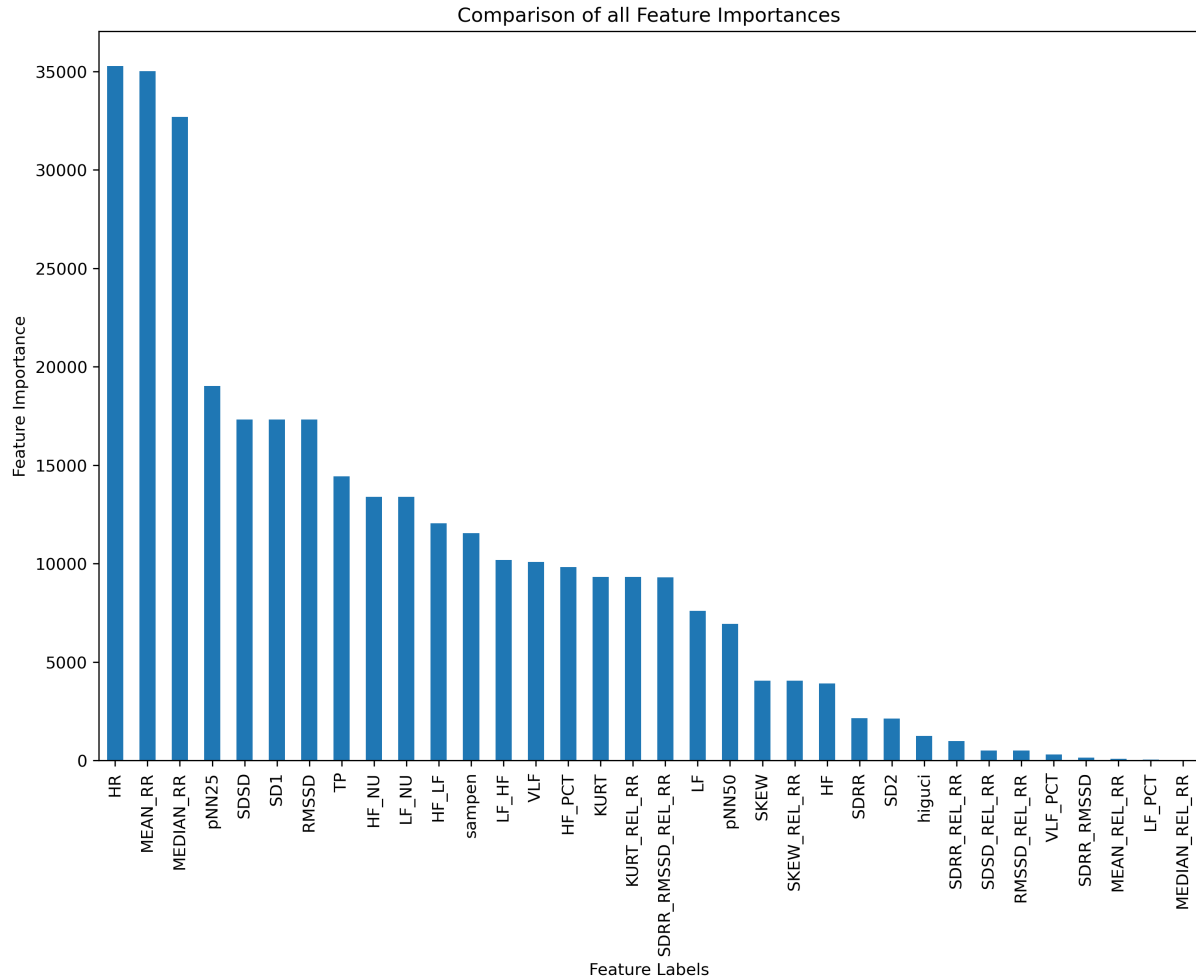
Figure 3.6: Feature importance Swell-KW dataset

The results of the feature importance for the Swell-KW are illustrated in Fig. 3.6. As we can see, features such as HR, Mean RR, and Median RR are important to classify the correct classes. Most of the features consisting of relative values are not as important as the corresponding features that use the entire data window.

3.4.3 Data Processing for transfer learning

To do transfer learning with our dataset and the Swell-KW dataset, all features needed to be calculated the same way they were for the Swell-KW dataset.

Given a file with HRV data that has been preprocessed with the interpolation, we will do the feature calculation. This will be done to all the HRV files to create a train and test dataset, which we can use for transfer learning. How the features are calculated are given by formulas in Table 3.3, and using the same method as described in Section 3.4.1.

Training, Validation, and Test Data

In ML, we divide the data into three sections: training, validation, and test data. Most of the data will be training data, and then 20% will be test data. Validation data is 10% of the training data.

The data is divided in the Python code. Firstly, the train and test data is divided using the `train_test_split` from the *scikit-learn (scklearn)* library in Python. The training data is then split into training and validation during the `fit` function, also from the scikit-learn library. This function is used to train the models, given the datasets and the labels for training.
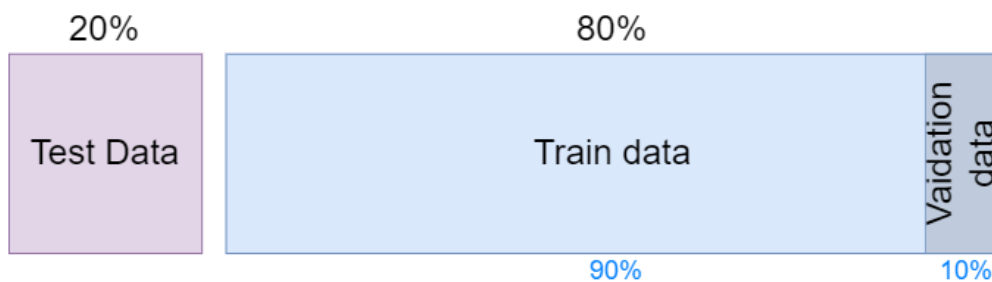
Figure 3.7: Distribution of test, train, and validation data

3.5   Metrics for Evaluation

This section will discuss the metrics that will be used to evaluate the methodology's effectiveness and performance. The evaluation of a ML model will use some generic metrics, such as accuracy, f-number, and confusion matrix.

3.5.1   Evaluation of Classification

A confusion matrix will be employed to evaluate the classification models generated. In a two-class classification, the matrix will be a 2×2 matrix, on the form shown in Table 3.4, where the diagonal presents the number of correctly classified samples, while the other elements are the number of incorrect classifications.

		Predicted	
		Positive	Negative
Actual	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

Table 3.4: Illustration of a 2-class confusion matrix

Additionally, the models will be evaluated using general performance metrics, such as accuracy. Accuracy is given by Eq. (3.1).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.1}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

Just using accuracy is not enough, as a model can give out great accuracy while just predicting one class. Therefore, we also use another metric called F-number. The F-number provides a single score that balances precision and recall in one number and is given by Eq. (3.2) or Eq. (3.3).

$$F - score = \frac{1}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{2 * Precision * Recall}{Precision + Recall} \tag{3.2}$$

$$F - score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{3.3}$$

As we can see, Eq. (3.2) used both recall and precision in its equation. Recall (Eq. (3.4)) is how many positive cases the classifier correctly predicted over all the positive cases in the data. In other words, how many are predicted as positive out of all the positive examples?

$$Recall = \frac{TP}{TP + FN} \tag{3.4}$$

Precision (Eq. (3.5)) is how many positive predictions are correct. So, out of all the predicted examples as positive, how many are actually positive?

$$Precision = \frac{TP}{TP + FP} \tag{3.5}$$

The optimal value is 1, meaning all predictions were correct. If the dataset was unbalanced, other metrics might highlight the classification results.

Chapter 4

Results

4.1 Correlation between subjective parameters and HRV

A correlation matrix is a tool used to quantify the strength and direction of linear relationships between variables. The matrix is symmetric, with variables listed in rows and columns, and each cell contains the correlation coefficient representing the relationship between the corresponding pair of variables.

The correlation coefficient ranges from -1 to 1. A positive value indicates a positive correlation, meaning that as one variable increases, the other also tends to increase. A negative value signifies a negative correlation, suggesting that as one variable increases, the other tends to decrease. A correlation of 0 implies no linear relationship between the variables.

When interpreting a correlation matrix, it's crucial to remember that correlation does not imply causation. Even if two variables are strongly correlated, it does not necessarily mean that one variable causes the other to change. Additionally, outliers can strongly influence correlation coefficients, so examining the data for potential anomalies is essential.

As we can see in the figure 4.1, the subjective parameter subjective score does not correlate with any of the other parameters in the matrix.

35
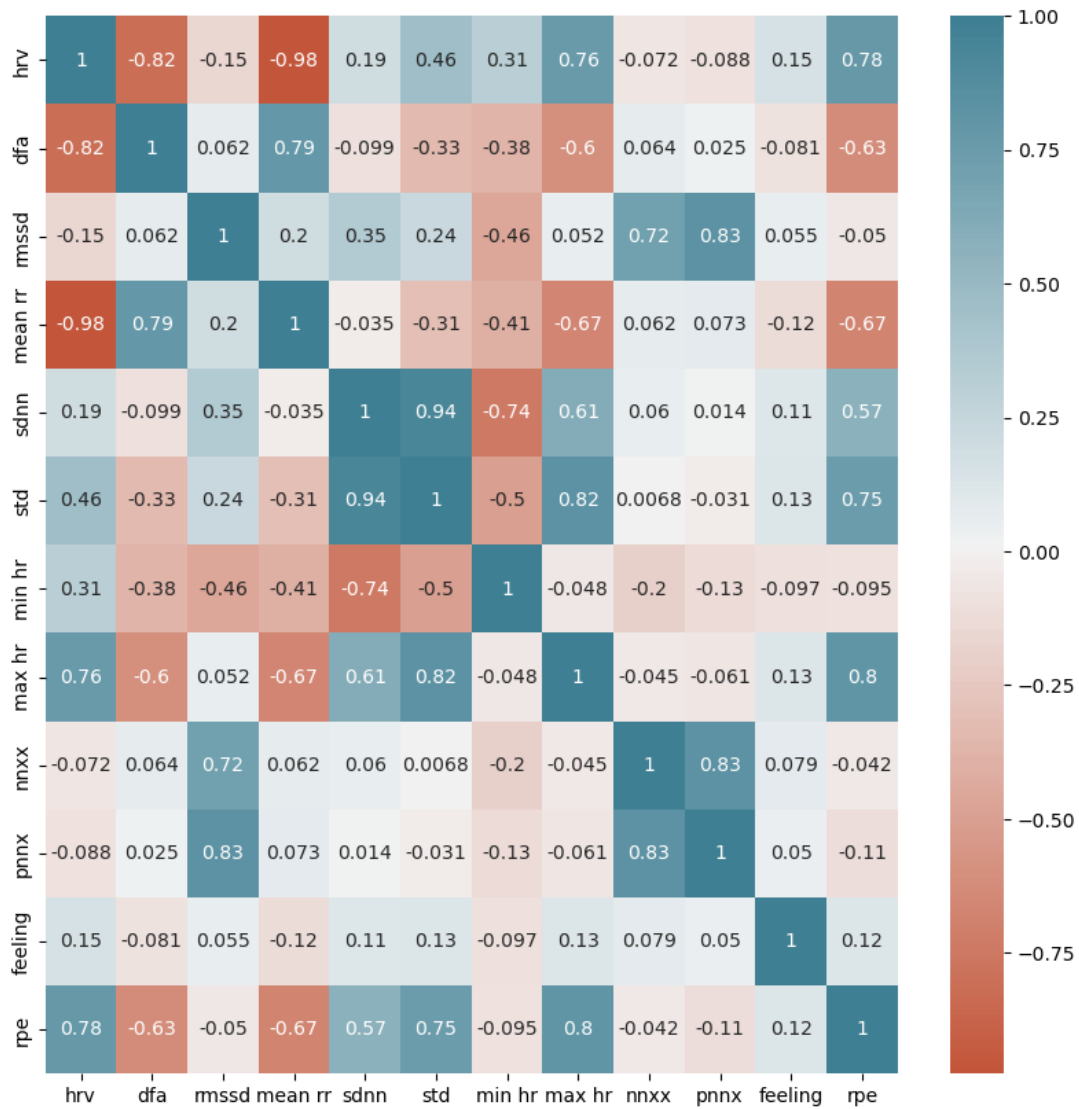
Figure 4.1: Correlation Matrix with various parameters, both subjective and objective

Given the equation to calculate DFA Alpha 1, DFA Alpha 1 strongly correlates with HRV and mean RR values, as expected. RPE is strongly correlated to max hr and HRV, and as RPE is dependent on the session's intensity, it makes sense that they are correlated. RPE is also somewhat inversely correlated to DFA Alpha 1; given that DFA is correlated to HRV, this is expected.

We could add more features to the correlation matrix, but this work did not consider it necessary.

4.2 Prediction of subjective parameter subjective score using LSTM

Using a LSTM model, it was attempted to perform a binary classification. Attempting to predict and differentiate between subjective score weak and strong, where subjective score 1 and 2 were together as class bad, and subjective score 3, 4 and 5 were combined to class good. This distribution of the various scores are visualized in Table 4.1.

Table 4.1: Distribution of subjective score for binary classification

Class bad	Class good
1 2	3 4 5
Weak	Strong

This LSTM model could not predict the correct subjective score as the model only predicted a stong subjective score for all cases. This is visualized in the confusion matrix in Fig. 4.2.
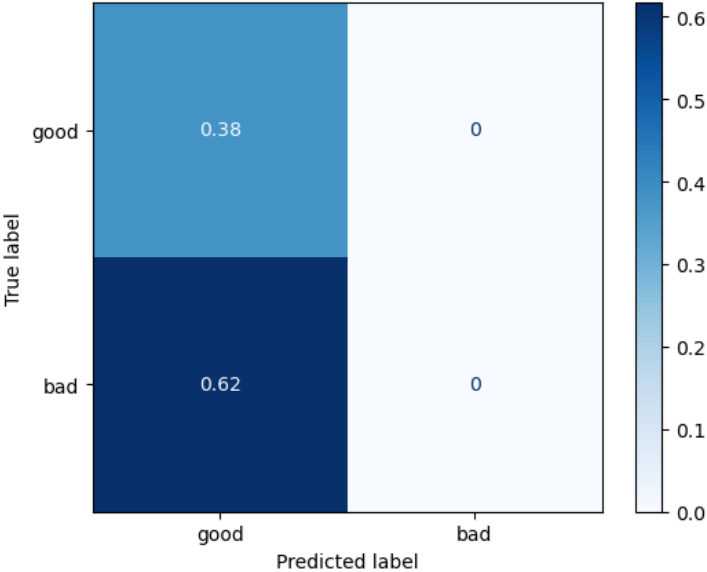
Figure 4.2: Confusion matrix for LSTM model

As explained in **??**, recall is how many of the cases in the class the classifier correctly predicted over all the cases in the data. The precision is how many of the predictions of the class are correct.

With that in mind, we can take a look at the classification report (Table 4.2). Here we can see

that the recall for class 0, which is a good subjective score, is equal to 1. The F-score, however, is low. This is caused by bad precision. For class 1, which is the bad subjective score, we have precision and recall as 0, meaning there are no predictions for that class.

Table 4.2: Classification Report of LSTM model

	precision	recall	f1-score	support
0	0.38	1.00	0.55	224602
1	0.00	0.00	0.00	362232
accuracy	0.80	1839	0.38	586834
macro avg	0.19	0.50	0.28	586834
weighted avg	0.15	0.38	0.21	586834

4.3 Prediction of subjective score using transfer learning

4.3.1 Binary Transfer Learning Model

The first attempt at transfer learning was to do a binary classification where subjective scores 1 and 2 were combined, and subjective scores 3, 4, and 5 were combined.

Table 4.3: Distribution of subjective score for binary classification using transfer learning

Class 0	Class 1
1 2	3 4 5
Weak	Strong

As seen in Fig. 4.3, subjective scores 1 and 2, otherwise known as weak subjective scores, were not able to be predicted once. All predictions were classified as strong subjective score. The dataset is imbalanced as most of the sessions given is scored as a 3 (average) and 4 (slightly above average) for both athletes. This makes the train and test dataset imbalanced as it will be a large percentage of the data in these classes and very few in the 1 and 5 classes. This can make it harder for the model to determine what differentiates the classes. As the dataset is imbalanced, the model found it difficult to differentiate between classes as they are quite similar in values for the features.
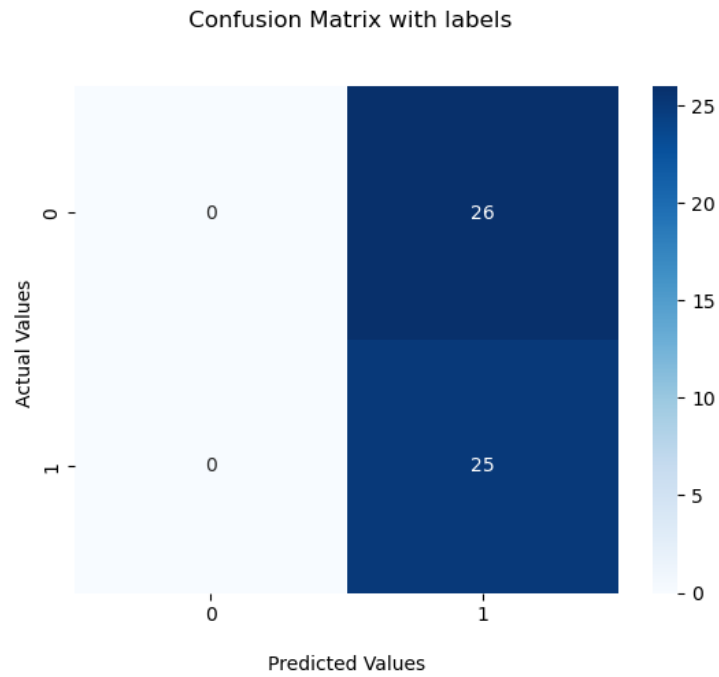
Figure 4.3: Confusion matrix for binary classification using transfer learning

4.3.2 Outer points classification using transfer learning

In the second model, we tested just the outer points of the subjective subjective scores, so very strong vs very weak, to see if it could detect each side of the scale.

Table 4.4: Distribution of subjective score for binary classification, outer points

Class 0				Class 1
1	2	3	4	5
Weak				Strong

So, using transfer learning, just comparing 1 and 5, then we got an accuracy of 0.8004. Fig. 4.4 shows the confusion matrix of the model, and as we can see, it mostly classifies correctly.
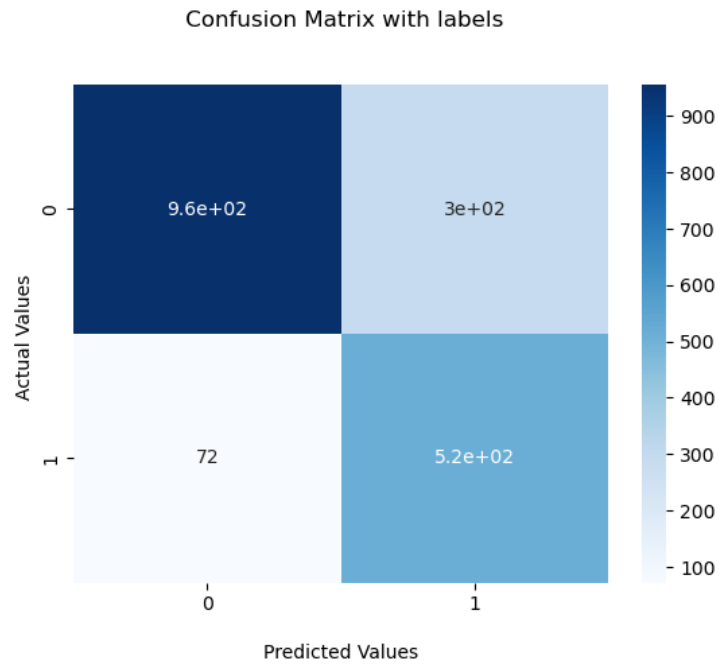
Figure 4.4: Confusion matrix of ends of scale, 1 vs 5.

In Fig. 4.5, the learning curve is displayed. The accuracy of both the train and test sets increases drastically in the beginning before stabilizing.
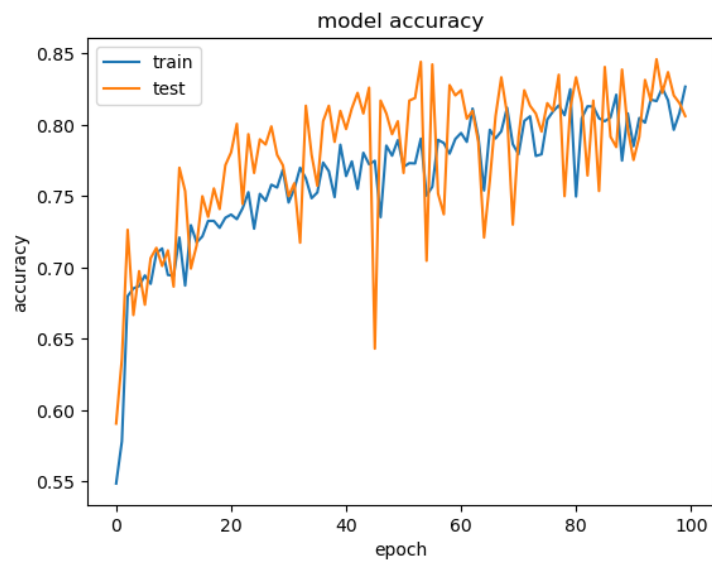
Figure 4.5: Accuracy of train and test set, 1 vs 5.

Table 4.5: Classification Report of 1 vs 5

	precision	recall	f1-score	support
0	0.76	0.93	0.84	1028
1	0.88	0.64	0.74	811
accuracy	0.80	1839		
macro avg	0.82	0.78	0.79	1839
weighted avg	0.81	0.80	0.79	1839

4.3.3 Three class classification using transfer learning using all of data set

After trying to differentiate the classes 1 and 5, it was attempted to differentiate the 5 classes into three classes, then 1 with 2, 3 alone, and 4 with 5.

Table 4.6: Distribution of subjective score for three class classification, all scores

Class 0		Class 1	Class 2	
1	2	3	4	5
Weak		Average	Strong	

This worsened the accuracy drastically, and then got an accuracy at 0.364. As we can see in Fig. 4.6, the model only predicts the worst subjective score and gives us a bad f1-score for all three classes.

Table 4.7: Classification report, three classes 1,2 vs 3 vs 4,5.

	precision	recall	f1-score	support
0	0.36	1.00	0.53	2858
1	nan	0.00	nan	2350
2	nan	0.00	nan	2639
accuracy	0.36	7847	0.79	1839
macro avg	0.36	0.33	0.53	7847
weighted avg	0.36	0.36	0.53	7847
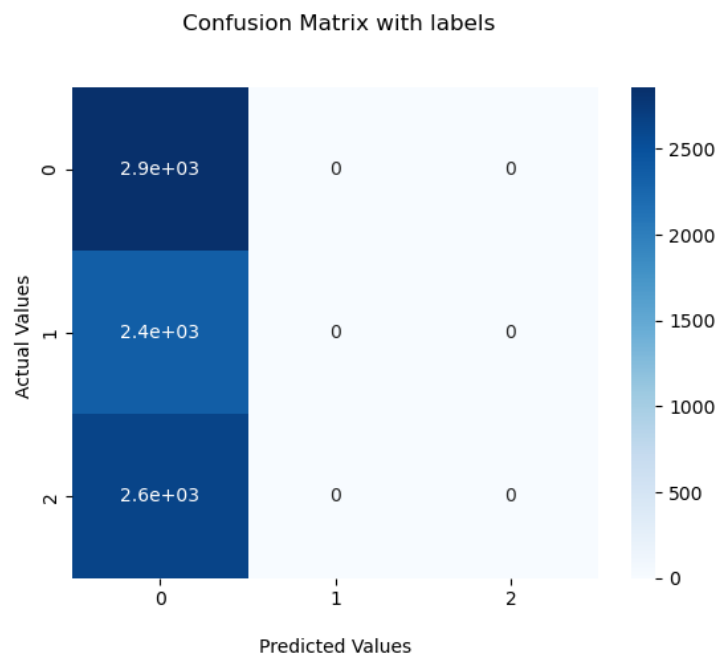
Figure 4.6: Confusion matrix, three classes, 1,2 vs 3 vs 4,5.

4.3.4 Three class classification using transfer learning using part of dataset

As the accuracy decreased from the end of the scale to three classes, it was attempted to do the same with just the end of the scale and the middle class. So we tried to classify class 1, 3 and 5.

Table 4.8: Distribution of subjective score for three class classification, selected scores

Class 0		Class 1		Class 2
1	2	3	4	5
Weak		Average		Strong

This resulted in an accuracy of 0.33, with only predicting class 1, which, as seen in Table 4.8, is equal to an average feeling, and the score is 3.
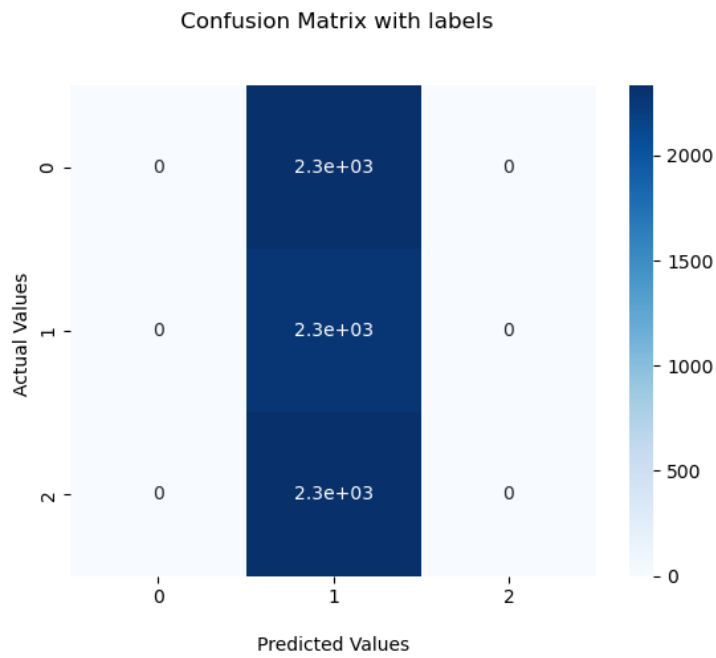
Figure 4.7: Confusion matrix, three classes, 1 vs 3 vs 5.

4.3.5 Balancing dataset

In an attempt to balance out the datasets, we randomly selected a number of files of each subjective score. The strong subjective score consists of subjective scores 3, 4, and 5, while the weak score consists of subjective scores 1 and 2.

Table 4.9: Distribution of subjective score for binary classification, balanced dataset

Class 0		Class 1		
1	2	3	4	5
Weak		Strong		

With a more balanced train set the model predicted with an accuracy at 0.84. As we can see in Fig. 4.8 we have predictions for both classes.
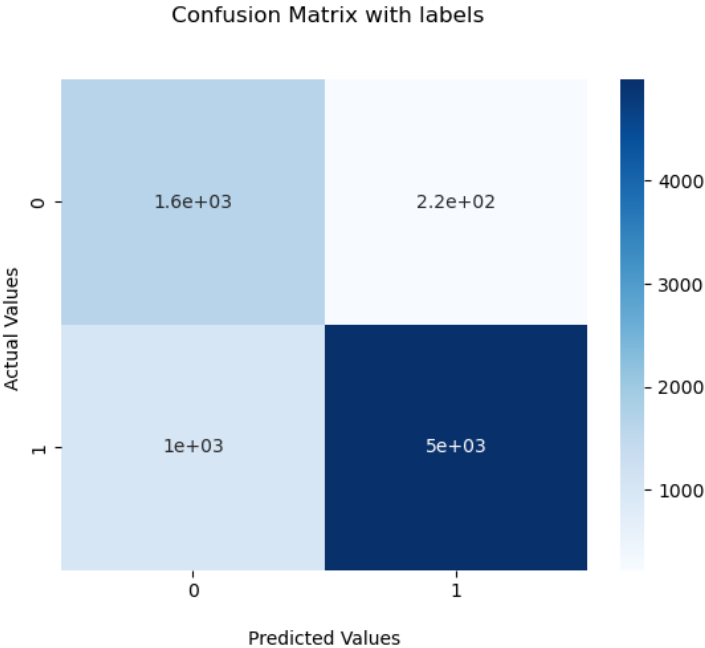
Figure 4.8: Confusion Matrix for balanced dataset

The precision and recall for both classes is better than in previous models. As we can see from Table 4.10, we have a lower recall value for class 0. This indicates that the model is not able to predict this class quite correctly. The precision for both classes is better than before, giving us an acceptable F1-score. Improvements still need to be made, but these results give hope for future work.

Table 4.10: Classification report, three classes 1,2 vs 3 vs 4,5.

	precision	recall	f1-score	support
0	0.88	0.62	0.73	2639
1	0.83	0.96	0.89	5208
accuracy			0.84	7847
macro avg	0.86	0.79	0.81	7847
weighted avg	0.85	0.84	0.84	7847

4.3.6 Comparison of Classification Results

As we can see from the results in the previous mentioned models, we get a bad accuracy for all instances using the original dataset. It is possible to differentiate and classify the outer subjective scores, but when we include the average subjective score the model is not able to predict the correct class. For all models except second model where just outer points were used, the model using transfer learning was not able to predict more than one class.

As the models are not able to predict more than one class, they will predict the class which is most frequent in the training set. This is for all cases the class containing the average subjective score 3. For instances where three classes were attempted, the models differ which class all predictions are made to.

In an attempt to balance the dataset, and then use the same model as before we get significantly better results. This can tell us that given a more balanced dataset it is possible to classify correctly. More about this will be discussed in Chapter 5.

Chapter 5

Discussion

The deployment of any machine learning model will need to go through an analysis and discussion of its results. In this chapter, we will discuss the results presented in Chapter 4 and then evaluate the model. We will discuss the effect of data limitations, the model's accuracy, and other critical points that shape the overall performance of the machine learning model.

Let's review the aim of this work before proceeding to the discussion. This work aimed to investigate methods for predicting subjective score for an athlete after a session and evaluate the effectiveness and feasibility of the prediction. Classification using LSTM and transfer learning with CNN were then used as models.

The objective is motivated by the wish to predict how the athlete's subjective score so that Olympiatoppen can use the results to determine the same athlete's readiness-to-train.

5.0.1 Lack of sufficient data in dataset

A high-quality dataset is crucial for data analysis when it comes to ensuring the accuracy and reliability of the findings. The dataset used in this project was gathered throughout the year 2023 by two athletes and contains 290 sessions of various lengths. The data consists of HRV for each training session, which then has gone through a feature extraction, resulting in 34 features used.

In general, machine learning and deep learning often want as much data as possible. One of the primary challenges in this project was the limited dataset, both in number of sessions

and number of athletes. Another challenge in the training was the uneven distribution of each class in the model. 80% of the sessions were classified as a score of 3 and 4 by the athlete, whereas the remaining 20% of the sessions were either 1,2 or 5.

The success of many machine learning algorithms is often dependent upon the size and distribution of classes in the training data. In our case, the limitation of relevant data led to challenges in capturing the underlying patterns. This lack of diversity may have affected the model's ability to generalize and predict.

To address this limitation, transfer learning was attempted as a technique. It's important to acknowledge that these methods come with their own set of assumptions and potential biases. As the SWELL-KW model was trained while resting or working, it has a significantly lower heart rate than the training sessions of OLT dataset.

As briefly mentioned in Section 4.3.5, using a more balanced training set will give better results. As most sessions had an average score it would be difficult for the model to find patterns for the less frequent scores. Generating a larger and more balanced dataset will help the possibility of predicting correctly.

5.0.2 Model Accuracy and Performance

The accuracy of our machine learning model is an essential metric in evaluating its effectiveness. During the evaluation phase, the model demonstrated terrible accuracy on the test dataset, where the model was predicting the same score for all sessions. Later on, when just using sessions on each side of the scale, then score 1 vs score 5, the model achieved results that aligned well with our expectations. However, it's crucial to recognize that these results are not desirable as they do not use or predict the scores in the middle, even though it can give hope that the prediction is possible if the model has more data to train and find patterns. Accuracy alone might not provide a comprehensive picture of the model's performance.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

This work aimed to investigate methods for predicting athlete's subjective feelings and evaluate their effectiveness for potential use in a future system.

The experiments conducted in this study focused on using transfer learning from an existing dataset to the provided dataset from OLT. Multiple attempts were made using both binary and three-class classifications. The CNN transfer learning model provided less than desirable results. It was discovered that using a more balanced and bigger dataset would most probably be able to correctly predict classes.

While improvements and optimizations are still needed, the findings support the feasibility of correctly predicting athletes' feelings. The results have also highlighted the importance of investigating different types of machine learning, as some methods will work better than others. The results showcase the importance of transfer learning for smaller datasets.

6.2 Future Work

The end goal of this thesis is to evaluate methods for predicting athletes' subjective feelings after a workout session.

Possible future work will be to add additional features to the model and transfer learning, such as DFA Alpha 1. However, adding this feature would require more time or computational processing, as the calculation of this feature can take a while for each session. Other

features could also be added if they can help the model predict more accurate results.

For future work, the dataset could expand and add more athletes. A larger dataset is not only desirable but adding more athletes can give the model more insight into the feature importance of how people determine their subjective feeling.

For future work, it would also be beneficial to investigate other methods to predict subjective feelings rather than just transfer learning or other machine learning methods. One possibility is to investigate Fourier Bessel, a method in the frequency domain that can provide better approximations to HRV analysis.

Bibliography

[1] Olsen, "Overview and Estimation of Detrended Fluctuation Analysis Alpha 1 for Usage in Training," eng, NTNU, Project Thesis, May 2023, p. 20.

[2] T. Ziemssen, M. Reimann, J. Gasch, and H. Rüdiger, "Trigonometric regressive spectral analysis: An innovative tool for evaluating the autonomic nervous system," en, *Journal of Neural Transmission*, vol. 120, no. 1, pp. 27–33, Sep. 2013, ISSN: 1435-1463. DOI: `10. 1007 / s00702 - 013 - 1054 - 5`. [Online]. Available: `https : / / doi . org / 10 . 1007 / s00702-013-1054-5` (visited on 01/21/2024).

[3] *Wearables Aren't Dead, They're Just Shifting Focus as the Market Grows 16.9% in the Fourth Quarter, According to IDC*, en, Mar. 2017. [Online]. Available: `https : / / www . businesswire . com / news / home / 20170302005298 / en / Wearables - Arent - Dead - Theyre - Just - Shifting - Focus - as - the - Market - Grows - 16.9 - in - the - Fourth - Quarter - According - to - IDC` (visited on 01/21/2024).

[4] T. t. p. g. i. S. a. n. l. f. t. i. g. b. c. o. c. D. t. v. u. cycles and S. C. D. M. D. D. T. R. i. t. Text, *Topic: Smartwatches*, en. [Online]. Available: `https : // www . statista . com / topics / 4762/smartwatches/` (visited on 01/21/2024).

[5] F. De Arriba-Pérez, M. Caeiro-Rodríguez, and J. M. Santos-Gago, "Collection and Processing of Data from Wrist Wearable Devices in Heterogeneous and Multiple-User Scenarios," en, *Sensors*, vol. 16, no. 9, p. 1538, Sep. 2016, Number: 9 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 1424-8220. DOI: `10.3390/s16091538`. [Online]. Available: `https : // www . mdpi . com / 1424 - 8220 / 16 / 9 / 1538` (visited on 01/21/2024).

[6] R. Jerath, M. Syam, and S. Ahmed, "The Future of Stress Management: Integration of Smartwatches and HRV Technology," en, *Sensors*, vol. 23, no. 17, p. 7314, Jan. 2023, Number: 17 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 1424-8220. DOI: `10 . 3390 / s23177314`. [Online]. Available: `https : / / www . mdpi . com / 1424 - 8220/23/17/7314` (visited on 01/21/2024).

[7] *Structure of the Heart | SEER Training*. [Online]. Available: `https://training.seer.cancer.gov/anatomy/cardiovascular/heart/structure.html` (visited on 02/07/2024).

[8] *Heart Anatomy*, en-US. [Online]. Available: `https://www.texasheart.org/heart-health/heart-information-center/topics/heart-anatomy/` (visited on 02/07/2024).

[9] *Your Heart: Powerhouse of the Body*, en. [Online]. Available: `https://my.clevelandclinic.org/health/body/21704-heart` (visited on 02/07/2024).

[10] J. D. Pollock and A. N. Makaryus, "Physiology, Cardiac Cycle," eng, in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2024. [Online]. Available: `http://www.ncbi.nlm.nih.gov/books/NBK459327/` (visited on 02/07/2024).

[11] P. T. Paharia, *What is the Nervous System?* en, Section: Health Articles, Feb. 2010. [Online]. Available: `https://www.news-medical.net/health/What-is-the-Nervous-System.aspx` (visited on 02/07/2024).

[12] *Organization of the Nervous System | SEER Training*. [Online]. Available: `https://training.seer.cancer.gov/anatomy/nervous/organization/` (visited on 02/07/2024).

[13] *Electrocardiogram (ECG or EKG) - Mayo Clinic*. [Online]. Available: `https://www.mayoclinic.org/tests-procedures/ekg/about/pac-20384983` (visited on 05/04/2023).

[14] Y. Sattar and L. Chhabra, "Electrocardiogram," eng, in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2023. [Online]. Available: `http://www.ncbi.nlm.nih.gov/books/NBK549803/` (visited on 05/04/2023).

[15] *Electrocardiogram (ECG or EKG)*, en. [Online]. Available: `https://www.heart.org/en/health-topics/heart-attack/diagnosing-a-heart-attack/electrocardiogram-ecg-or-ekg` (visited on 05/04/2023).

[16] L. Ortiz Martin, P. Picazo-Sanchez, P. Peris-Lopez, and J. Tapiador, "Heartbeats Do Not Make Good Pseudo-Random Number Generators: An Analysis of the Randomness of Inter-Pulse Intervals," *Entropy*, vol. 20, p. 94, Jan. 2018. DOI: `10.3390/e20020094`.

[17] A. Abdou and S. Krishnan, "Horizons in Single-Lead ECG Analysis From Devices to Data," *Frontiers in Signal Processing*, vol. 2, 2022, ISSN: 2673-8198. [Online]. Available: `https://www.frontiersin.org/articles/10.3389/frsip.2022.866047` (visited on 05/11/2023).

[18] M. P. Witvliet, E. P. M. Karregat, J. C. L. Himmelreich, J. S. S. G. de Jong, W. A. M. Lucassen, and R. E. Harskamp, "Usefulness, pitfalls and interpretation of handheld single-lead electrocardiograms," en, *Journal of Electrocardiology*, vol. 66, pp. 33–37, May 2021, ISSN: 0022-0736. DOI: `10.1016/j.jelectrocard.2021.02.011`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S0022073621000455` (visited on 05/11/2023).

[19] J. Francis, "ECG monitoring leads and special leads," *Indian Pacing and Electrophysiology Journal*, vol. 16, no. 3, pp. 92–95, Jul. 2016, ISSN: 0972-6292. DOI: `10.1016/j.ipej.2016.07.003`. [Online]. Available: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5067828/` (visited on 12/30/2023).

[20] M. Malik and A. J. Camm, "Heart rate variability," en, *Clinical Cardiology*, vol. 13, no. 8, pp. 570–576, 1990, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/clc.4960130811, ISSN: 1932-8737. DOI: `10.1002/clc.4960130811`. [Online]. Available: `https://onlinelibrary.wiley.com/doi/abs/10.1002/clc.4960130811` (visited on 05/06/2023).

[21] N. Singh, K. J. Moneghetti, J. W. Christle, D. Hadley, D. Plews, and V. Froelicher, "Heart Rate Variability: An Old Metric with New Meaning in the Era of using mHealth Technologies for Health and Exercise Training Guidance. Part One: Physiology and Methods," *Arrhythmia & Electrophysiology Review*, vol. 7, no. 3, pp. 193–198, Aug. 2018, ISSN: 2050-3369. DOI: `10.15420/aer.2018.27.2`. [Online]. Available: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6141929/` (visited on 05/04/2023).

[22] *RR Intervals, Heart Rate, and HRV Howto*. [Online]. Available: `https://archive.physionet.org/tutorials/hrv/` (visited on 01/03/2024).

[23] J.-G. Dong, "The role of heart rate variability in sports physiology (Review)," *Experimental and Therapeutic Medicine*, vol. 11, no. 5, pp. 1531–1536, May 2016, Publisher: Spandidos Publications, ISSN: 1792-0981. DOI: `10.3892/etm.2016.3104`. [Online]. Available: `https://www.spandidos-publications.com/10.3892/etm.2016.3104` (visited on 05/06/2023).

[24] *DFA alpha 1 HRV based (an)aerobic threshold estimation*, en. [Online]. Available: `https://aiendurance.com/blog/dfa-alpha-1-thresholds-from-heart-rate-variability` (visited on 03/29/2023).

[25] *DFA a1 and exercise intensity FAQ*, en. [Online]. Available: `http://www.muscleoxygentraining.com/2021/01/dfa-a1-and-exercise-intensity-faq.html` (visited on 03/31/2023).

[26] *Time Series Analysis: Definition, Types & Techniques | Tableau*, en-US. [Online]. Available: `https://www.tableau.com/learn/articles/time-series-analysis` (visited on 02/07/2024).

[27] E. Alpaydin, *Machine learning: the new AI*. MIT press, 2016.

[28] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[29] *Support Vector Machine (SVM) Explained*, en. [Online]. Available: `https://se.mathworks.com/discovery/support-vector-machine.html` (visited on 12/16/2023).

[30]   R. Gandhi, *Support Vector Machine — Introduction to Machine Learning Algorithms*, en, Jul. 2018. [Online]. Available: `https : / / towardsdatascience . com / support - vector-machine-introduction-to-machine-learning-algorithms-934a444fca47` (visited on 12/16/2023).

[31]   K. P. Thanaraj, B. Parvathavarthini, U. J. Tanik, V. Rajinikanth, S. Kadry, and K. Kamalanand, "Implementation of deep neural networks to classify eeg signals using gramian angular summation field for epilepsy diagnosis," *arXiv preprint arXiv:2003.04534*, 2020.

[32]   B. J. Wythoff, "Backpropagation neural networks: A tutorial," *Chemometrics and Intelligent Laboratory Systems*, vol. 18, no. 2, pp. 115–155, 1993.

[33]   R. Hecht-Nielsen, "Theory of the backpropagation neural network," in *Neural networks for perception*, Elsevier, 1992, pp. 65–93.

[34]   J. E. Dayhoff and J. M. DeLeo, "Artificial neural networks: Opening the black box," *Cancer: Interdisciplinary International Journal of the American Cancer Society*, vol. 91, no. S8, pp. 1615–1635, 2001.

[35]   J. Teuwen and N. Moriakov, "Convolutional neural networks," in *Handbook of medical image computing and computer assisted intervention*, Elsevier, 2020, pp. 481–501.

[36]   S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 international conference on engineering and technology (ICET)*, Ieee, 2017, pp. 1–6.

[37]   R. DiPietro and G. D. Hager, "Chapter 21 - Deep learning: RNNs and LSTM," in *Handbook of Medical Image Computing and Computer Assisted Intervention*, ser. The Elsevier and MICCAI Society Book Series, S. K. Zhou, D. Rueckert, and G. Fichtinger, Eds., Academic Press, Jan. 2020, pp. 503–519, ISBN: 978-0-12-816176-0. DOI: `10.1016/B978-0-12-816176-0.00026-0`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/B9780128161760000260` (visited on 02/07/2024).

[38]   M. Saeed, *An Introduction to Recurrent Neural Networks and the Math That Powers Them*, en-US, Sep. 2022. [Online]. Available: `https : / / machinelearningmastery . com/an-introduction-to-recurrent-neural-networks-and-the-math-that-powers-them/` (visited on 02/07/2024).

[39]   Z. A. Khan, T. Hussain, A. Ullah, S. Rho, M. Lee, and S. W. Baik, "Towards Efficient Electricity Forecasting in Residential and Commercial Buildings: A Novel Hybrid CNN with a LSTM-AE based Framework," en, *Sensors*, vol. 20, no. 5, p. 1399, Jan. 2020, Number: 5 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 1424-8220. DOI: `10.3390/s20051399`. [Online]. Available: `https://www.mdpi.com/1424-8220/20/5/1399` (visited on 02/04/2024).

[40] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," *Physica D: Nonlinear Phenomena*, vol. 404, p. 132 306, Mar. 2020, ISSN: 0167-2789. DOI: 10.1016/j.physd.2019.132306. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167278919305974 (visited on 02/07/2024).

[41] J. A. L. Marques, F. N. B. Gois, J. P. d. V. Madeiro, T. Li, and S. J. Fong, "Chapter 4 - Artificial neural network-based approaches for computer-aided disease diagnosis and treatment," in *Cognitive and Soft Computing Techniques for the Analysis of Healthcare Data*, ser. Intelligent Data-Centric Systems, A. K. Bhoi, V. H. C. de Albuquerque, P. N. Srinivasu, and G. Marques, Eds., Academic Press, Jan. 2022, pp. 79–99, ISBN: 978-0-323-85751-2. DOI: 10.1016/B978-0-323-85751-2.00008-6. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780323857512000086 (visited on 02/04/2024).

[42] J. Brownlee, *A Gentle Introduction to Transfer Learning for Deep Learning*, en-US, Dec. 2017. [Online]. Available: https://machinelearningmastery.com/transfer-learning-for-deep-learning/ (visited on 02/04/2024).

[43] Z. Zhao, L. Alzubaidi, J. Zhang, Y. Duan, and Y. Gu, "A comparison review of transfer learning and self-supervised learning: Definitions, applications, advantages and limitations," *Expert Systems with Applications*, vol. 242, p. 122 807, May 2024, ISSN: 0957-4174. DOI: 10.1016/j.eswa.2023.122807. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417423033092 (visited on 02/04/2024).

[44] P. Virtanen, R. Gommers, T. E. Oliphant, *et al.*, "SciPy 1.0: Fundamental algorithms for scientific computing in Python," en, *Nature Methods*, vol. 17, no. 3, pp. 261–272, Mar. 2020, Number: 3 Publisher: Nature Publishing Group, ISSN: 1548-7105. DOI: 10.1038/s41592-019-0686-2. [Online]. Available: https://www.nature.com/articles/s41592-019-0686-2 (visited on 12/18/2023).

[45] C. R. Harris, K. J. Millman, S. J. van der Walt, *et al.*, "Array programming with NumPy," en, *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020, Number: 7825 Publisher: Nature Publishing Group, ISSN: 1476-4687. DOI: 10.1038/s41586-020-2649-2. [Online]. Available: https://www.nature.com/articles/s41586-020-2649-2 (visited on 12/18/2023).

[46] *Highlights — pyHRV - OpenSource Python Toolbox for Heart Rate Variability 0.4 documentation*. [Online]. Available: https://pyhrv.readthedocs.io/en/latest/index.html (visited on 01/21/2024).

[47]   *Getting started — hrvanalysis 1.0.0 documentation.* [Online]. Available: `https://aura-healthcare.github.io/hrv-analysis/readme.html` (visited on 01/21/2024).

[48]   J. D. Hunter, "Matplotlib: A 2D graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007, Publisher: IEEE COMPUTER SOC. DOI: `10.1109/MCSE.2007.55`.

[49]   *TrainingPeaks | Train With Confidence.* [Online]. Available: `https://www.trainingpeaks.com/` (visited on 01/21/2024).

[50]   *Python-fitparse Documention — python-fitparse 0.0.1-dev documentation.* [Online]. Available: `https://pythonhosted.org/fitparse/` (visited on 01/21/2024).

[51]   *Overview and SWELL movie.* [Online]. Available: `https://www.swell-project.net/about-swell.html` (visited on 12/17/2023).

[52]   "SWELL-Knowledge Worker Dataset," *Radboud University*, [Online]. Available: `https://cs.ru.nl/~skoldijk/SWELL-KW/Dataset.html` (visited on 12/17/2023).

[53]   M. Vollmer, "A Robust, Simple and Reliable Measure of Heart Rate Variability using Relative RR Intervals," Sep. 2015. DOI: `10.1109/CIC.2015.7410984`.

[54]   M. Malik, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *Circulation*, vol. 93, pp. 1043–1065, Mar. 1996.

[55]   *ACM-Research/stress-diagnosis-neural-networks*, original-date: 2022-01-03T20:19:03Z, Apr. 2022. [Online]. Available: `https://github.com/ACM-Research/stress-diagnosis-neural-networks` (visited on 02/07/2024).

Appendix

A   Declaration about usage of artificial intelligence

NTNU | Fakultet for informasjons-
teknologi og elektroteknikk

Deklarasjon om KI-hjelpemidler

Har det i utarbeidingen av denne rapporten blitt anvendt KI-baserte hjelpemidler?

☐ Nei

☒ Ja

Hvis *ja*: spesifiser type av verktøy og bruksområde under.

Tekst

☒ **Stavekontroll**. Er deler av teksten kontrollert av:
Grammarly, Ginger, Grammarbot, LanguageTool, ProWritingAid, Sapling, Trinka.ai eller lignende verktøy?

☐ **Tekstgenerering**. Er deler av teksten generert av:
ChatGPT, GrammarlyGO, Copy.AI, WordAi, WriteSonic, Jasper, Simplified, Rytr eller lignende verktøy?

☐ **Skriveassistanse**. Er en eller flere av ideene eller fremgangsmåtene i oppgaven foreslått av:
ChatGPT, Google Bard, Bing chat, YouChat eller lignende verktøy?

Hvis *ja* til anvendelse av et tekstverktøy - spesifiser bruken her:

Tekst kontrollert av Grammarly

Kode og algoritmer

☐ **Programmeringsassistanse**. Er deler av koden/algoritmene som i) fremtrer direkte i rapporten eller ii) har blitt anvendt for produksjon av resultater slik som figurer, tabeller eller tallverdier blitt generert av: *GitHub Copilot, CodeGPT, Google Codey/Studio Bot, Replit Ghostwriter, Amazon CodeWhisperer, GPT Engineer, ChatGPT, Google Bard* eller lignende verktøy?

Hvis *ja t*il anvendelse av et programmeringsverktøy - spesifiser bruken her:

Bilder og figurer

☐ **Bildegenerering**. Er ett eller flere av bildene/figurene i rapporten blitt generert av:
Midjourney, Jasper, WriteSonic, Stability AI, Dall-E eller lignende verktøy?

Hvis *ja* til anvendelse av et bildeverktøy - spesifiser bruken her:

☐ **Andre KI verktøy**. har andre typer av verktøy blitt anvendt? Hvis ja spesifiser bruken her:

☒ Jeg er kjent med NTNUs regelverk: *Det er ikke tillatt å generere besvarelse ved hjelp av kunstig intelligens og levere den helt eller delvis som egen besvarelse.* Jeg har derfor redegjort for all anvendelse av kunstig intelligens enten i) direkte i rapporten eller ii) i dette skjemaet.

_____
Underskrift/Dato/Sted