

# Assessing the physical capabilities of sportsman using wearables and machine learning

De Clerck Jeroen

Master's dissertation submitted in order to obtain the academic degree of Master  
of Science in de informatica

Supervisor:

Councillors: Prof. dr. ir Toon De Pessemier, Kris Vanhecke

Academic year 2019-2020

# Assessing the physical capabilities of sportsman using wearables and machine learning

Jeroen De Clerck

Student number: 01206141

Supervisor:

Counsellors: Prof. dr. ir. Toon De Pessemier, Kris Vanhecke

Master's dissertation submitted in order to obtain the academic degree of  
Master of Science in de informatica

Academic year 2019-2020

## Abstract (Dutch)

### METHODE

Het doel van deze masterthesis is om Machine Learning te gebruiken voor een digitale personal coach te ontwikkelen. 2 onderdelen van deze coach zijn geïsoleerd: een model dat de huidige activiteit van een gebruiker kan classificeren en een model dat de cardiovasculaire fitness van de gebruiker kan meten. De belangrijkste metriek gebruikt is hartslagvariabiliteit. HSV is ruim onderzocht in de medische sector in de context van hartziektes en hartaandoeningen. Deze thesis poogt deze bestaande research te hergebruiken en toe te passen op cardiovasculaire fitness en machine learning.

Gezonde jongvolwassenen werden gevraagd om een fitness-regime te fietsen op een hometrainer, hun hartslage werd gemeten en een aantal meta-data punten werden genoteerd. In totaal bestaat de dataset uit 23 testpersonen, voor een totaal van ~95000 hartslagen.

### RESULTATEN

De volgende netwerken werden getest voor de classificatie van de huidige activiteit-staat: LSTM (~58% accuraatheid), Deep LSTM (~65% accuraatheid), DeepHeart (~62% accuraatheid), DeepHeartV2 (~75% accuraatheid), TCN (~85% accuraatheid). De dataset voor het trainen van het fitheid-niveau model bleek te klein te zijn voor een resultaat te bereiken.

### CONCLUSIE

De voorgestelde netwerken slaagden er in om tot ~85% accuraatheid te bereiken voor de activiteit-staat classification. Dit is een redelijk positief resultaat, maar nog niet toepasbaar voor een matuur product.

Echter, hoewel het onmogelijk is om de validiteit van de gebruikte methoden volledig aan te tonen totdat meer data verzameld is, beide onderdelen werden onderzocht en gevalideerd met behulp van vaak-geciteerde en collegiaal getoetste studies.

## Abstract (English)

### METHOD

The goal of this master thesis is to use Machine Learning to develop a digital personal coach for sport activities. 2 subgoals have been isolated: A model capable of classifying the activity state of the user and a model to gauge the cardiovascular fitness level of the user. The primary measurement used is Heart Rate Variability. HRV has been a subject of research in the medical community and can be used to predict various cardiovascular diseases or conditions. This thesis attempts to reappropriate the existing HRV research for cardiovascular fitness and machine learning.

Healthy young adults were asked to perform a specific fitness training session on a stationary bicycle, their heart rate response was measured and various metadatapoints were noted. In total 23 test subjects were added to the data set, resulting in ~95.000 timesteps of heartbeat data.

### RESULTS

The following models were tested for activity state classification: LSTM (~58% accuracy), Deep LSTM (~65% accuracy), DeepHeart (~62% accuracy), DeepHeartV2 (~75% accuracy), TCN (~85% accuracy). The number of participants in the study was too little to develop the base-level fitness regression.

### CONCLUSION

The proposed networks managed to achieve up to ~85% accuracy on activity state classification. This is a fairly positive result, but not yet applicable for real-world scenarios.

However, while it is impossible to fully prove the validity of the applied methods until more data is gathered, both subgoals were explored and validated using heavily cited, peer-reviewed studies.

## Extended abstract

### BACKGROUND

People who practice sports occasionally, often have no or very limited support and guidance. People can download a fitness scheme or app from the internet, but these schemes are rarely personalized to the physical capabilities of that person. Moreover, these schemes are usually static and do not take into account the real progress that a user has made or personal goals. Nowadays, the technological advances allow a more accurate monitoring of the user. Smart watches have become more and more popular. Heart rate sensors are becoming cheap and commonplace. Computational power of wearables and smartphones is increasing, which makes running complex tasks such as light-weight machine learning more of a reality.

### METHOD

The goal of this master thesis is to research the possibility of using Machine Learning to develop a digital personal coach for sport activities. 5 Subgoals were determined to be required to achieve this goal, 2 of which have been isolated for development.

- A model capable of classifying the state of the user. Are they resting, active, recovering, under intense exertion, ...
- A model to gauge the cardiovascular fitness level of the user.

The primary measurement used is Heart Rate Variability. This is the minuscule variability in time between heartbeats and is the result of a multitude of complex interactions between various hormonal and nervous systems of the body. HRV has been a subject of research in the medical community and can be used to predict various cardiovascular diseases or conditions. Current generation heart rate monitors are able to monitor HRV to a respectable degree compared to medical grade electrocardiogram machines, so the timing is right to attempt to reappropriate the existing HRV research for cardiovascular fitness and machine learning.

In this master thesis the Polar H7 (a HRV-capable chest strap) is used to measure heart rate data from test subjects. Healthy young adults were asked to perform

a specific fitness training session on a Kettler ergometer x3, their heart rate response was measured and various meta-datas were noted. A Polar plug-in adapter was used to synchronise the measurements between the chest strap and the stationary bicycle. In total 27 test subjects were gathered, of which 23 were able to fully complete the session, resulting in ~95.000 timesteps of heartbeat data.

From the heart rate measurements extra features were engineered based on the established relevancy to physical fitness. These features were then labeled and used as training data for the heart rate classifier. For the base-level fitness regression, a formula is established to give a “fitness score” based on results of the test subject. With some post-processing the heart rate measurements are reduced to a fixed amount of datapoints as input. Multiple deep neural networks are tested and evaluated for the classifier and regression tasks.

## RESULTS

For the activity state classification task, the proposed networks stem from either natural language processing or earlier experimentations involving machine learning and heart rate data. The following accuracies were achieved:

Model	Accuracy
LSTM	~58%
Deep LSTM	~65%
DeepHeart	~62%
DeepHeartV2	~75%
TCN	~85%

The number of participants in the study was too little to develop the base-level fitness regression. Additional test subjects would be required for a adequate conclusion.

## CONCLUSION

The proposed networks managed to achieve up to ~85% accuracy on activity state classification. This is a fairly positive result, but not yet applicable for real-world scenarios.

However, while it is impossible to fully prove the validity of the applied methods until more data is gathered, both subgoals were explored and validated using heavily cited, peer-reviewed studies.

## Verlengd abstract

### ACHTERGROND

Personen die af en toe sporten, krijgen vaak weinig tot geen ondersteuning en begeleiding. Deze personen kunnen een fitness-schema of app downloaden van het internet, maar deze schemas zijn zelden gepersonaliseerd op basis van de fysieke capabiliteiten van de gebruiker. Daarbovenop zijn deze schemas vaak statisch en houden ze geen rekening met de vooruitgang die de gebruiker boekt of de persoonlijke doelen van de gebruiker. Hedendaagse technologische vooruitgang heeft er voor gezorgd dat accurate meting van de gebruiker gemakkelijk en bereikbaar zijn. Hartslagsensoren worden goedkoper en wijder gebruikt, de rekenkracht van deze draagbare toestellen en smartphones neemt gestaag toe, wat het steeds realistischer maakt om complexe taken (zoals simpele machine learning taken) hierop uit te voeren.

### METHODE

Het doel van deze masterthesis is om Machine Learning te gebruiken voor een digitale personal coach te ontwikkelen. 5 onderdelen werden geïsoleerd, waarvan er uiteindelijk 2 uitgewerkt zijn:

- een model dat de huidige activiteit van een gebruiker kan classificeren. Is de gebruiker in rust, actief, aan het recupereren, onder intense inspanning, ...?
- een model dat de cardiovasculaire fitness van de gebruiker kan meten.

De belangrijkste metriek gebruikt is hartslagvariabiliteit. Dit is de minuscule variabiliteit in tijd tussen hartslagen en is het resultaat van een groot aantal complexe interacties tussen verschillende hormonale systemen en zenuwstelsels. HSV is ruim onderzocht in de medische sector in de context van hartziektes en hartaandoeningen. Nieuwe generatie hartslagmonitoren zijn in staat om deze HSV te meten op een niveau dat medische electrocardiogram-apparatuur bijna evenaart. Het is de opportune tijd om deze bestaande research pogingen te hergebruiken en toe te passen op cardiovasculaire fitness en machine learning.

In de thesis wordt de Polar H7 (een HSV-capabele borstband hartslagmonitor) gebruikt om hartslag data van proefpersonen te meten. Gezonde jongvolwassenen werden gevraagd om een fitness-regime te fietsen op een Kettler Ergometer

X3 hometrainer, hun hartslagrespons werd gemeten en een aantal meta-data punten werden genoteerd. Een Polar plugin adapter werd gebruikt om deze data te synchroniseren tussen de borstband en de hometrainer. In totaal werden 27 proefpersonen verzameld, waarvan 23 in staat waren om het volledige regime te voltooien, met ca. 95.000 hartslagen aan data als resultaat.

Van deze hartslagdata werden een aantal features verwerkt, gebaseerd op de gevestigde relevante metrieken voor fysieke fitness. Deze features werden vervolgens gelabeled en gebruikt als trainings data voor de hartslag classificeerder. Voor de fitness-regressie werd een formule ontwikkeld om een “fitness score” toe te kennen aan de proefpersonen op basis van hun resultaten van het regime. Met een kleine hoeveelheid post-processing werd de hartslagdata van een regime verwerkt tot een vaste hoeveelheid datapunten die kan dienen als input. Verschillende netwerken worden getest en geëvalueerd voor beide taken.

## RESULTATEN

Voor de hartslag classificeerder werden netwerken gebruikt die ontwikkeld zijn voor Natural Language Processing, of netwerken die in eerdere experimenten m.b.t. hartslag data gebruikt zijn. De volgende accuraatheid werd behaald:

Model	Accuraatheid
LSTM	~58%
Deep LSTM	~65%
DeepHeart	~62%
DeepHeartV2	~75%
TCN	~85%

De dataset voor het trainen van het fitheid-niveau model bleek te klein te zijn voor een resultaat te bereiken. Meer proefpersonen zouden nodig zijn om tot een conclusie te komen voor deze taak.

## CONCLUSIE

De voorgestelde netwerken slaagden er in om tot ~85% accuraatheid te bereiken voor de activiteit-staat classification. Dit is een redelijk positief resultaat, maar nog niet toepasbaar voor een matuur product.

Echter, hoewel het onmogelijk is om de validiteit van de gebruikte methoden volledig aan te tonen totdat meer data verzameld is, beide onderdelen werden onderzocht en gevalideerd met behulp van vaak-geciteerde en collegiaal getoetste studies.

## Lay summary

### BACKGROUND

People who practice sports occasionally, often have no or very limited support and guidance. People can download a fitness scheme or app from the internet, but these schemes are rarely personalized to the physical capabilities of that person. Moreover, these schemes are usually the same for every participant and do not take into account the real progress that a user has made or personal goals. Nowadays, the technological advances allow a more accurate monitoring of the user. Smart watches have become more and more popular. Heart rate sensors are becoming cheap and commonplace. Processing power of these smartwatches and smartphones is increasing, which makes running complex tasks such as light-weight machine learning more of a reality.

### METHOD

The goal of this master thesis is to research the possibility of using AI to develop a digital personal coach for sport activities. Of the components determined to be required to achieve this goal, 2 were researched in this document.

- Software capable of determining how active the user is. Are they resting, active, recovering, under intense exertion, ...
- Software capable of calculating how fit the user is and giving them a “fitness score”.

The primary measurement used is Heart Rate Variability. This is the minuscule difference in time between heartbeats. HRV has been a subject of research in the medical field and can be used to predict various heart diseases or conditions. Current generation heart rate monitors are able to monitor HRV to a respectable degree compared to medical grade monitors, so the timing is right to attempt to use the existing HRV research for calculating fitness in the human body using machine learning.

In this master thesis the Polar H7 (a HRV-capable chest strap) is used to measure heart rate data from test subjects. Healthy young adults were asked to perform a specific fitness training session on a Kettler ergometer x3, their heart rate response was measured and various extra information about the subject was noted. In total

27 test subjects were gathered, of which 23 were able to fully complete the session, resulting in ~95.000 heartbeats data.

## ON MACHINE LEARNING

Machine learning is a type of artificial intelligence where a stack of mathematical “neuron” layers is made to learn by example. Each of these layers is nothing more than an amount of cells, where each cell is able to receive various inputs, perform a calculation, and pass the result of this calculation on to one or more neurons in the next layer. The sequence of layers are called a “network” and are structured so that the output of the very last layer is a class (in the case of classifying networks) or a score (in case of regression networks). There are more possible use-cases of networks but the thesis is limited to these two.

If a network is designed to find the difference between a dog and a cat, we can show it 100 cat pictures and 100 dog pictures. The algorithm does not know which ones are which, but after letting the network decide on its own we can find out which pictures were predicted correctly and which weren't. Through a concept called “backpropagation” we are capable of showing the network where it went wrong so it can adjust its network to better predict this so-called “training data”. By repeating this training session over and over with massive amounts of data, the network will eventually be able to accurately classify pictures as “Dog” or “Cat”, even pictures that it has never seen before.

In this thesis, the input data is not “Dog” or “Cat” pictures, but instead sequences of the timing measurements of 128 heartbeats. The output is not the type of animal, but rather how active the user was during time of measurement.

In the case of the fitness scorer, the output of the AI is a number roughly corresponding to how fit a user is.

## RESULTS

For determining how active the user is, we based ourselves on existing networks, particularly ones designed and used in text recognition, or ones that were made to find heart conditions. Each model is graded on the percentage of sequences it could correctly classify:

Model	Accuracy
LSTM	~58%
Deep LSTM	~65%
DeepHeart	~62%
DeepHeartV2	~75%
TCN	~85%

The number of participants in the study was too little to create a network that can give a fitness score. Additional test subjects would be required.

## CONCLUSION

The proposed networks achieved up to ~85% accuracy. This is a fairly positive result, but not yet up to par for a real-life fitness coach.

However, while it is impossible to fully prove if the applied methods work until more data is gathered, both subgoals were heavily researched and validated using peer-reviewed studies.

# Acknowledgements

This is for acknowledging all of the people who helped out

*De auteur geeft de toelating deze masterproef voor consultatie beschikbaar te stellen en delen van de masterproef te kopiëren voor persoonlijk gebruik. Elk ander gebruik valt onder de bepalingen van het auteursrecht, in het bijzonder met betrekking tot de verplichting de bron uitdrukkelijk te vermelden bij het aanhalen van resultaten uit deze masterproef.*

# Table of Contents

Abstract (Dutch) . . . . .	i
Methode . . . . .	i
Resultaten . . . . .	i
Conclusie . . . . .	i
Abstract (English) . . . . .	ii
Method . . . . .	ii
Results . . . . .	ii
Conclusion . . . . .	ii
Extended abstract . . . . .	iii
Background . . . . .	iii
Method . . . . .	iii
Results . . . . .	iv
Conclusion . . . . .	iv
Verlengd abstract . . . . .	vi
Achtergrond . . . . .	vi
Methode . . . . .	vi
Resultaten . . . . .	vii
Conclusie . . . . .	vii
Lay summary . . . . .	ix
Background . . . . .	ix
Method . . . . .	ix
On Machine Learning . . . . .	x
Results . . . . .	x
Conclusion . . . . .	xi
<b>Acknowledgements</b>	<b>xii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Goal . . . . .	1
1.2 Scope of the thesis . . . . .	2
1.3 Why only heart rate . . . . .	3

1.4	Methodology . . . . .	3
1.4.1	Data Gathering . . . . .	3
1.4.2	Machine learning . . . . .	4
<b>2</b>	<b>Medical Background</b>	<b>6</b>
2.1	Anatomy of the heart . . . . .	6
2.2	Innervation of the heart . . . . .	7
2.3	Physiological function of the heart . . . . .	8
2.3.1	Nerval stimulation . . . . .	8
2.3.2	Ectopic Beats . . . . .	11
2.3.3	Heart Rate . . . . .	11
2.3.4	Heart Rate variability . . . . .	12
2.3.4.1	Time Domain . . . . .	13
2.3.4.2	Frequency Domain . . . . .	13
2.3.5	Heart Rate Recovery . . . . .	14
2.3.6	Cardiac Output . . . . .	15
2.4	Measuring heart rate . . . . .	15
2.4.1	Phonocardiogram (PCG) measurement . . . . .	15
2.4.2	Electrocardiogram (ECG) measurement . . . . .	16
2.4.3	Photoplethysmogram (PPG) measurement . . . . .	16
2.5	Physical fitness . . . . .	17
2.5.1	Fitness metrics . . . . .	18
2.5.1.1	$\text{VO}_{2\text{max}}$ & Aerobic capacity . . . . .	18
2.5.1.2	Baecke Score . . . . .	19
2.5.1.3	Blood Lactate levels . . . . .	19
2.5.1.4	Heart Rate variability . . . . .	19
2.5.1.5	Heart Rate Recovery . . . . .	20
2.5.2	Fitness Training . . . . .	21
2.5.2.1	Detraining . . . . .	22
2.5.2.2	Overreaching spectrum . . . . .	22
2.5.2.3	Overtraining syndrome . . . . .	23
<b>3</b>	<b>Data Gathering</b>	<b>27</b>
3.1	RR interval filtering and correction . . . . .	28
3.2	Heart Rate metrics feature engineering for the classification task . . . . .	29
3.2.1	RR interval . . . . .	30
3.2.2	Current Heart Rate . . . . .	30
3.2.3	Current Heart Rate Variability . . . . .	30
3.2.4	Maximum Heart Rate . . . . .	30

3.2.5	Resting Heart Rate . . . . .	30
3.2.6	HRV Time domain . . . . .	31
3.2.7	HRV Frequency domain . . . . .	31
3.2.8	Classification . . . . .	32
3.3	Subject Variables for the regression task . . . . .	32
3.3.1	Fitness score . . . . .	32
3.3.2	Input data . . . . .	33
<b>4</b>	<b>Activity State Classification</b>	<b>36</b>
4.1	Long Short Term Memory . . . . .	37
4.2	Deep LSTM . . . . .	38
4.3	DeepHeart . . . . .	38
4.4	DeepHeart v2 . . . . .	39
4.5	Temporal Convolutional Network . . . . .	39
4.6	Labeling methods . . . . .	40
4.7	Seq2Seq Classification . . . . .	41
<b>5</b>	<b>base-level Fitness regression</b>	<b>43</b>
<b>6</b>	<b>Conclusion</b>	<b>44</b>
6.1	Learning points . . . . .	44
6.2	Future work . . . . .	44
<b>Appendix A: Network Summaries</b>		<b>46</b>
LSTM network summary	. . . . .	46
Deep LSTM network summary	. . . . .	46
DeepHeart network summary	. . . . .	47
DeepHeartV2 network summary	. . . . .	48
TCN network summary	. . . . .	49
<b>Appendix B: Abbreviations</b>		<b>51</b>
<b>References</b>		<b>52</b>

# Chapter 1

## Introduction

### 1.1 Goal

The goal of this master thesis is to research the possibility of using Machine Learning to develop a digital personal coach for sports activities. It should be able to measure and track the performance of the user and suggest appropriate training sessions that result in an increase in cardiovascular fitness. In ideal conditions, a fully mature fitness coach should, therefore:

- Be able to determine a “base fitness level” of the user during the initialisation of the coaching program. This would be a machine learning regression task on labeled, batched sequence data to determine a certain “fitness level score”
- Be able to determine the increase or decrease of the cardiovascular fitness of the user through unlabeled monitoring. This would be a machine learning regression task on unlabeled, batched sequence data to update the previously acquired “fitness level score”
- Permanently track the user and classify their current “activity state”, optionally be able to automatically count periods of increased activity as a workout session. This would be a machine learning multi-label classification task on unlabeled, real-time sequence data to determine the current exertion level on a short-term timescale (30-60 seconds).
- Be able to track a user during their training session and provide feedback to keep the user in the optimal exertion range. This is a machine learning multi-

label classification task on unlabeled, real-time sequence data to determine the current exertion level on a very short-term timescale (5-10 seconds).

- Compile the result of a workout session to gauge the intensity and the effect on the body of the user. At this point, it should be able to determine the fatigue state of the user so that the coach can take this into account. This would be a machine learning classification task on labeled, batched sequence data to determine a fatigue state.
- Suggest the user what kind of workout to perform next and how long to recover from recent workouts. This would be a recommender system task that can accurately predict so-called overreaching in the user.

## 1.2 Scope of the thesis

To limit the scope of this master thesis, only 3 of these goals will be researched in-depth, 2 of which have been combined into 1 multipurpose model.

- A model capable of classifying the state of the user. Are they resting, active, recovering, under intense exertion, ...
- A model to gauge the cardiovascular fitness level of the user. Re-applying this model during different stages of the coaching period would be able to gauge the increase and decrease of the fitness level.

The primary measurement used is Heart Rate Variability (HRV). This is the minute variability in the time between heartbeats and is the result of a multitude of complex interactions between various hormonal and nervous systems of the body. HRV has been a subject of research in the medical community and can be used to predict various cardiovascular diseases or conditions. Current generation heart rate monitors can monitor HRV to a respectable degree compared to medical-grade electrocardiogram machines, so the timing is right to attempt to reappropriate the existing HRV research for cardiovascular fitness and machine learning. This however excludes the increasingly popular “smartwatches” that monitor heartrate using the less accurate light-based PPG method, as this method is incapable of measuring HRV.

### 1.3 Why only heart rate

While modern smartphones are almost always equipped with various sensors that can facilitate data-gathering (such as accelerometers, pedometers, GPS-trackers, ...), this thesis will focus on heart rate and its variability only. HRV has been a subject of research in the medical community and has been shown to be a marker of conditions such as epilepsy, Parkinson, multiple sclerosis (Cygankiewicz & Zareba 2013), stress resilience (Dong et al. 2018), and even sudden death from cardiac causes (Bassan et al. 2005).

Closer related to the subject of a fitness coach, it has been shown that habitual aerobic exercise plays a role in maintaining augmented HRV (De Meersman 1993), that HRV is likely to be sensitive to training adaptation (Buchheit 2014) and that HRV might be a better measure of cardiovascular fitness than post-exercise Heart Rate Recovery (Buchheit & Gindre 2006).

Limiting the subject to Heart Rate Variability reduces complexity and introduces less unknown variables, which should result in a clear conclusion of whether HRV is a suitable subject for further research and continued development.

### 1.4 Methodology

#### 1.4.1 DATA GATHERING

Participants were asked to strap on a Polar H7 HRV-capable chest heart rate monitor and perform a specific fitness training session on a Kettler Ergometer X3. Beat-to-beat intervals were measured and stored using the Elite HRV Android app, which can export these values into text files where each line represents the milliseconds between 2 beats.

The Kettler Ergometer X3 is equipped with an electromagnetic braking system that adapts to the cycling strength of the user. Brake horsepower (BHP) is determined by output torque and rotational speed, so the Ergometer measures the rotational speed of the user and provides adequate braking power so that the user is forced to provide the preset BHP to keep the rotational speed constant. This BHP is provided in Wattage (Joule per second).

The Ergometer is also able to count the energy spent during the various stages of the session. As the power is equal to energy over time, determining energy is

simply a matter of multiplying the power determined earlier with the time spent at this power output.

The session consisted of the following stages:

- a 2-minute stage in sedentary rest
- a 5-minute stage in which the subject provides 50 Watt (0.067 BHP) at 50 rotations per minute
- a 5-minute stage in which the subject provides 100 Watt (0.134 BHP) at 50 rpm
- a 2-minute stage of recovery in sedentary rest
- a 5-minute stage in which the subject tries to maintain 115-120 bpm Heart Rate (equals to around 60% of the theoretical maximum heart rate of young adults)
- a 2-minute stage of recovery in sedentary rest
- a 5-minute stage in which the subject tries to maintain 155-160 bpm Heart Rate (equals to around 80% of the theoretical maximum heart rate of young adults)
- a 1-minute stage in which the subject cycles at maximum exertion
- a 13-minute stage of recovery, of which 4 minutes are kept for analysis

Additionally, the following meta-data is noted:

- the Rating of Perceived Exertion (RPE) of each session
- the energy spent during the three “constant heart rate” sessions
- the distance traveled during the three “constant heart rate” sessions
- the Baecke questionnaire scores of the subject

The reasoning behind the methodology of this data gathering session is explained in-depth in chapter 3. In total 27 test subjects were gathered, of which 23 were able to fully complete the session, resulting in ~95.000 timesteps of heartbeat data.

#### 1.4.2 MACHINE LEARNING

Various sequence learning and established Natural Language Processing techniques are tested and compared for the activity state classification task:

- a standard Long Short Term Memory (LSTM) network

- a deeply layered LSTM network
- an experimental “DeepHeart” network that was successfully used to diagnose various medical conditions (Ballinger et al. 2018)
- a bespoke evolution of DeepHeart
- a Temporal Convolutional Network (TCN)

Various Convolutional and residual networks are tested for the base-level fitness regression task.

These networks are written in Python 3.8.3, created using Keras, which uses the Tensorflow 2 library to train these networks on a GTX 1070Ti CUDA-enabled graphics card.

## Chapter 2

# Medical Background

### 2.1 Anatomy of the heart

The heart is the organ responsible for pumping blood into the arteries and veins of the human body. It is located within the thorax (chest) in between the lungs, oriented slightly towards the left-hand side in its so-called mediastinum. The shape is slightly triangular, pointing towards the feet. The superior portion is called the base, and the inferior portion is called the apex.

The pumping action is provided by the 4 chambers; the left and right side each have an *atrium* and a *ventricle*. The atria act as a receiving chamber that contracts to push blood to the ventricles, which act as the primary pumping mechanism and propel it to either the lungs or the rest of the body. The atria receive blood on a nearly continuous basis, preventing flow from stopping while the ventricles are contracting.

Both sides of the heart have a distinct function. The right-hand side delivers deoxygenated blood into the pulmonary trunk, which leads toward the capillaries in the lungs where gas exchange occurs: Carbon dioxide exits the blood and oxygen enters. Highly oxygenated blood returning from these capillaries passes through a series of vessels that join together to conduct blood into the left atrium, which pumps the blood into the left ventricle, which in turn pumps oxygenated blood into the aorta and on to the many branches of the systemic circuit. Eventually, these vessels will lead to the various capillaries, where exchange with the cells of the muscles occurs. Oxygen and nutrients exit the capillaries to be used by the cells in their metabolic processes, and carbon dioxide and waste products, such as

lactic acid and ADP, will enter the blood.

Not every muscle in the heart is equal. The atria only need to pump blood into the ventricles, so they are smaller. The muscle of the left ventricle is much thicker and better developed than that of the right ventricle, to overcome the high resistance required to pump blood into the long systemic circuit. The right ventricle does not need to generate as much pressure, since the path to the lungs is shorter and provides less resistance. With an average chamber volume of 70mL and an average heart rate of 75bpm, each chamber is capable of pumping 5250 mL blood per minute.

(Betts et al. 2013)

## 2.2 Innervation of the heart

Normal cardiac rhythm is established by the sinoatrial (SA) node, also known as the pacemaker of the heart. It initiates the *sinus rhythm*, or normal electrical pattern followed by contraction of the heart. The impulse is conducted from the SA node through a specialized internodal pathway, to the atrioventricular (AV) node. In addition, the interatrial band conducts the impulse directly from the right atrium to the left atrium. The wave of the electrical impulse spreads across the atria from the top to the bottom portions, pumping blood down into the ventricle. It takes approximately 50 milliseconds to travel from SA to AV node (figure 2.2, step 2 and 3).

The next step in the chain is the atrioventricular bundle. There is a critical pause before the AV node transmits the impulse to the AV bundle, partially attributable to the small diameter of the cells of the node which slows the impulse. This results in approximately 100ms delay for the impulse to pass through the node. This pause is critical to the heart function, as it allows the atrial muscles to complete their contraction that pumps blood into the ventricles before the impulse is transmitted to the cells of the ventricle itself (figure 2.2, step 4).

The AV bundle divides into two AV bundle branches, both descending and reaching the apex of the heart where they connect with the Purkinje fibers. This passage takes approximately 25 ms. The Purkinje fibers are additional conductive fibers that spread the impulse to the muscles in the ventricles. They extend from the apex of the heart toward the wall between the atrium and the ventricle, continuing to the base of the heart. The Purkinje fibers have a fast inherent conduction

rate, and the electrical impulse reaches all the ventricular muscle cells in about 75 ms. Since the electrical stimulus begins at the apex, the contraction also begins at the apex and travels toward the base of the heart, similar to squeezing a tube of toothpaste from the bottom. This allows the blood to be pumped out of the ventricles. The total time elapsed from the initiation of the impulse in the SA node until the relaxation of the ventricles is approximately 225ms (figure 2.2, step 5).

The SA node, without external control, would initiate a heart impulse approximately 80–100 times per minute. While possible in some healthy individuals, rates lower than 50 beats per minute would indicate a condition called bradycardia. With extreme stimulation by the SA node, the AV node can transmit impulses maximally at 220 per minute. This establishes the typical maximum heart rate in a healthy young individual. Higher rates are possible in hearts that are damaged or stimulated by certain drugs, but the heart can no longer effectively pump blood at these rates.

(Betts et al. 2013)

## 2.3 Physiological function of the heart

### 2.3.1 NERVAL STIMULATION

The human body is a complex balancing act of various electrical and/or hormonal impulses.

The nervous system can be split up in a hierarchy. Starting from the full nervous system we can split 2 subsystems based on their location; the central nervous system (CNS) in the brain and spinal cord, and the peripheral nervous system (PNS) in the rest of the body. The peripheral nervous system can again be split up into 2 subsystems based on the control that we have over it: the somatic nervous system which handles voluntary signals and the autonomic nervous system (ANS) which handles involuntary signals. The autonomic nervous system can again be split up based on the context in which signals are sent: the sympathetic nervous system (SNS) which handles “fight-and-flight” reactions and the parasympathetic nervous system (PSNS) which handles “rest-and-digest” reactions.

Nervous control starts with two paired centers in the brainstem. The cardio accelerator regions stimulate activity via sympathetic stimulation, and the cardioin-

hibitory centers decrease heart activity via parasympathetic stimulation. During rest, both centers provide stimulation to the heart, with a slight predomination for cardioinhibitory stimulation as the SA node left to its own devices would initiate a sinus rhythm of approximately 100 bpm. This means that simply stopping parasympathetic stimulation would let the heart rate increase to approx. 100 bpm, but further increase requires sympathetic stimulation.

These (para-)sympathetic stimulations are fired upon various sensoric impulses, which are summed up in table 2.1 and table 2.2.

(Betts et al. 2013)

Table 2.1: Major Factors increasing Heart Rate and force of contraction. (Betts et al. 2013)

Factor	Effect Trigger
n. Cardioaccelerator	Release of norepinephrine by cardioinhibitory nerves
Proprioceptors In	Increased firing rates of positional sensors
Chemoreceptors	detection of decreased levels of O <sub>2</sub>
Chemoreceptors	detection of increased levels of H <sup>+</sup> , CO <sub>2</sub> and lactic acid
Baroreceptors	falling blood volume/pressure
Limbic system	Anticipation of physical exercise or strong emotions
Catecholamines	Increased epinephrine and norepinephrine release
Thyroid hormones	Increased T <sub>3</sub> and T <sub>4</sub> in the blood (released by thyroid)
Calcium	Increase in calcium ions in the blood
Potassium	Decrease in potassium ions in the blood
Sodium	Decrease in sodium ions in the blood
Body temperature	Increase in body temperature
Stimulants	Presence of nicotine, caffeine or other stimulants

Table 2.2: Major Factors decreasing Heart Rate and force of contraction. (Betts et al. 2013)

Factor	Effect Trigger
n. Cardioinhibitor	Release of acetylcholine by cardio accelerator nerves
Proprioceptors In	Decreased firing rates of positional sensors
Chemoreceptors	detection of increased levels of O <sub>2</sub>
Chemoreceptors	detection of decreased levels of H <sup>+</sup> , CO <sub>2</sub> and lactic acid
Baroreceptors	rising blood volume/pressure
Limbic system	Anticipation of relaxation by the limbic system
Catecholamines	Increased (nor-)epinephrine release by the adrenal glands
Thyroid hormones	Decreased T <sub>3</sub> and T <sub>4</sub> in the blood (released by thyroid)
Calcium	Increase in calcium ions in the blood
Potassium	Increase in potassium ions in the blood
Sodium	Increase in sodium ions in the blood
Body temperature	Decrease in body temperature
Tranquilizers	Presence of opiates, tranquilizers or other depressants

### 2.3.2 ECTOPIC BEATS

An ectopic or premature heartbeat is when the heart either skips a beat or adds an extra beat. These are usually not a cause for concern (in fact occur commonly in adults), and are not necessarily an indication of irregular heart functionality. Increased stress or anxiety can induce ectopic beats in healthy individuals, but these can also be caused by lack of sleep, various recreational drugs (caffeine, alcohol, smoking, ...) or medicinal drugs (asthma inhalers, antibiotics, ...)

There are two types of ectopic heartbeats:

- premature atrial contractions, which originate in the upper chambers, or atria.
- premature ventricular contractions, which originate in the lower chambers, or ventricles.

### 2.3.3 HEART RATE

The maximum heart rate that a healthy individual can attain is around 220 bpm. This maximum heart rate is a hard limit due to physiological phenomena (the maximum throughput of the SA node) and will remain constant for a subject, regardless of increased/decreased fitness level or other factors. This maximum heart rate will slowly deteriorate with age and can be approximated by taking the 220 bpm maximum and subtracting the age of the subject.

The lowest heart rate that a healthy individual can attain is called the Resting Heart Rate (RHR). This is the heart rate of the subject when it is at complete rest, relaxed but awake, with no external stress factors such as recent exercise, environmental temperature, ... For an average person, the RHR lies within the range of 60 to 100 beats per minute, but a few exceptionally trained aerobic athletes demonstrate resting heart rates in the range of 30–40 beats per minute. This is lower for fit individuals because the body has adapted to maintain basic functionality on these lower heart rates.

Performing any kind of movement will place a circulatory demand on the heart, and the heart rate will change to meet this demand. If demand is higher than what  $HR_{max}$  allows, the heart rate will plateau at  $HR_{max}$  until the body can no longer sustain the exercise. For a constant exercise that doesn't reach this boundary, the demand will also remain constant, so equilibrium will be reached at a certain heart

rate. An important distinction to make is that of the lactate threshold. Lactate is a substance in the blood that increases in concentration when performing heavy exercise.

Heart rate will quickly reach the equilibrium as described as long as blood lactate levels remain under a threshold, the value of which depends on the individual. At higher levels of exertion blood lactate levels will rise but plateau after 10-20 minutes above the resting lactate levels. This means that the heart rate equilibrium will only be reached after 10-20 minutes of sustained exercise; this is called the slow component of cardiovascular kinetics. At severe levels of exertion lactate levels will not plateau but rise steeply until the subject is too exhausted to continue, no equilibrium will be reached.

(Zakynthinaki 2015)

#### 2.3.4 HEART RATE VARIABILITY

The time between heartbeats (the RR interval or Inter Beat Interval - IBI) is not constant but fluctuates ever so slightly between beats. The difference between these intervals is not random but is the result of a multitude of complex interactions between various organs of the body and is best described as “mathematical chaos”. Two overlapping processes can be isolated (Shaffer & Ginsberg 2017):

- the interaction between parasympathetic and sympathetic branches of the ANS;
- the regulatory mechanisms that also control respiratory sinus arrhythmia (RSA) (in which the sinus firing rate increases with inspiration and decreases with expiration), baroreceptor reflex (negative-feedback control of blood pressure), and rhythmic changes in vascular tone (degree of restriction in blood vessels).

HRV diminishes with age and is a marker for parasympathetic activity, with a higher value being in general considered “more healthy”. Lower HRV values have been linked to Myocardial Infarction and Myocardial Dysfunction while raising HRV through exercise to battle cardiovascular mortality risk has been proven beneficial. (Malik et al. 1996)

HRV can be measured in various ways, which can be split into either time domain analysis or frequency domain analysis. These measurements also give different

results when different measurement epochs are used, so when studying HRV the epoch and metric must be carefully considered.

#### 2.3.4.1 Time Domain

Time Domain analysis is in function of statistical difference in time measurement between heartbeats. Various metrics exist (Shaffer & Ginsberg 2017):

- SDNN: The standard deviation of the RR intervals. 24 Hour monitoring of this value is a strong predictor of morbidity and mortality. In short-term measurement while at rest the primary source of variation can be attributed to RSA.
- Various derivatives of SDNN such as SDSD (standard deviation of the difference between successive IBI), SDRR (standard deviation of IBI without filtering abnormal beats), SDANN (standard deviation of the mean of IBI of 5-minute segments during 24hr recording), and SDNNI (mean of the standard deviation of IBI of 5-minute segments during 24hr recording).
- NN50: The number of adjacent NN intervals that differ from each other by more than 50 ms, requires a 2-minute epoch. NN Interval being all the RR intervals with “abnormal” beats filtered out. In practice (and certainly in our case) these two definitions are synonyms.
- PNN50: The proportion of NN50 over the total number of IBI, requires a 2-minute epoch and is closely correlated to Parasympathetic Nervous System activity.
- RMSSD: Root mean square of the difference between successive IBI. RMSSD is strongly backed by research and is considered the most relevant and accurate measure of ANS activity over the short-term.
- lnRMSSD: log of RMSSD which has been shown to be an acceptable measurement for short-term recordings (Esco & Flatt 2014).

#### 2.3.4.2 Frequency Domain

Various oscillations have been measured with a frequency ranging from seconds to  $>24$  hours and can be roughly grouped into different bands (see table 2.3). The LF and HF bands are significant because their oscillations can be affected by breathing rhythm. More specifically, the LF band is affected by slow breaths (3 to 9 respirations per minute) and the HF band is affected by fast breathing (9 to

24 respirations per minute).

The VLF frequency band has shown many uses for heart conditions and all-cause mortality, but very little is known of its usefulness regarding cardiovascular fitness. The LF band mainly reflects baroreceptor activity and is affected by both parasympathetic and sympathetic activity. The HF band mainly reflects RSA because it corresponds to the HRV related to the respiratory cycle. It also reflects parasympathetic activity. The ratio of LF to HF power (LF/HF ratio) is poorly understood and various interpretations have been given. The most useful interpretation to our goal is that it estimates the ratio between SNS and PSNS activity.

Table 2.3: frequency bands of Heart Rate variability.

Name	Period	Frequency
Ultra Low (ULF)	5 min - 24 hrs	11 Hz - 3333 Hz
Very Low (VLF)	25 sec - 300 sec	0.0033 Hz - 0.04 Hz
Low (LF)	7 sec - 25 sec	0.04 Hz - 0.15 Hz
High (HF) or Respiratory band	2.5sec - 7 sec	0.15 Hz - 0.4 Hz

### 2.3.5 HEART RATE RECOVERY

After cessation of exercise, the return of the heart rate towards its baseline is modulated by the ANS. The resulting speed at which the baseline is reached is called Heart Rate Recovery (HRR). As mentioned before, heart rate is influenced by the two distinct somatic sub-systems; PSNS and SNS. This balancing act has a significant effect on HRR because the impulses of these systems are delivered independently of each other. At first, the PSNS (responsible for reducing heart rate) will kick into gear resulting in a sharp decrease in HR. The SNS (responsible for increasing heart rate) reacts slower and winds down the impulses given to the heart gradually. The combination will result in a HRR curve, split into a fast phase and a slow phase. The transition between these phases denotes the border where the SNS takes over the mediation of HR from the PSNS (Bartels et al. 2018). Studies show that this curve becomes steeper as the subject transitions to adulthood, so a window of at least 2, preferably 3 minutes between measurements is recommended for achieving age-independent HRR metrics (Suzic Lazic et al. 2017) (Molina et al. 2016).

### 2.3.6 CARDIAC OUTPUT

In the context of fitness, one of the most important factors is Cardiac Output (CO). CO is the measurement of the volume of blood pumped by each ventricle per minute and is calculated by multiplying HR with Stroke Volume (SV) (the amount of blood pumped by each ventricle). It is fairly easy to conclude that an increase in CO means an increase in oxygenated blood reaching the muscles, which increases the potential maximum energy output of a muscle. Note that SV is not the same as the volume of a ventricle. There is always a certain volume of blood (the so-called End Systolic Volume or offload) present in the ventricles. During the cardiac cycle, the ventricles fill up to a maximum volume (called the End Diastolic Volume or preload) after which the SV is ejected to return to the ESV. An average heart at rest has approximately 130 mL EDV and 50-60 mL ESV, which results in an SV of 70-80 mL.

A higher heart rate will result in a higher Cardiac Output, but its not just a matter of multiplying Heart Rate with Stroke Volume. SV will initially be able to keep up with an increase in HR, but at around 120 bpm there is not enough time between contractions for the ventricles to fill up with blood. Up until 160bpm, the rate at which SV decreases is lower than the rate at which HR increases, so CO will continue to increase. As HR rises above 160 bpm, this balance shifts and CO starts to decrease. Therefore, individuals performing aerobic exercises are cautioned to monitor their HR to ensure they stay within the target heart rate range. The target HR is loosely defined as the range in which both the heart and lungs receive the maximum benefit from the aerobic workout and is dependent upon age.

(Betts et al. 2013)

## 2.4 Measuring heart rate

### 2.4.1 PHONOCARDIOGRAM (PCG) MEASUREMENT

The phonocardiograph will measure heart pulse data by listening to the sound of the beating heart. In a normal, healthy heart, there are two distinctly audible heart sounds. The first sound is created by the closing of the valves between the atrium and the ventricle during ventricular contraction. The second heart sound is the sound of the closing of the valves between the ventricle and the aortae

during ventricular relaxation (Betts et al. 2013). This method of measurement has some novel applications and is showing potential to be as accurate as optical measurement, but has not yet been brought to market as a consumer wearable (Sharma et al. 2019) (Abbasi-Kesbi et al. 2018).

#### 2.4.2 ELECTROCARDIOGRAM (ECG) MEASUREMENT

By careful placement of surface electrodes on the body, it is possible to record the complex, compound electrical signal of the heart. This tracing of the electrical signal is the electrocardiogram (ECG). There are five prominent points on the ECG (see figure 2.3): the P wave, the three points of the QRS complex, and the T wave. The small P wave represents the impulse in the atria. The atria begin contracting approximately 25 ms after the start of the P wave. The large QRS complex represents the impulse in the ventricles, which will begin to contract as the R wave reaches its peak. Lastly, the T wave represents the end of the impulse in the ventricles (the end of the impulse in the atria is obfuscated by the QRS complex). Segments are defined as the regions between two waves. Intervals include one segment plus one or more waves. In this thesis, the most notable interval is the RR interval or the interval between the peaks of the QRS complex. (Betts et al. 2013)

The standard electrocardiograph uses multiple electrodes to obtain 3, 5, or 12 leads. The greater the number of leads an electrocardiograph uses, the more information the ECG provides. The term “lead” typically describes the voltage difference between two of the electrodes. Electrocardiographs are also available in a small, portable, battery-operated device known as a Holter monitor, or simply a Holter, that continuously monitors heart electrical activity. Chest strap heart rate monitors (such as the polar H-series used here) are rudimentary electrocardiograms.

#### 2.4.3 PHOTOPLETHYSMOGRAM (PPG) MEASUREMENT

Various biological tissue types (bone, blood, pigments in the skin, ...) reflect and absorb light in different ways. The pumping action of the heart has a large effect on the pressure and corresponding presence of blood in the arteries. The peak of this pressure wave is called “systole” and the valley is called “diastole”. Note that this pulse is very strong in the arteries, being directly connected to the output of the pump. Deeper inside the system this pulse gradually weakens to static pressure

in the veins. The pulsing of blood flow and blood pressure can be determined easily using a pulse rate monitor. The principle behind PPG sensors is the optical detection of this pulse using a sensor system consisting of a light source and a detector. The sensor monitors changes in the light intensity via reflection from or transmission through the tissue. The changes in light intensity can be mapped to the diastole and systole to provide information on the pulse rate of the heart. (Tamura et al. 2014)

PPG sensors are fairly simple devices and have found their way to various wrist-based wearables, where they measure the pulse rate of the arteries in the wrist.

In HRV analysis, the RR intervals obtained from ECG measurements are required. In principle, any signal providing IBI timing data could be used, as long as the measurements are sufficiently accurate. (Pinheiro et al. 2016) suggests that data from PPG measurements could potentially be used to achieve the same goals as this thesis, but PPG is inherently less accurate than ECG. In the cited study the distinction is made between Heart Rate Variability from ECG measurement and Pulse Rate Variability from PPG measurement. Especially high-frequency measurements during periods of intense activity suffer from high inaccuracy (Menghini et al. 2019).

## 2.5 Physical fitness

To model our fitness coach we need to know the distinct differences between a fit and unfit subject. The distinction has to be made between a trained and untrained subject because these concepts are inherently linked but are not synonyms. To be “fit” is to have a large aerobic capacity, the body has adapted to be able to deliver more oxygen to the muscles which improve aerobic ability. To be “trained” is having the extra muscle and nervous impulses to be able to use this oxygen, converting it into more power than an untrained subject.

Another aspect that is closely linked to “fitness” is cardiovascular health. It is common knowledge that being more fit results in a lesser chance to contract cardiovascular problems such as infarcts, but being in good cardiovascular health doesn’t necessarily imply being fit. Attempting to alleviate poor cardiovascular health through fitness training is however a widely acknowledged and applied therapy.

## 2.5.1 FITNESS METRICS

Due to the large amounts of ways to parametrize “fitness”, calculating a fitness score involves carefully selecting which metrics to incorporate and which to discard.

### 2.5.1.1 $\text{VO}_{2\max}$ & Aerobic capacity

Aerobic capacity is the ability to take in oxygen and deliver it to skeletal muscle during exercise. It uses  $\text{VO}_{2\max}$  (or Peak Oxygen Uptake), the maximum rate of oxygen consumption in  $\text{mL}/(\text{kg}^*\text{min})$  - the volume of oxygen delivered per minute, per kilogram of body mass - , as a measure because it reflects the capabilities of the circulatory and respiratory systems. It is a gold standard measure for comparing individual subjects and training methods because it correlates with cardiorespiratory fitness and endurance capacity. Erythropoietin (EPO) can significantly increase  $\text{VO}_{2\max}$ , so it is no surprise that it has been used extensively as a Performance-Enhancing Drug in aerobic sports such as cycling. Note that higher  $\text{VO}_{2\max}$  doesn't necessarily mean a higher volume of oxygen in the blood. In fact, most  $\text{VO}_{2\max}$  increases can be attributed to a higher capacity of the body to extract the oxygen already present in the blood.

Accurately measuring  $\text{VO}_{2\max}$  is impossible without specialized medical equipment, but a meta-study revealed that  $\text{VO}_{2\max}$  may be simulated/extrapolated from RPE during a Perceptually Regulated Exercise Test (PRET) test. RPE can be noted using the Borg Scale, a standardised 6-20 scale with 6 meaning no exertion at all and 20 meaning maximal exertion. These ratings can informally be described as “exercise resulting in a heart rate of rating times 10”. (Coquart et al. 2016) Furthermore, performance during a Cooper test (maximum distance covered by foot during 12 minutes) can be extrapolated to  $\text{VO}_{2\max}$  using the formula:

$$\text{VO}_{2\max} \approx \frac{\text{dist} - 504.9}{44.73}$$

Where  $\text{dist}$  is the distance covered in meters in 12 minutes (Florindo & Latorre 2003).

### **2.5.1.2 Baecke Score**

The Baecke Score is a tried and tested (Florindo & Latorre 2003) self-administered questionnaire to evaluate the daily activity level of the subject. It is a quick way to determine a rudimentary score for the fitness level of an individual but requires validation from measurements to be a reliable metric.

The questionnaire includes items about daily work life, sport, and leisure time activities. The questions classify the subject's activity in each category from inactive to very active. Questions about sport and leisure time activities also include the type of activity, the frequency of performance, and the number of months per year that the activity is performed. All items result in a separate score from 1 to 5 points that incorporates activity duration, frequency, and an intensity score based on energy costs. Summing the work score, sports score, and leisure score results in an overall unitless activity score ranging from three to fifteen.

### **2.5.1.3 Blood Lactate levels**

Fatigue has an apparent impact on the slow phase of the blood lactate level increase: the higher the intensity that can be sustained without a significant presence of the slow component, the better the prospects for endurance. Endurance training will have the effect of elevating the value of the lactate threshold of an individual and thus eliminate the slow component for some exercise intensities. This way exercise intensities which were initially severe for the particular person might become heavy or even moderate following training (Zakynthinaki 2015).

### **2.5.1.4 Heart Rate variability**

Heart Rate Variability is the variation in time between each heartbeat. This variation is affected by the autonomic nervous system and will increase from PSNS activity and decrease from SNS activity. A healthy ANS is one that can quickly activate or deactivate either stimulus, so being able to quickly change from low to high variation and vice-versa is a measure of good physical and mental health and high stress resilience (Dong et al. 2018). Habitual aerobic exercise appears to aid the maintenance of elevated HRV in active subjects when compared with age- and weight-matched sedentary subjects (De Meersman 1993), and correlates with  $\text{VO}_{2\text{max}}$  (Buchheit & Gindre 2006).

Not all HRV measures are created equal; HF, SDNN, and RMSSD were statistically significantly higher in athletes than in controls, but other resting HRV parameters were not statistically different between groups (Danieli et al. 2014). It is equally important to compare HRV measures not in a vacuum but from longitudinal monitoring. In the case of elite athletes training for an upcoming event, another study found weekly and 7-day rolling averaged Ln rMSSD and the Ln rMSSD to R–R interval ratio as practically useful to represent a meaningful change in fatigue and/or fitness. In the same study, increasing HRV values were shown to be a possible sign of positive adaptation and/or coping with training load, while reductions in HRV in the week/days before pinnacle events may represent increasing freshness and readiness to perform (Plews et al. 2013).

#### 2.5.1.5 Heart Rate Recovery

HRR has become a significant, non-invasive measure of cardiovascular-parasympathetic influence and general cardiovascular health. A meta-analysis suggests that slowed HRR is an independent predictor of all-cause mortality (Qiu et al. 2017). HRR is statistically significantly faster in athletes (Danieli et al. 2014), and is associated with cardiac adaptation to physical activities of various durations and intensities (Durmić et al. 2019). HRR differs a lot from subject to subject, depending on the type of sport the subject is trained in, their age, the intensity of the training regime, etc. (Durmić et al. 2019) It is important to use the change in HRR over time as a measure of increased fitness, instead of using it as a baseline fitness metric.

HRR has a fast phase and a slow phase (see also section 2.3.5). Analysis revealed that the fast phase accounts for almost all the HR decay after intermediate intensity session (~98%), remaining only a minor portion of the decay falling in the slow phase. On the other hand, slightly different contributions are observed in high-intensity sessions (91% and 9%). For these reasons, percentage values of slow and fast phase might provide complementary information when comparing different exercise intensities. This same study also demonstrated that high-intensity exertion delays both parasympathetic reactivation and sympathetic withdrawal, resulting in a flatter curve (Bartels et al. 2018). The fast phase HRR correlates with parasympathetic modulation (Danieli et al. 2014).

Studies have shown no correlation between HRR and HRV (Esco et al. 2010) (Molina et al. 2016), but a correlation exists between HRR and Baecke score

(Buchheit & Gindre 2006), which means that having an active lifestyle results in faster HRR. Fast phase HRR (specifically, the Heart Rate Index, being  $HR_{max}$  minus Heart rate after 1 minute of recovery) also correlates with  $VO_{2max}$ . As  $VO_{2max}$  is primarily influenced by sympathetic activity, this means that an increase in fitness not only means higher parasympathetic modulation but also enhanced sympathetic activity. In general, HRR can be used as a measure of ANS adaptation (Durmić et al. 2019).

### 2.5.2 FITNESS TRAINING

Fitness training is based on three principles (Powers & Howley 1995):

- a training effect occurs when the tissue is stressed at a level beyond it is accustomed to (overload)
- a training effect is limited to the tissue involved in the activity (specificity)
- the effect of training is quickly lost when training is stopped (reversibility).

These principles aren't limited to muscular tissue, but can also be applied to aerobic capacity.  $VO_{2max}$  increases after consistent training sessions of 20 to 60 minutes, at an intensity of 50% to 85%  $VO_{2max}$ . Recall that  $VO_{2max}$  is measured as volume over time (per kg of body mass), which means that an increased Cardiac Output (due to higher preload and lower afterload) accounts for 50% of the  $VO_{2max}$  increase. The remaining 50% is due to an increased oxygen intake by organ tissue, resulting from a greater capillary density in the trained tissue (Powers & Howley 1995).

If an amount of muscle fibers are insufficient to accomplish a task, more fibers will be recruited, which increases the SNS response, HR response, and ventilation response to sustain the metabolic rate of these extra fibers. Fitness training results in an increased presence of mitochondrial cells in the muscle, which means that overall fewer muscle fibers need to be recruited which decreases the aforementioned responses (Powers & Howley 1995).

Carefully applied cycles of overload and adequate recovery are an important concept in fitness training. If the subject maintains or increases their exercise performance after the overload training period (despite having high perceived fatigue), he/she is considered Acutely Fatigued (AF). If the subject has decreased performance, he/she is considered overreached. Overreach can be split up into Functionally Overreached (FOR) and Non-Functionally Overreached (NFOR).

Detecting FOR or AF is not possible through HR monitoring, as no difference can be observed between the two states. A study concludes that heart rate is inadequate to prescribe and monitor intensified training. Subjective ratings of fatigue and readiness to train on visual analog scales were most powerful to distinguish between AF and FOR. In the same study, these 2 parameters correctly predicted 78% of the subjects as AF or FOR (Haaf 2018).

#### 2.5.2.1 Detraining

Detraining is the rapid decline of the benefits gained from training. A fast initial decrease of  $\text{VO}_{2\text{max}}$  is largely attributed to a decrease of blood plasma volume, which decreases Stroke Volume (It is shown that a plasma infusion can largely alleviate detraining symptoms within 12 days of the cessation of training). Another factor of aerobic capacity decreasing is the drop of muscular mitochondrial cells. A single week of detraining resulted in the loss of 50% of the gains during a 5-week training period (Powers & Howley 1995). Capillary density remained roughly constant during detraining.

#### 2.5.2.2 Overreaching spectrum

FOR is when a subject reaches a short-term decrement of performance, without signs of maladaptation as a consequence. When followed by a period of adequate recovery this state will reach an overall result in performance enhancement (termed supercompensation). NFOR is a dangerous cascade of insufficient recovery and excessive demand, resulting in a downward spiral marked by negative hormonal and psychological alterations. When left unchecked it will result in overtraining syndrome (OTS). Early-stage NFOR can be compensated by a quick response in recovery strategies; more rest and better-adapted sleep schedule, diet, and social activities (Kellmann et al. 2018).

Supercompensation is a popular term, but a meta-analysis (Haaf 2018) (Bellinger 2020) shows that similar improvement can be gained purely by AF instead of going into FOR:

“One of the challenges of research on the training-overtraining spectrum is to distinguish between AF and FOR. Some important differences exist between these two stages of the overtraining spectrum. Firstly, intensified training which results in AF leads to increased performance. In FOR athletes, however, the increase in

performance after intensified training and a taper period is smaller or even absent (Aubry et al. 2014). Secondly, disturbed sleep and an increased illness incidence in FOR compared to AF athletes has been shown (Hausswirth et al. 2014). Thirdly, continuation of imbalanced training- recovery in FOR athletes results eventually in NFOR or OTS. These stages must be prevented because of the severe symptoms and long recovery times (Meeusen et al. 2013). In conclusion, it is the difference between AF and FOR that is very relevant for athletes, because it represents the thin border between effective and too much training.” (Haaf 2018)

“It is clear that endurance athletes are required to undertake periodised increases in training load to provide an overload stimulus [31] and induce physiological adaptations to training. However, there is no evidence to suggest that inducing a state of FOR is necessary or required to promote these adaptations and improve exercise performance.” (Bellinger 2020)

### **2.5.2.3 Overtraining syndrome**

Overtraining, meaning the negative effects of training sessions that are too long and/or too strenuous, are a significant detriment to the physical and mental health of the subject. The majority of aerobic training injuries can be attributed to over-training. The general symptoms of overtraining are an increased heart rate, continued muscle soreness, loss of appetite, chronic fatigue, decrease in performance, psychological staleness, etc. (Kellmann et al. 2018) (Powers & Howley 1995) Compared to NFOR, recovering from OTS requires a continuous restoration consisting of long rest and recovery periods lasting from weeks to months accompanied by a reduced performance. (Kellmann et al. 2018)

Diagnosis of OTS can only be made after proper clinical evaluation and exclusion of the much more common medical conditions of endurance athletes. Common conditions with presentations similar to OTS include asthma, anemia, hypothyroidism, immunodeficiency, hypocortisolemia, chronic fatigue syndrome, and depression among others. To accurately diagnose OTS, additional blood screening tests are required to look for markers such as the following: testosterone:cortisol ratio, overnight urinary cortisol:cortisone ratio, maximal heart rate at lactate threshold, and/or decreased maximal lactate concentration (Kreher 2016). It is therefore advised to take preventive measures to avoid OTS altogether, a general rule in prevention is that training intensity or duration should not increase more than 10% week over week.

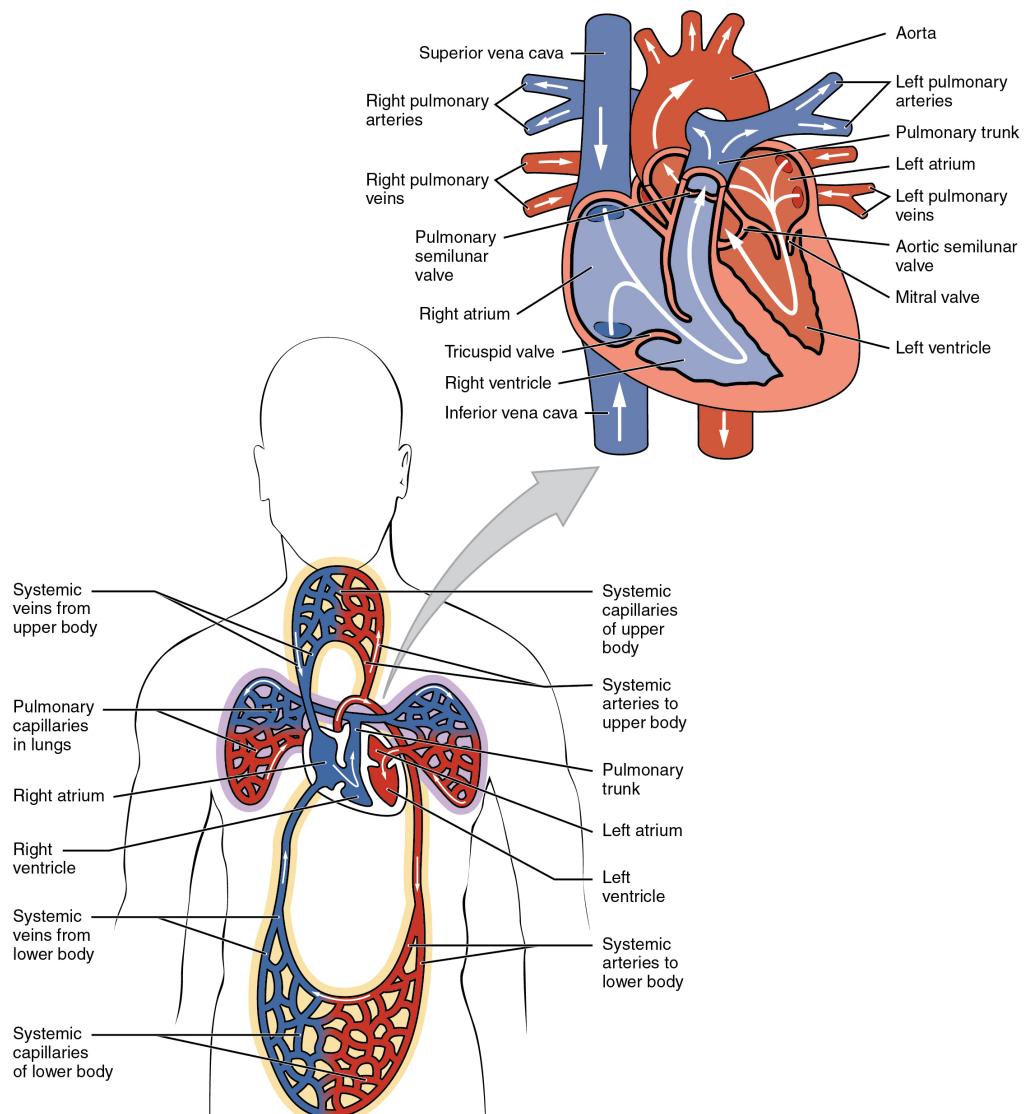


Figure 2.1: An anatomical overview of the heart (Betts et al. 2013)

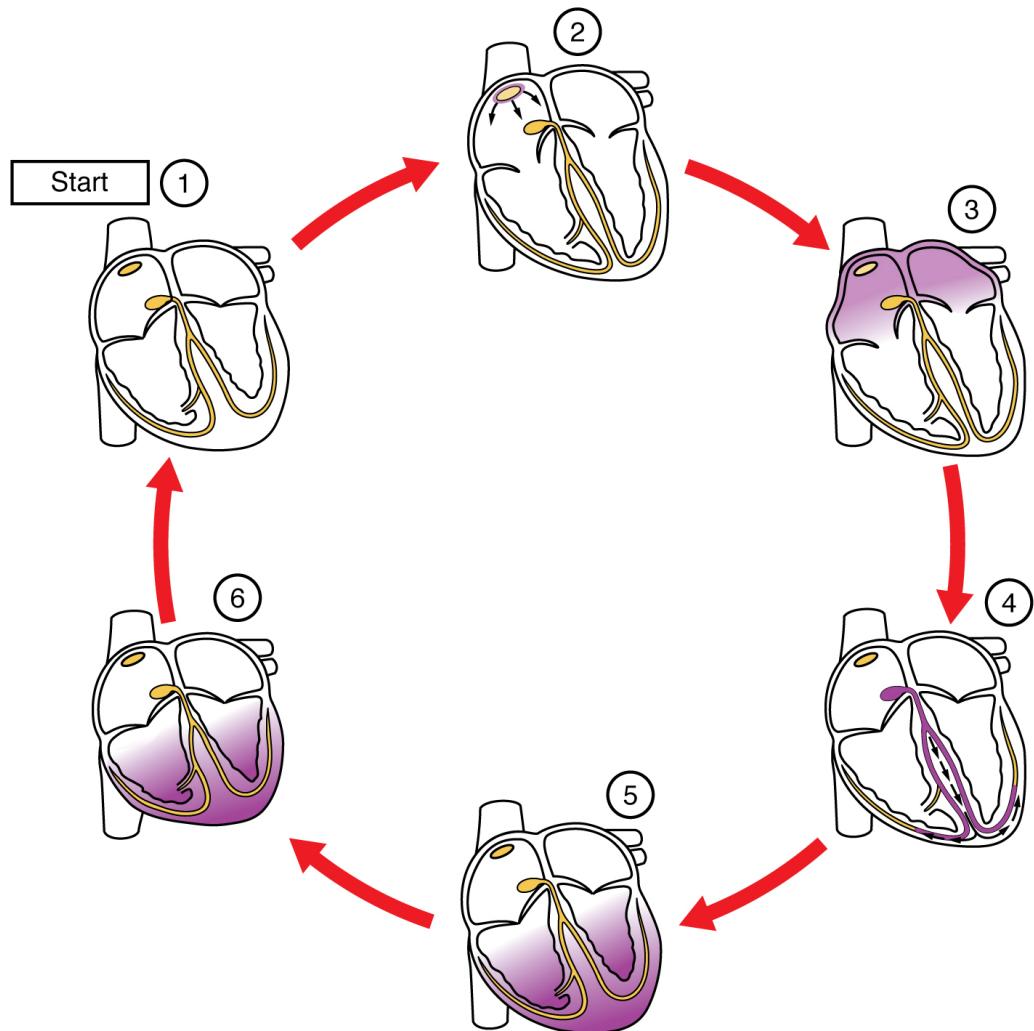


Figure 2.2: the innervation and pumping mechanism of the heart (Betts et al. 2013)

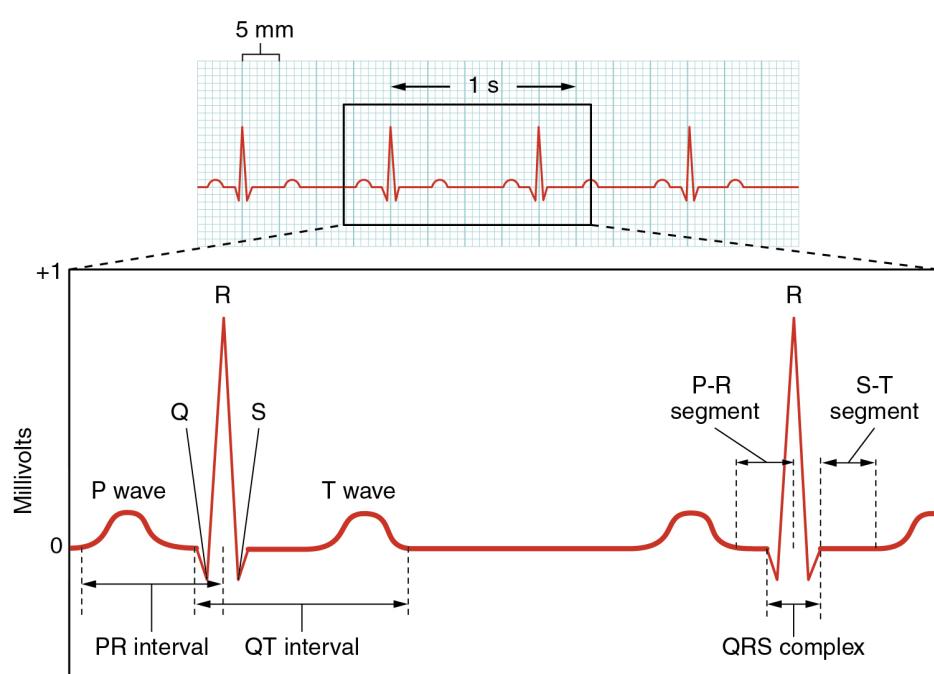


Figure 2.3: an example ECG tracing

## Chapter 3

# Data Gathering

Using the knowledge gained in the previous chapter, various metrics are carefully picked based on how relevant they are to cardiovascular fitness. In an ideal scenario, we could measure oxygen levels in the blood, but equipment for this kind of data gathering is expensive and bulky, which makes it hard to perform at the scale required for machine learning applications.

As a reminder, the subjects were asked to perform the following fitness training session:

- a 2-minute stage in sedentary rest
- a 5-minute stage in which the subject provides 50W (0.067 BHP) at 50 rpm
- a 5-minute stage in which the subject provides 100W (0.134 BHP) at 50 rpm
- a 2-minute stage of recovery in sedentary rest
- a 5-minute stage in which the subject tries to maintain 115-120 bpm Heart Rate (equals to around 60% of the maximum heart rate of young adults)
- a 2-minute stage of recovery in sedentary rest
- a 5-minute stage in which the subject tries to maintain 155-160 bpm Heart Rate (equals to around 80% of the maximum heart rate of young adults)
- a 1-minute stage in which the subject cycles at maximum exertion
- a 4-minute stage of recovery

Test subjects were sourced from the student population and were of mixed gender, various fitness levels, aged between 19 and 28 years old. Test subjects were asked to abstain from alcohol for at least 24 hours and to abstain from caffeine for at least 12 hours. This is an acceptable compromise between the personal life of the

test subjects and the half-life time of the substances in the circulatory system. In total 27 test subjects were gathered, of which 23 were able to fully complete the session, resulting in ~95.000 timesteps of heartbeat data.

The RR intervals are measured and stored using the Elite HRV Android app, which can then be exported to text files where each line contains an integer representing the milliseconds between each beat. These intervals are then sent through a filter and passed on for feature engineering.

### 3.1 RR interval filtering and correction

Not all heartbeats are captured perfectly. Any ectopic beats that occur need to be filtered out and beats that were erroneously measured or completely skipped by the heartbeat sensor or the Bluetooth connection need to be accounted for.

Before passing on to feature engineering, each interval is stored in a seven-beat window. When this window is full, the fourth beat is checked (Current Beat Timing, CBT). If it differs from the previous beat with either more than 50ms or more than 10ms and 3 times the current HRV, a closer look is taken.

The filterer will assume one of five scenarios happened. For each scenario, it creates a hypothetical solution and calculates how realistic this solution is. It then picks the best solution and applies a correction to the current seven-beat window.

The filter checks for multiple possibilities:

- there is a sudden change in HRV, the measured CBT is correct. The future 3 beats are taken and averaged out. The score of this solution is the CBT minus this average.
- the sensor failed to register a beat. The proposed solution is to replace the CBT with two beats with  $CBT/2$ . The score of this solution is the difference between the average of the seven-beat window excluding the CBT, and  $CBT/2$
- the sensor failed to register two successive beats. The proposed solution is to replace the CBT with three beats with  $CBT/3$ . The score of this solution is the difference between the average of the seven-beat window excluding the CBT, and  $CBT/3$
- the sensor misregistered a beat. The proposed solution is to average out the

CBT and its successor. The score of this solution is the difference between the average of the seven-beat window excluding the CBT and its successor, and the averaged timing value

- the sensor misregistered *and* missed a beat. The proposed solution is to average out the CBT and its successor and to add an additional beat with this averaged timing. The score of this solution is the difference between the average of the seven-beat window excluding the CBT and its successor, and the averaged timing value.

At first, if the score of the sudden change solution is less than 50ms, it means that there is no faulty measurement, just a sudden change in Heart Rate Variability. However, if this is not the case, the filterer checks for the lowest score of the four proposed solutions. If this score is lower than the score of the “sudden change”, it is accepted and the seven-beat window is replaced by the solution. If the cause of the change in IBI timing is still unexplained, the filtering gives up and will simply accept the faulty measurement.

### 3.2 Heart Rate metrics feature engineering for the classification task

The feature engineering results in every RR interval being accompanied by 10 extra features:

- HR
- HRV
- SDNN
- SDSD
- pNN50
- RMSSD
- HF power (60s epoch)
- LF power (300s epoch)
- VLF power (300s epoch)
- LF/HF ratio (300s epoch)

Figure 3.1 plots these datapoints out for a sample user over time. It is immediately clear that some correlations exist, which means that we should be able to obtain a working Machine Learning classifier.

### 3.2.1 RR INTERVAL

The most basic measurement from which all others are derived, the RR interval is the exact time between heartbeats in milliseconds. These values are stored in various sliding windows (of a constant time interval, 10 seconds, 60 seconds, 120 seconds, and 300 seconds respectively) for further derivation.

### 3.2.2 CURRENT HEART RATE

The beat-to-beat heart rate is the standard way to measure cardiac activity. It is simply the number of full contractions of the heart per minute. Because we are starting from heart rate variability metrics, it is easily derived by  $(60000/\text{RR interval in milliseconds})$ , but to reduce the inherent variability of heart rate variability, the linearly weighted moving average of the last 10 seconds of measurements is taken.

### 3.2.3 CURRENT HEART RATE VARIABILITY

A very simple way to determine the current overall HRV is to simply make the difference between the last 2 beats, but this proved to be a very unstable measurement, again due to the inherent variability of HRV. So just like HR, the linearly weighted moving average of the last 10 seconds of measurements is taken.

### 3.2.4 MAXIMUM HEART RATE

Maximum heart rate being a constant value for a subject regardless of their fitness level makes it an important metric to track. During the data gathering training session, subjects were asked to perform a maximal exertion test, so the maximum heart rate achieved during the session is assumed to be their maximum heart rate.

### 3.2.5 RESTING HEART RATE

During the data gathering training session, subjects started with a 2-minute sedentary rest period. Theoretically, the resting heart rate can be determined during this window, but some subjects showed lower heart rates during certain recovery segments in the middle of the session. This can be attributed to the increased

stress level the subject attained from the anticipation of being tested, which is absent during the recovery segments.

Therefore, the resting heart rate is determined to be the global minimum heart rate attained during the session.

### 3.2.6 HRV TIME DOMAIN

Standard Deviation of Inter Beat Intervals is the most basic way to analyse HRV in the time domain, so it should be included in our model inputs. The standard deviation of the successive differences between IBIs is also an easy and basic measurement. (Danieli et al. 2014) also shows that RMSSD and SDNN are higher in athletes so both metrics are included.

pNN50 was chosen over NN50 because the number of beats in the 120-second window varies greatly depending on the current heart rate of the subject. (at maximum exertion, the RR interval can drop to 350ms resulting in a window of 300-350 beats, while at rest the RR interval can be as high as 1000ms, resulting in a window of 120 beats)

### 3.2.7 HRV FREQUENCY DOMAIN

For extracting frequency domain statistics Welch's *spectral density estimation* is used. The resulting periodogram is returned as a list of equally spaced segments, where each element represents the power of that particular segment. For each frequency band, we can sum the corresponding segments to achieve the power of that band.

(Danieli et al. 2014) shows that HF power is higher in athletes, so it is included. Because it is also a measure of parasympathetic activity, we analyse HF power in a fairly short 60-second epoch otherwise we would not be able to accurately assess HF power during the relatively short recovery and resting segments.

Due to the low period of VLF and LF power (oscillation periods of up to 300 seconds and up to 25 seconds respectively), their power is calculated from 300-second epochs. LF/HF ratio is also derived from this 300-second IBI window.

### 3.2.8 CLASSIFICATION

Due to the setup of the training session, it is easy to classify each heartbeat. A global time counter keeps track of the absolute time passed since the start of recording, which makes it easy to classify each stage of the session:

- the resting stage is classified as “resting” (class 0)
- the active stages are classified as “active” (class 1)
- the recovery stages are classified as “recovery” (class 2)

## 3.3 Subject Variables for the regression task

For each subject, some extra meta-data is gathered for use in the fitness regression task.

### 3.3.1 FITNESS SCORE

The following meta-data is noted (fig 3.3):

- the resting heart rate of the subject
- the maximum heart rate of the subject during the 50W 50 rpm stage
- the maximum heart rate of the subject during the 100W 50 rpm stage
- the maximum heart rate of the subject
- the energy spent during the three “constant heart rate” stages
- the Baecke questionnaire scores of the subject

out of these variables, a fitness score is calculated from the sum of 3 values:

- The calories expended during the maximum exertion stage, multiplied by 2. Maximum exertion is a good metric for  $VO_{2max}$ , but as we can't measure oxygen levels in the blood we use the expended energy as a substitute.
- The sum of the calories expended during the two constant Heart Rate stages, multiplied by the Baecke score divided by 15. The energy expenditure of these two stages also reflects the capacity for the subject to produce energy under constant load, but this measure is less reliable and not backed by research. By incorporating the Baecke score at this point we can make this

measure less important while also grading the subject on their lifestyle which has been shown to correlate with cardiovascular fitness.

- The average percentage of the heart rate reserve used by the subject during both constant load stages. heart rate reserve is the difference between the maximum heart rate and resting heart rate of the user. The lower the amount used of this reserve, the better the subject has adapted to exerting this load

The resulting score (3.2) is a unitless value between 60 and 170. This score is very dependent on gender and should not be used to compare different subjects, as it is intended to monitor the increase or decrease in cardiovascular fitness of the subject.

### 3.3.2 INPUT DATA

Because of the limitations of neural networks, the size of the input must be identical for all subjects. This means that the variable-length heartbeat data can't simply be passed to the machine learning algorithm. The data of the training session must be engineered to fit a constant-length input.

The maximum heart rate, resting heart rate, VLF, LF & HF power, and LF/HF ratio are extracted from the complete session. For every active stage of the session the VLF power, LF power, HF power, LF/HF ratio, maximum Heart rate, and rMSSD of that 5-minute segment is extracted. For every recovery stage, these same parameters are extracted with the additional Heart Rate Recovery after 1 minute and the Heart Rate Recovery after 2 minutes.

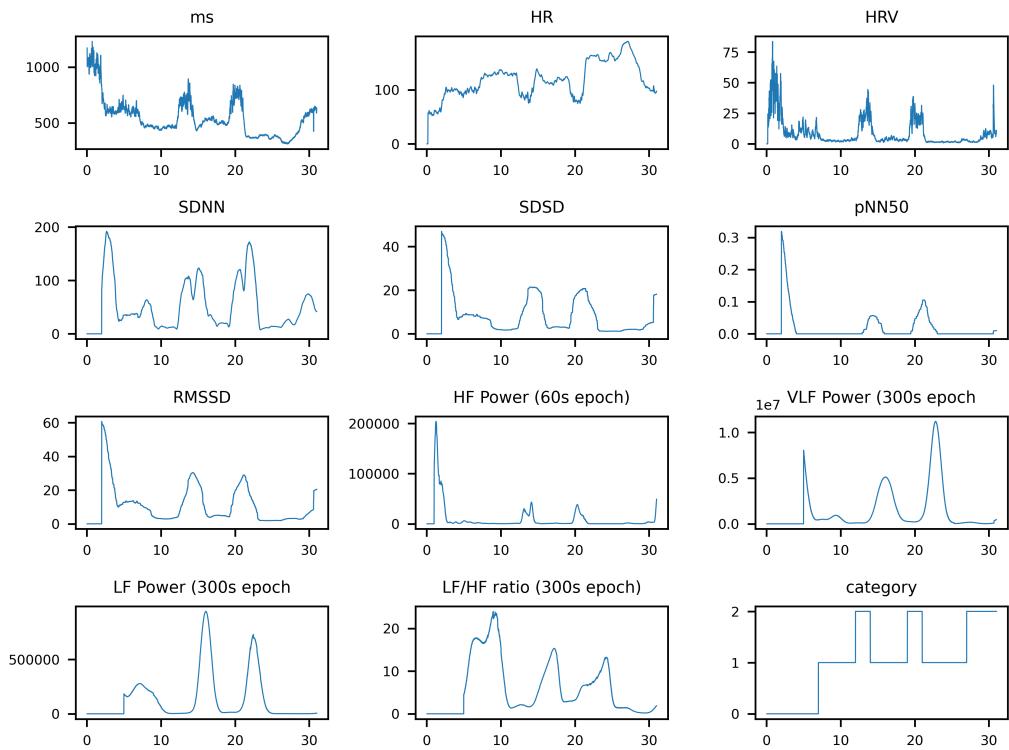


Figure 3.1: Feature extraction from heartbeat data

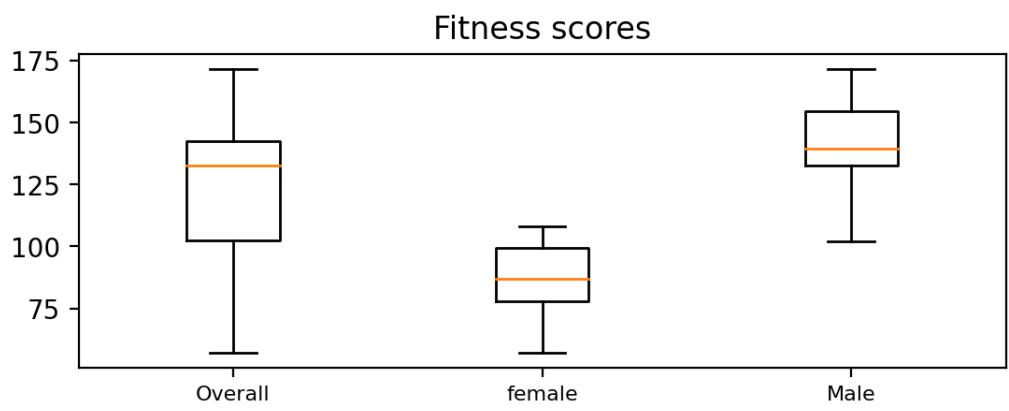


Figure 3.2: Fitness scores of the test subjects

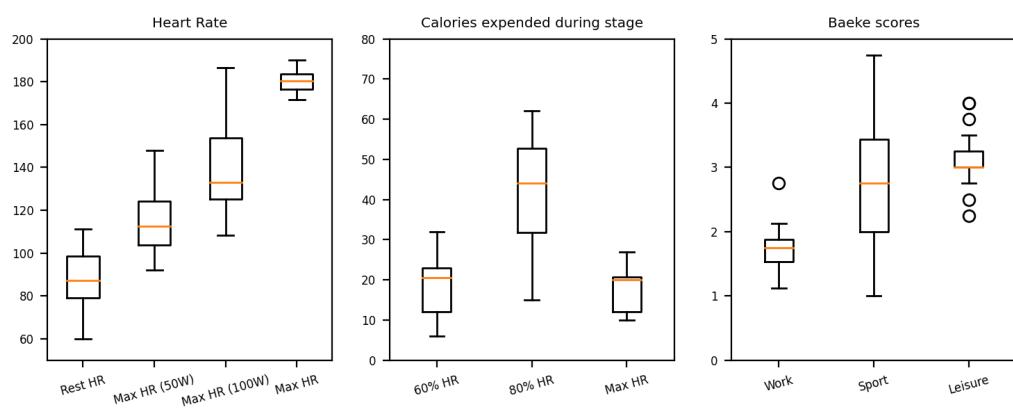


Figure 3.3: Fitness scores sub-variables of the test subjects

## Chapter 4

# Activity State Classification

For the activity state classification task, the proposals stem from either natural language processing or earlier experimentations involving machine learning and heart rate data. The raw input data of all users are concatenated, resulting in 95.000 data points. Indexes spaced evenly apart based on the desired sequence length are generated, and a random offset is generated based on the length of the remaining beats that don't fit inside a sequence.

For each epoch, an attempt is made to assign a class to all these sequences. If all beats in the tail of the sequence are the same class, the entire sequence gets assigned to this class. The length of this tail is a parameter that is explored later in the chapter, but the default case is a tail length of half the sequence length.

```
if np.all(data[seq_end-(seq_len/2):seq_end,-1] == clss):
    X[clss] = np.append(X[clss], [data[seq_start:seq_end,:-1]])
```

After these classes are determined, an equal amount of sequences for each class is bundled as the data for that epoch. After each epoch, the indexes are reshuffled and the offset is regenerated so that each epoch will see slightly different sequences in a different order.

```
# 2 sequence lengths of spare space:
# one seq_len for the actual sequence and
# one seq_len for when the offset is equal to seq_len
indexes = [i*seq_len for i in range(math.floor(len(data) / seq_len)-1)]
max_offset = len(data) % seq_len
```

```

if max_offset == 0:
    max_offset = seq_len
    indexes = indexes[:-1]

current_offset = random.randrange(max_offset)

```

Using this method, each epoch would result in approximately 30 batches, each containing 9 sequences of 128 timesteps. Batch size is chosen as a multiple of the number of classes so that the total amount of sequences is divisible by the batch size. All the following networks use some form of softmax activation in their final layer and are trained with the “adam” optimizer using categorical cross-entropy as the loss function. This combination is chosen because it is optimized for multi-class labeling problems.

Another optimisation made is to dampen the learning rate once the network is reaching an equilibrium. Due to the relatively small dataset, the accuracy can vary wildly from epoch to epoch. By lowering the learning rate towards the end of the training run, a more consistent final result is achieved.

Each network is trained three times using the same settings and the result is averaged out, but in almost all cases the result of training fell within a margin of error.

## 4.1 Long Short Term Memory

Long short-term memory (LSTM) is a type of recurrent neural network (RNN) that is used to analyse sequence data. A common use-case is connected handwriting recognition and speech recognition. An LSTM consists of multiple cells, each containing a memory cell and gates that make the cell capable of remembering values over arbitrary time intervals. These cells can pass a value to their successor cell (the hidden state), which makes them capable of handling sequences of data. The LSTM itself was proposed as an evolution of the standard RNN to solve the vanishing gradient problem, in which gradients that are back-propagated can explode or implode in size.

The LSTM used has a cell for each beat in the input sequence, each with 11 hidden states (one for each feature). This layer outputs the hidden states of the very last

cell. A densely connected layer containing three neurons, each signifying one of the three output classes, is used as the output layer.

This network manages to achieve some correct labeling with an accuracy of ~58% on the validation set. A completely random labeler would correctly label 33% of sequences, so LSTM is better but still far from ideal.

## 4.2 Deep LSTM

As an evolution of a standard LSTM network, a network with three stacked LSTM layers is evaluated. Instead of the LSTM layer returning the hidden states of the last cells, it outputs all the hidden states of all cells. This is then passed to a second LSTM layer which then passes all of its hidden states to the third and final LSTM layer. This third layer passes the hidden outputs of its very last cell to a dense layer in the same way as the shallow LSTM. This results in a network containing 3072 trainable parameters. The network achieves ~65% accuracy on the validation set, which is marginally better than a simple LSTM.

## 4.3 DeepHeart

DeepHeart is a network architecture proposed by (Ballinger et al. 2018) to classify HRV data into 5 categories (normal, diabetes, high blood pressure, and sleep apnea). It is designed to give a classification for every timestep (Heartbeat in this case), which makes it a sequence-to-sequence algorithm. The model consists of the following layers:

- A convolutional layer with a filter size of 12, followed by a 0.2 dropout and a max pooling layer with a pool size of 2
- Two convolutional layers with a filter size of 5, each followed by a 0.2 dropout and a max pooling layer with a pool size of 2
- 4 Bidirectional LSTM layers, which means that each layer consists of 2 LSTM chains, one standard, and one that goes backward in time. This has the benefit that these layers can analyze sequences in the past as well as in the future. Of these four layers the first three output all their hidden states to the following layer and the fourth outputs the hidden state of the last cell
- Here a slight adaptation is made to the original design. Instead of a “conv-

lution of filter length 1” with tanh activation, a densely connected layer is used as output for the classification, to make the network more suitable to multilabel classification for disjunct classes.

This results in a network containing 576.771 trainable parameters. The network achieves only ~62% accuracy on the validation set, which makes it perform on the same level as the fairly simple and computationally less intensive simple LSTM network.

#### 4.4 DeepHeart v2

DeepHeartV2 is a proposed evolution of DeepHeart for this thesis with the following changes:

- In the original DeepHeart, the sequences are convoluted before being passed to multiple layers of RNNs. This has a significant impact on the productivity of the LSTM layers as much of the temporal information is filtered out before it reaches the memory cells of the LSTM. In the original model, there were only 11 timesteps left in these layers. In this version of DeepHeart, the 4 stacked LSTM layers are swapped with the 9 layers of the convolutional network.
- The convolutional layers now use causal padding, a technique also used in WaveNet (Oord et al. 2016)
- A final LSTM is added before the densely connected output layer.
- The bidirectional LSTMs each have 11 hidden outputs to match the number of features in the input data.

This results in a network containing 221.405 trainable parameters. The network achieves an accuracy of ~75% on the validation set, which makes it perform significantly better than the original DeepHeart.

#### 4.5 Temporal Convolutional Network

A Temporal Convolution Network (also known as a Causal Convolutional Network) is an adaptation of the well known Convolutional Neural Network, where there

is no recurrent connection between “cells”, but instead the input layer (receiving the sequence timesteps) gets step-wise convolved down to an output layer. These convolutions must respect the ordering in which the data is modeled.

This network architecture requires either a large number of layers or wide filtering at every layer to achieve the large “receptive field” of the input layer. A way to increase this field without significantly increasing computational size is by using “Dilated Causal Convolution”. This involves ignoring certain connections between layers based on their location and depth within the network. Visually it can be explained by imagining a tree structure within the layers (fig 4.1).

In this implementation the keras-tcn package is used, which can stack multiple of these networks together as residual blocks (fig 4.2) which means these blocks have the potential to be skipped over. The TCN used in this thesis has a filter size of 2, contains 4 stacks of residual blocks, each containing 5 layers with resp. 1, 2, 4, 8, and 16 dilations, resulting in a receptive field of 128 timesteps. This layer outputs the states of its 128 timesteps to the densely connected layer containing 3 neurons, using softmax activation results in the classification that we want.

This results in a network containing 1.308.291 trainable parameters which makes it the most complex model by a wide margin. The network achieves ~85% accuracy on the validation set, which makes it perform the best out of all models tested thus far.

## 4.6 Labeling methods

As mentioned before in the introduction of this chapter, the method used to assign a class is a parameter in itself. If all beats in the tail of the sequence are the same

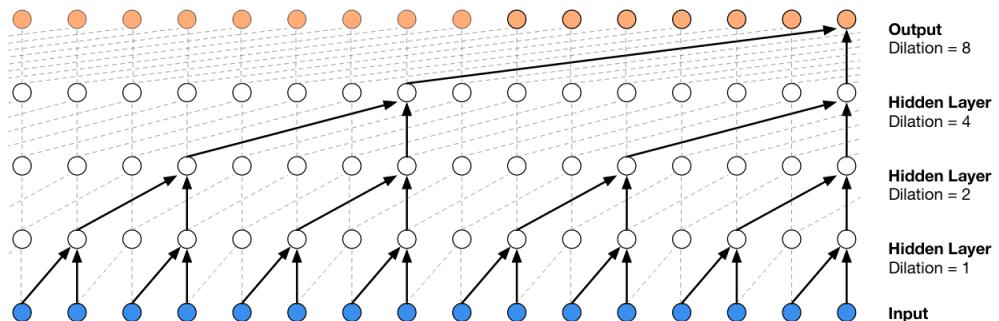


Figure 4.1: Dilated Causal Convolution (Oord et al. 2016)

class, the entire sequence gets assigned to this class. If this is not the case, the proposed sequence is instead discarded, this means that entire sections in the dataset are not used altogether. Decreasing this tail length would increase the dataset because fewer sequences are discarded.

It is however clear that a sufficiently low “checking parameter” will result in a sequence that is hard to predict. If the checking parameter is  $1/seq\_len$ , the first  $seq\_len-1$  can have a different class than the eventual sequence class.

The impact of this parameter is explored by training the 2 best performing models (DeepHeartV2 and TCN) on various tail lengths (TL).

**Table 4.1:** The resulting accuracy of decreasing classification tail length.

Model	TL 50%	TL 25%	TL 10%	TL 5%
DeepHeart V2	75%	70%	69%	66%
TCN	85%	82%	81%	80%

Only a mild accuracy drop is observed in the TCN, while DeepHeartV2 shows a larger drop. What is apparent is that the accuracy drop is manageable, especially after a further investigation showed that it became harder for the model to achieve an optimal result with lower tail lengths. Some training runs showed nearly identical performance with TL 5% as with TL 50%, but the lower the TL parameter the lower the number of training sessions that achieved this accuracy, dragging the average result down. The training data size increased from approx. 30 batches per epoch for 50% TL to approx 35 batches per epoch for 5% TL, which is only a slight increase.

## 4.7 Seq2Seq Classification

Sequence-to-sequence classification (where each element in a sequence receives a classification, instead of just the full sequence receiving a single classification) was considered but proved to be too inaccurate to be viable as a solution.

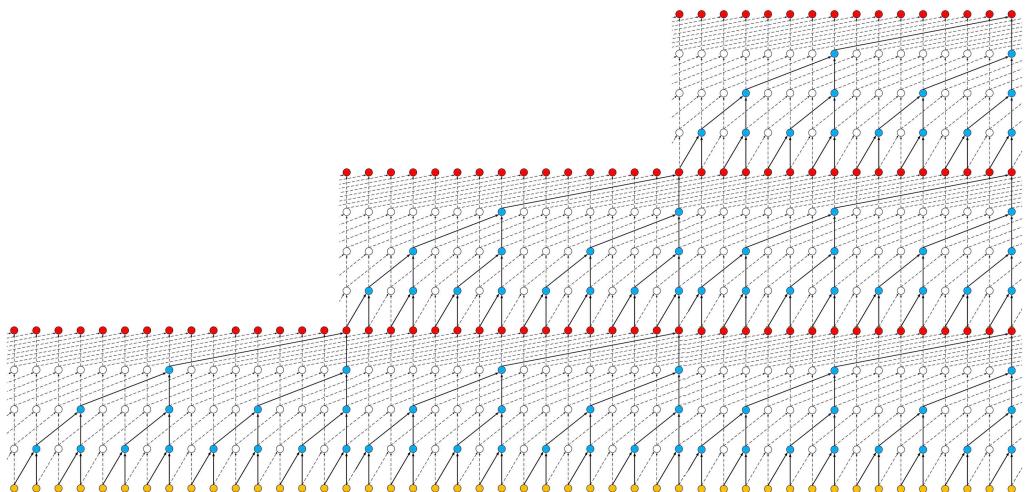


Figure 4.2: Residual blocks in the TCN (From the keras-tcn github page)

## Chapter 5

# base-level Fitness regression

The base-level fitness regression is a fairly simple task. For each test subject, the data from their fitness training session was extracted and the base-level fitness was scored based on meta-data gained from the Baecke questionnaire and the stationary bicycle.

The goal would now be to create a model that can perform regression on just the data extracted from the Heart Rate sensor. If this method can be shown to be a valid way to achieve a fitness score, it means that it is possible to measure cardiovascular fitness based on heart rate data alone.

A simple densely connected neural network consisting of a handful of densely connected layers should be able to handle this task, after which possible improvements like Convolutional Neural Networks and others can be compared. However, it quickly became clear that the dataset of 23 subjects is too small to provide any result. Data resampling such as Bootstrapping (Creating new input samples by randomly sampling elements from other existing input samples) or Jackknifing (determining the distribution of each sample element and randomly generating new input samples from these distributions) would make it difficult to calculate a corresponding base-level fitness score, so would not have been beneficial for proving the validity of the regression task.

Validation loss rose to astronomical values and predictions on validation data were orders of magnitude larger or smaller than the expected 60-200 value range. An attempt was made to use K-Fold Cross Validation which mildly increased the training quality, but still, no adequate result was reached.

# Chapter 6

## Conclusion

### 6.1 Learning points

The proposed Machine Learning networks managed to achieve up to ~82% accuracy on sequenced heartbeat data. This is a fairly positive result, but not yet applicable for real-world scenarios.

Due to the current ministerial decree(s) with the ongoing covid-19 pandemic, the number of participants to the study was unfortunately too limited to see any significant results. The datasets used during training of both subgoals resulted in short epochs, requiring high epoch counts (up to 500 epochs) to achieve a stable result. For the base-level fitness regression, this became even more extreme resulting in the failure of the experiment.

However, while it is impossible to fully prove the validity of the applied methods until more data is gathered, both subgoals were explored and validated using heavily cited, peer-reviewed studies. The base-level fitness goal in particular is a fairly simple computational task (regression), so the the eventual success of the methodology seems likely.

### 6.2 Future work

There are several potential directions for extending this thesis.

Being able to use PPG data from smartwatches would lower barrier of entry, users

and test subjects would not have to strap on a special device for testing and could be monitored permanently. Using Pulse Rate Variability measurements instead of Heart Rate Variability measurements requires only minor changes to the code used in this thesis, but requires a completely new test setup.

Currently, the Activity Stage model does not use the potential gains from having the heart beat data streamed to the model. Accuracy could potentially improve by being able to use the previously labeled timesteps instead of only the unlabeled timesteps. So-called “Context Aware Learning” would be able to not only classify the activity state of the user based on short-term sequence data, but it would know a longer term “context” in which the data was generated (An increased heart rate during a High Intensity Interval Training versus running up the stairs during a regular work day) (Lore & Reddy 2019). Certain techniques are being discovered to not only classify this context, but also keep track of the “drift” of the context, mostly within the field of Anomaly Detection in streamed data w.r.t. network security (Mulinka & Casas 2018) (Mulinka et al. 2019)

An interesting concept was found but not expanded upon. The FitzHugh Nagumo model is the model of an a biological neuron proposed to account for cardiac impulse propagation. A study was able to achieve promising results by creating a Neural Network Layer based on this model, but these results were unable to be replicated in the experimental models of this thesis (Madl 2016).

The base-level Fitness regression task is made to combine 2 subgoals (base-level fitness & increase/decrease of fitness) into 1 implementation. The disadvantage now is that tracking increase/decrease of fitness requires performing a predefined training session, which makes the entire coach much less integrable with daily life. Developing a custom solution for this subgoal made to work with unlabeled sequence data would solve this issue.

The implementation of the three remaining subgoals (tracking and coaching the user during a session, compiling a result from past sessions and recommending sessions that optimize fitness training) have not been addressed, but are required for a fully mature fitness coach. Various background information (such as fatigue states) have not been fully applied in the addressed subgoals, but might prove useful when implementing the remaining subgoals.

# Appendix A: Network Summaries

LSTM network summary

Layer (type)	Output Shape	Param #
<hr/>		
input_1 (InputLayer)	[(None, 128, 11)]	0
<hr/>		
lstm (LSTM)	(None, 11)	1012
<hr/>		
dense (Dense)	(None, 3)	36
<hr/>		
Total params: 1,048		
Trainable params: 1,048		
Non-trainable params: 0		
<hr/>		

Deep LSTM network summary

Layer (type)	Output Shape	Param #
<hr/>		
input_1 (InputLayer)	[(None, 128, 11)]	0
<hr/>		
lstm (LSTM)	(None, 128, 11)	1012

```

-----
lstm_1 (LSTM)           (None, 128, 11)      1012
-----
lstm_2 (LSTM)           (None, 11)          1012
-----
dense (Dense)           (None, 3)           36
=====
Total params: 3,072
Trainable params: 3,072
Non-trainable params: 0
-----
```

## DeepHeart network summary

```

-----
Layer (type)            Output Shape        Param #
-----
input_1 (InputLayer)     [(None, 128, 11)]   0
-----
conv1d (Conv1D)          (None, 117, 128)    17024
-----
dropout (Dropout)        (None, 117, 128)    0
-----
max_pooling1d (MaxPooling1D) (None, 58, 128)  0
-----
conv1d_1 (Conv1D)        (None, 54, 128)     82048
-----
dropout_1 (Dropout)      (None, 54, 128)     0
-----
max_pooling1d_1 (MaxPooling1 (None, 27, 128)  0
-----
conv1d_2 (Conv1D)        (None, 23, 128)     82048
-----
dropout_2 (Dropout)      (None, 23, 128)     0
-----
max_pooling1d_2 (MaxPooling1 (None, 11, 128)  0
-----
```

bidirectional (Bidirectional (None, 11, 128)		98816
-----		
bidirectional_1 (Bidirection (None, 11, 128)		98816
-----		
bidirectional_2 (Bidirection (None, 11, 128)		98816
-----		
bidirectional_3 (Bidirection (None, 128)		98816
-----		
dropout_3 (Dropout)	(None, 128)	0
-----		
dense (Dense)	(None, 3)	387
=====		
Total params: 576,771		
Trainable params: 576,771		
Non-trainable params: 0		
-----		

## DeepHeartV2 network summary

Layer (type)	Output Shape	Param #
-----		
input_1 (InputLayer)	[(None, 128, 11)]	0
-----		
bidirectional (Bidirectional (None, 128, 22)		2024
-----		
bidirectional_1 (Bidirection (None, 128, 22)		2992
-----		
bidirectional_2 (Bidirection (None, 128, 22)		2992
-----		
bidirectional_3 (Bidirection (None, 128, 22)		2992
-----		
dropout (Dropout)	(None, 128, 22)	0
-----		
conv1d (Conv1D)	(None, 128, 128)	33920
-----		
dropout_1 (Dropout)	(None, 128, 128)	0

```

-----  

max_pooling1d (MaxPooling1D) (None, 64, 128) 0  

-----  

conv1d_1 (Conv1D) (None, 64, 128) 82048  

-----  

dropout_2 (Dropout) (None, 64, 128) 0  

-----  

max_pooling1d_1 (MaxPooling1 (None, 32, 128) 0  

-----  

conv1d_2 (Conv1D) (None, 32, 128) 82048  

-----  

dropout_3 (Dropout) (None, 32, 128) 0  

-----  

max_pooling1d_2 (MaxPooling1 (None, 16, 128) 0  

-----  

bidirectional_4 (Bidirection (None, 22) 12320  

-----  

dense (Dense) (None, 3) 69  

=====  

Total params: 221,405  

Trainable params: 221,405  

Non-trainable params: 0
-----
```

## TCN network summary

```

-----  

Layer (type) Output Shape Param #  

=====  

input_1 (InputLayer) [(None, 128, 11)] 0  

-----  

tcn (TCN) (None, 128) 1307904  

-----  

dense (Dense) (None, 3) 387  

=====  

Total params: 1,308,291  

Trainable params: 1,298,051
-----
```

Non-trainable params: 10,240

---

## Appendix B: Abbreviations

<b>HRV</b>	Heart Rate Variability
<b>RPE</b>	Rating of Perceived Exertion
<b>LSTM</b>	Long Short Term Memory
<b>TCN</b>	Temporal Convolutional Network
<b>SA node</b>	Sinoatrial node
<b>AV node</b>	Atrioventricular node
<b>CNS</b>	Central Nervous System
<b>PNS</b>	Peripheral Nervous System
<b>ANS</b>	Autonomic Nervous System
<b>SNS</b>	Somatic Nervous System
<b>PSNS</b>	ParaSympathetic Nervous System
<b>RSA</b>	Respiratory Sinus Arrhythmia
<b>CO</b>	Cardiac Output
<b>SV</b>	Stroke Volume
<b>RHR</b>	Resting Heart Rate
<b>HRR</b>	Heart Rate Recovery
<b>IBI</b>	Inter Beat Interval
<b>AF</b>	Acutely Fatigued
<b>FOR</b>	Functionally Overreached
<b>NFOR</b>	Non-Functionally Overreached
<b>OTS</b>	Overtraining Syndrome
<b>BHP</b>	Brake Horsepower

# References

- Abbasi-Kesbi, R., Valipour, A. & Imani, K., 2018. Cardiorespiratory system monitoring using a developed acoustic sensor. *Healthcare Technology Letters*, 5(1), pp.7–12.
- Ballinger, B. et al., 2018. Deepheart: Semi-supervised sequence learning for cardiovascular risk prediction. *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*, pp.2079–2086. Available at: <http://arxiv.org/abs/1802.02511>.
- Bartels, R. et al., 2018. Heart rate recovery fast-to-slow phase transition: Influence of physical fitness and exercise intensity. *Annals of Noninvasive Electrocardiology*, 23(3), pp.1–7.
- Bassan, M.M. et al., 2005. Heart-rate profile during exercise as a predictor of sudden death [2] (multiple letters). *New England Journal of Medicine*, 353(7), pp.734–735.
- Bellinger, P., 2020. Functional Overreaching in Endurance Athletes: A Necessity or Cause for Concern? *Sports Medicine*, (0123456789). Available at: <https://doi.org/10.1007/s40279-020-01269-w>.
- Betts, J.G. et al., 2013. *Anatomy & Physiology -OpenStax*, Available at: <https://openstax.org/details/books/anatomy-and-physiology>.
- Buchheit, M., 2014. Monitoring training status with HR measures: Do all roads lead to Rome? *Frontiers in Physiology*, 5 FEB(February), pp.1–20.
- Buchheit, M. & Gindre, C., 2006. Cardiac parasympathetic regulation: Respective associations with cardiorespiratory fitness and training load. *American Journal of Physiology - Heart and Circulatory Physiology*, 291(1).
- Coquart, J. et al., 2016. Submaximal, Perceptually Regulated Exercise Testing Predicts Maximal Oxygen Uptake: A Meta-Analysis Study. *Sports Medicine*, 46(6), pp.885–897.
- Cygankiewicz, I. & Zareba, W., 2013. *Heart rate variability* 1st ed., Elsevier B.V. Available at: <http://dx.doi.org/10.1016/B978-0-444-53491-0.00031-6>.
- Danieli, A. et al., 2014. Resting heart rate variability and heart rate recovery after submaximal exercise. *Clinical Autonomic Research*, 24(2), pp.53–61.
- De Meersman, R.E., 1993. Heart rate variability and aerobic fitness. *American Heart Journal*, 125(3), pp.726–731.

- Dong, S.Y. et al., 2018. Stress Resilience Measurement with Heart-Rate Variability during Mental and Physical Stress. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2018-July, pp.5290–5293.
- Durmić, T. et al., 2019. Usefulness of heart rate recovery parameters to monitor cardiovascular adaptation in elite athletes: the impact of the type of sport. *Physiology International*, 106(1), pp.81–94. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/30888216>.
- Esco, M.R. & Flatt, A.A., 2014. Ultra-short-term heart rate variability indexes at rest and post-exercise in athletes: Evaluating the agreement with accepted recommendations. *Journal of Sports Science and Medicine*, 13(3), pp.535–541.
- Esco, M.R. et al., 2010. The relationship between resting heart rate variability and heart rate recovery. *Clinical Autonomic Research*, 20(1), pp.33–38.
- Florindo, A.A. & Latorre, M. do R.D. de O., 2003. BaeckePAmeasure. *Revista Brasileira de Medicina do Esporte*, 9(3), pp.129–135.
- Haaf, T. ten, 2018. *On the physiological and psychological differences between functional over-reaching and acute fatigue*,
- Kellmann, M. et al., 2018. Recovery and performance in sport: Consensus statement. *International Journal of Sports Physiology and Performance*, 13(2), pp.240–245.
- Kreher, J., 2016. Diagnosis and prevention of overtraining syndrome: an opinion on education strategies. *Open Access Journal of Sports Medicine*, Volume 7, pp.115–122.
- Lore, K.G. & Reddy, K.K., 2019. Implicit Context-aware Learning and Discovery for Streaming Data Analytics. Available at: <http://arxiv.org/abs/1910.08438>.
- Madl, T., 2016. Deep neural heart rate variability analysis., (Nips). Available at: <http://arxiv.org/abs/1612.09205>.
- Malik, M. et al., 1996. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, 17(3), pp.354–381. Available at: <https://www.ahajournals.org/doi/full/10.1161/01.cir.93.5.1043>.
- Menghini, L. et al., 2019. Stressing the accuracy: Wrist-worn wearable sensor validation over different conditions. *Psychophysiology*, 56(11), pp.1–15.
- Molina, G.E. et al., 2016. Post-exercise heart-rate recovery correlates to resting heart-rate variability in healthy men. *Clinical Autonomic Research*, 26(6), pp.415–421.
- Mulinka, P. & Casas, P., 2018. Stream-based machine learning for network security and anomaly detection. *Big-DAMA 2018 - Proceedings of the 2018 Workshop on Big Data Analytics and Machine Learning for Data Communication Networks, Part of SIGCOMM 2018*, pp.1–7.
- Mulinka, P., Casas, P. & Vanerio, J., 2019. Continuous and Adaptive Learning over Big Streaming Data for Network Security. *Proceeding of the 2019 IEEE 8th International Conference on Cloud Networking, CloudNet 2019*, pp.40–43.
- Oord, A. van den et al., 2016. WaveNet: A Generative Model for Raw Audio., pp.1–15. Available at: <https://arxiv.org/abs/1609.03493>.

at: <http://arxiv.org/abs/1609.03499>.

- Pinheiro, N. et al., 2016. Can PPG be used for HRV analysis? *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2016-Octob, pp.2945–2949.
- Plews, D.J. et al., 2013. Training adaptation and heart rate variability in elite endurance athletes: Opening the door to effective monitoring. *Sports Medicine*, 43(9), pp.773–781.
- Powers, S. & Howley, E., 1995. *Exercise Physiology: Theory and Application to Fitness and Performances*,
- Qiu, S. et al., 2017. Heart rate recovery and risk of cardiovascular events and all-cause mortality: A meta-analysis of prospective cohort studies. *Journal of the American Heart Association*, 6(5).
- Shaffer, F. & Ginsberg, J.P., 2017. An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health*, 5(September), pp.1–17.
- Sharma, P., Imtiaz, S.A. & Rodriguez-Villegas, E., 2019. Acoustic Sensing as a Novel Wearable Approach for Cardiac Monitoring at the Wrist. *Scientific Reports*, 9(1), pp.1–13. Available at: <http://dx.doi.org/10.1038/s41598-019-55599-5>.
- Suzic Lazic, J. et al., 2017. Heart rate recovery in elite athletes: the impact of age and exercise capacity. *Clinical Physiology and Functional Imaging*, 37(2), pp.117–123.
- Tamura, T. et al., 2014. Wearable photoplethysmographic sensors—past and present. *Electronics*, 3(2), pp.282–302.
- Zakynthinaki, M.S., 2015. Modelling heart rate kinetics. *PLoS ONE*, 10(4), pp.1–26.