# Survival Analysis Using Time-Frequency Analysis of Heart Rate Variability During Exercise

**Carl Sjögren**

**2019**

**Master Thesis in Biomedical Engineering**

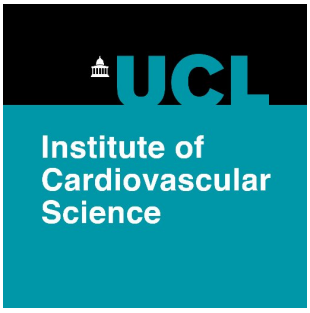
**Supervisor: Martin Stridh**

LUND UNIVERSITY

LTH

FACULTY OF ENGINEERING

**Department of Biomedical Engineering**



# Acknowledgements

This thesis project was conducted at University College London's Institute of Cardiovascular Science, in London, during September 2018 - March 2019. This collaboration would not be possible without Leif Sörnmo's advice when I asked for contacts in London.

I would also like to thank Martin Stridh, Associate Professor at LTH's Department of Biomedical Engineering, for his supervision and feedback with respect to work progression and report structure et cetera. A second pair of eyes was necessary when evaluating project aims, setting realistic goals and proof-reading the final thesis report.

At UCL I have received extensive and insightful support from my local supervisor Michele Orini, Research Associate at UCL. This thesis would never have finished if not for his many hours of trawling through code for finding bugs, mathematical guidance and overall supportive spirit. As I did this project on my own, the ability to discuss my ideas with someone with a firm grasp of the theory and a vested interest was invaluable.

Finally, I would like to extend my thanks to Christine Lehane, Associate Researcher at Copenhagen University, for help with statistics, inspiration for analyses and thesis proof-reading. I would also like to thank my colleagues at UCL who made it an enjoyable environment for me to work in.

Carl Sjögren

# Abstract

## Theory

Heart rate variability as an indicator of increased morbidity has been established in previous studies. It is defined by different frequency bands, corresponding to different biological mechanisms. This thesis aims to study heart rate variability indices in the time-frequency domain through extraction from the UK Biobank dataset.

## Method

Time-frequency indices for different time intervals during exercise were extracted. A time-frequency based filter analyzing the presence of high frequency noise was developed as a way of detecting and removing noisy sections of the ECG recordings. This noise reduction algorithm was coupled with time-domain noise reduction methods as a part of pre-processing the data before survival analysis. Intra-individual repeatability of indices was calculated and finally a combination of a random survival forest and Cox proportional hazards regression was used to evaluate said indices' value as indicators of cardiovascular risk.

## Results

The noise reduction performed by the combination of time and time-frequency domain noise rejection algorithms did not improve intra-patient repeatability. Selected time-frequency indices show high levels of intra-individual repeatability over a three year period. Out of the selected indices, one remained a significant predictor of cardiovascular event when combined with demographic data.

## Conclusion

The use of time-frequency based noise filtering shows promise for detection of high frequency noise artifacts, and should be further studied and tested on other kinds of recordings. The significant relationship between one time-frequency index of heart rate variability and increased risk of cardiovascular events needs to be further investigated.

# Contents

# 1 Introduction

The study of heart rate variability (HRV) is a popular field in biomedical signal processing. Heart rate variability is the term used to describe changes of the time intervals between heart beats, and has been shown to decrease as an individual ages, [1]. Several biological mechanisms in the human body add to this interval change, including blood pressure adjustment and mechanical effects from breathing, [2]. Studies have linked a reduction in heart rate variability with increased morbidity, [3], and can notice changes in patients suffering from post traumatic stress disorder, [4]. As such, the study of HRV allows us to gain insight into how our heart is being regulated and how it is behaving. It is ever changing, and can be affected during confrontation, stress or after having a big meal for lunch, but it is also affected by our breathing and the time of day. Simply put, our daily lives are reflected in our HRV, and it can easily be measured with non-invasive methods such as electrodes on the skin or pulse oximeters.

The work in this thesis is based on the UK Biobank Dataset, more specifically the ECG-recordings during exercise bike tests. Approximately 100,000 individuals have taken part in this exercise bike test, together with a small subset of people who have taken it twice. The study of heart rate variability is often confined to limited datasets, rarely containing more than a hundred subjects, which means that generalization of results becomes difficult. This is why a study on a larger scale population is of interest, no study on this scale has been conducted before with respect to HRV changes. The aim of this thesis is to build a pipeline that takes an original ECG-recording during exercise and performs a survival analysis, while evaluating time-frequency indices of HRV changes. Some parts of the pipeline are already in place, such as ECG-thresholding and time-frequency analysis functions, while some were created during the progression of this thesis. In addition to creating a model for survival analysis of exercise ECG-recordings, the reliability of the heart rate variability indices was also evaluated.

Because a person's heart rate changes during exercise, it is fitting to assume that the same applies to his or her heart rate variability. Because of this, time-frequency analysis methods were applied to the data, with the intention of finding features to describe changes in heart rate variability during different exercise stages. Physiological reasons for the phenomenon being studied and mathematical foundations for the performed analysis are described in Section 2. Following this, Section 3 takes the reader from start to finish in the analysis of the heart data, starting in the time domain, going through time-frequency analysis and culminating in a survival model with extracted time-frequency features. The results of this analysis are illustrated in Section 4, and finally discussed in Section 5.

# 2 Theory

## 2.1 The Heartbeat

The human heart pumps thousands of liters of blood each day, and helps to sustain itself and the remainder of the body with nutrients, oxygen and other important elements that the blood carries. This pumping action is performed by contractile cells that contract when triggered. This occurs when the cell's plasma membrane potential reaches a threshold value and leads to a chain of events called an action potential.

During an action potential, that is the electric charge driving muscular contraction, the cell first depolarizes, then reaches a plateau potential and eventually repolarizes through the flow in and out of different ions with positive or negative charge. This phenomenon is illustrated in Figure 1.

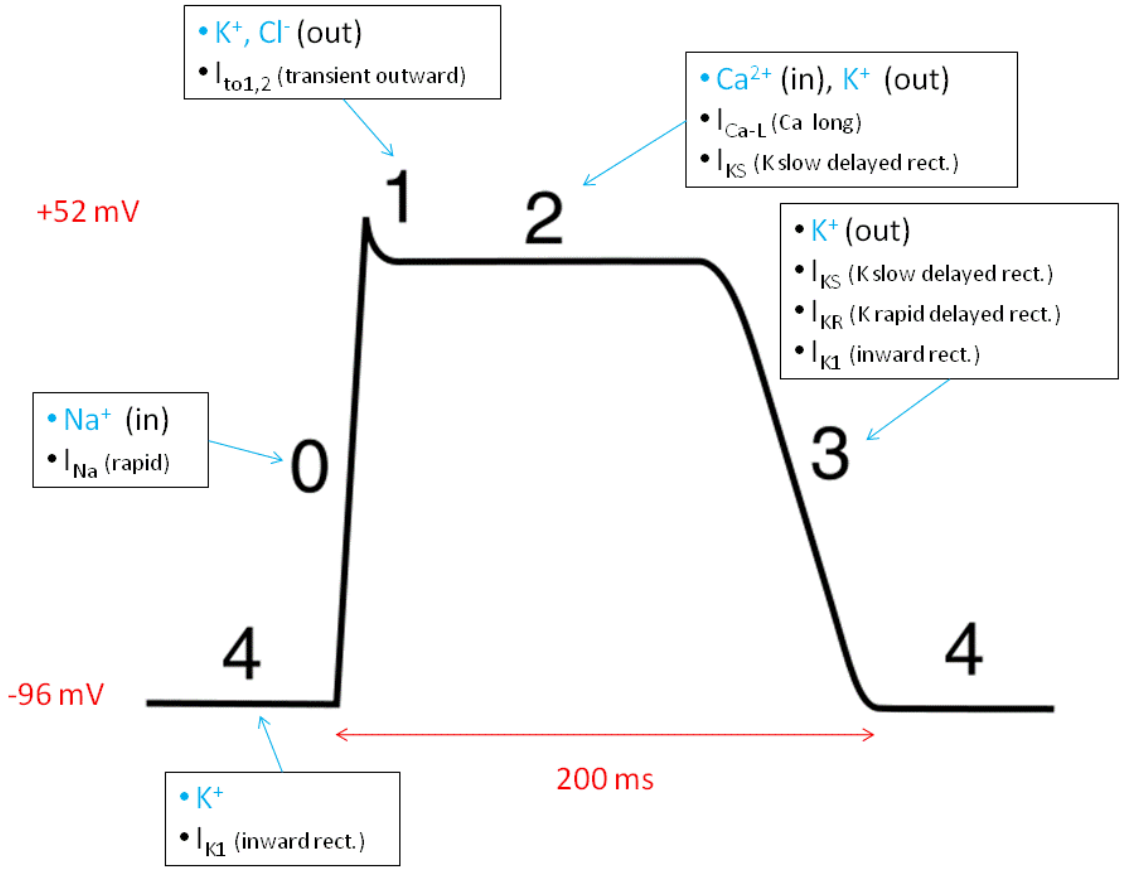



Figure 1: The cardiac action potential, [5].

The duration of this action potential is much greater than that of the normal skeletal muscle action potential, by a factor of 25-30. The contraction of the muscle occurs roughly 150 ms after the cell starts to depolarize.

The heart beats autonomously, and this autonomy is achieved by having nodal cells continuously depolarizing to initialize the spreading depolarization of cardiac muscle cells. A subset of the nodal cells set the frequency with which depolarization occurs, these cells are known as pacemaker cells and are located in the sinoatrial (SA) node and generate action potentials at a rate of about 60-80 action potentials per minute. The contraction signal propagates through atrioventricular (AV) node, where it is delayed for about 100 ms while atrial contraction begins, via conducting cells in the atrial walls. This delay is important as it allows the atria to contract earlier, filling up the ventricles before they contract. The AV node also contains pacemaker cells, and in the event that it does not receive any signal from the SA node, these cells will generate action potentials instead at a rate of 40-60 action potentials per minute.

From the AV node the action potential spreads via conducting cells through the AV bundle (referred to His bundle in Figure 2), eventually forking out into Purkinje fibers that convey the action potentials to cardiac contracting cells in the ventricular myocardium (muscular tissue), [6].

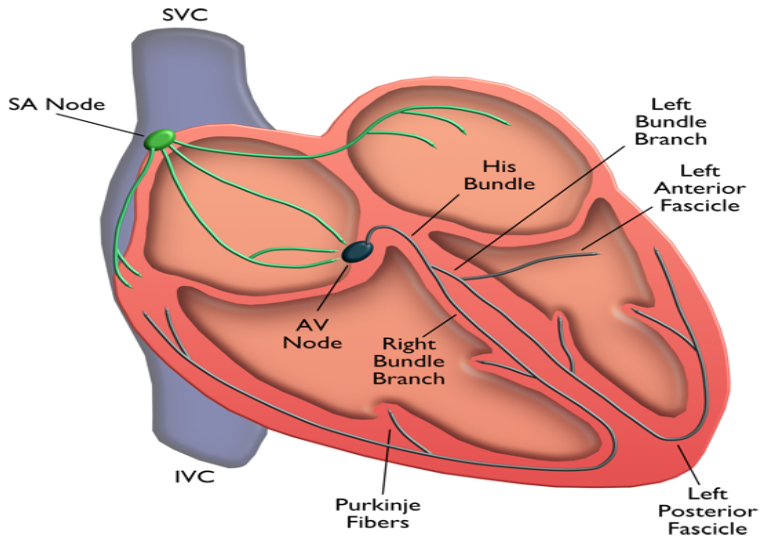



Figure 2: The cardiac conduction system, with the two atria above the ventricles [7].

## 2.2 The Electrocardiogram

The succession of depolarization-repolarization during the heartbeat can be measured with the help of electrodes placed on the surface of the body. The resulting recording is called an electrocardiogram, ECG, and is characterized by the signal displayed in Figure 3. As the figure suggests, there are different points of interest when measuring the ECG.

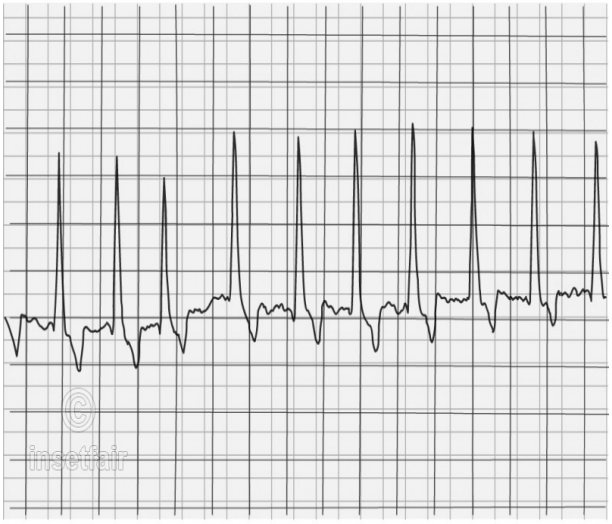



Figure 3: Succession of heart beats over time, [8].

An illustration of the heart beat's ECG morphology is illustrated in Figure 4. At first a P-wave occurs, which signifies depolarization of the atria with contraction beginning around 100 ms after the start of the P-wave. Following this is the QRS-complex, appearing when the ventricles depolarize. The contraction of the ventricles begins briefly after the R peak. Eventually the T-wave is registered, occuring as the ventricles repolarize. Both the timing and the size of these waves are of interest to study, as they can give information about reductions in heart muscle mass or faulty conduction, [6].

## 2.3 Heart Rate Regulation

The human body has control mechanisms that are not consciously controllable, referred to as the autonomic nervous system. In this system we have two subsystems: the sympathetic and the parasympathetic nervous systems. These two systems work constantly to try and keep the body in a steady state, which also applies to the heart. The sympathetic system's effects on the heart include increased heart rate, force of contraction

and increased arterial pressure. The parasympathetic system on the other hand has the opposite effect, reducing heart rate, contraction power and arterial pressure, [1].

### 2.3.1 The Sympathetic Nervous System

The sympathetic nervous system controls among other things the heart, blood vessels and adrenal medulla. When necessary, the sympathetic nervous system can stimulate production of epinephrine (adrenaline) in the adrenal medulla, constrict blood vessels which increases blood pressure, and increase the blood volume pumped with each heart beat or alternatively increase heart rate. If this system is activated too often it is considered pathological. This can be observed with regard to the occurrence of sympathetic bursts in patients with heart failure, where the occurrence can be more than twice that in healthy individuals.

But an overactive sympathetic nervous system is not only a result of cardiac disease. Sympathetic nerve activity also connects to kidney disease, type two diabetes, obesity, depression and more. In other words, increased sympathetic nerve activity is related to some of the main causes of death in our society, [9].

### 2.3.2 The Parasympathetic Nervous System

The parasympathetic nervous system is active during a state of rest, for example when sleeping. It also takes part in reducing heart rate by inhibiting the sympathetic nervous system and hyperpolarizing sinus nodal cells. The parasympathetic nervous system has essentially no effect on blood pressure in the body, with the exception of an increase in heart rate, [6].

## 2.4 The Baroreflex

The baroreflex, baro being latin for pressure, is a negative feedback system that works to maintain normal blood pressure levels using the sympathetic and parasympathetic nervous systems. The system relies on baroreceptors placed in the aortic sinuses, the walls of the carotid sinuses and in the wall of the right atrium. These pressure sensors trigger baroreceptor reflexes when blood pressure levels exceed or fall below normal values. If a high pressure is registered, the baroreceptors will have an increased output to the medulla oblongata in the medulla. An increased baroreceptor output leads in turn to inhibition of the cardioacceleratory and vasomotor centers of the brain, while stimulating the cardioinhibitory centers. This leads to a reduction in the rate and power with which the heart pumps blood and a dilation of peripheral arterioles, i.e, the arterioles in the limbs. A reduction of flow and increase in volume within which blood can flow thus lead to a reduction in overall blood pressure, [6].

In the opposite case, when the blood pressure levels drop below normal levels, the baroreceptor output is reduced and triggers opposite effects. Cardiac output increases and is combined with a constriction of the arteries, leading to an elevation in blood pressure. This reaction to decreased pressure levels is much slower than that of the

reaction to an increase in pressure level, [1].

## 2.5 Heart Rate Variability

These sympathetic and parasympathetic mechanisms contribute to the phenomenon that this thesis is studying, namely HRV. In studies where HRV is of interest, it is common to use the time between R-waves, also called RR-intervals, as a descriptor of the time interval between heart beats. Heart rate variability is the change in length of time in these R-R intervals. An example of an RR-interval can be seen in Figure 4.

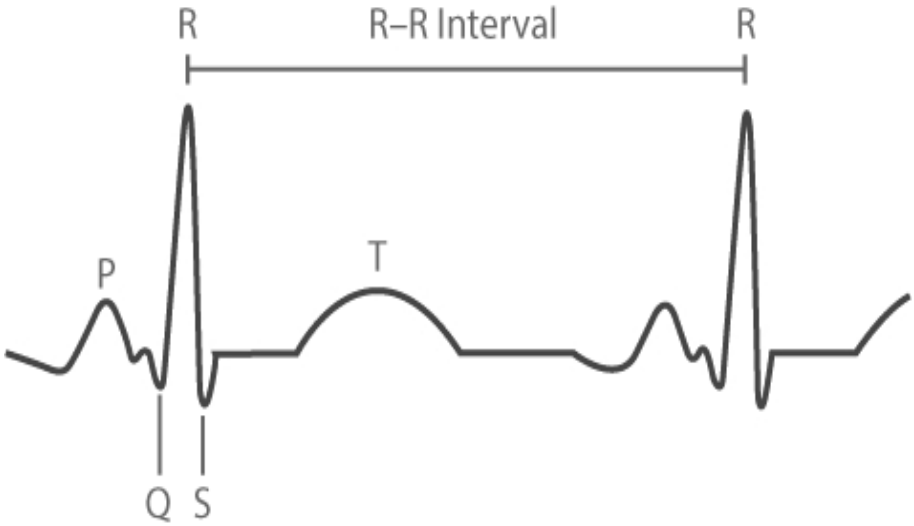



Figure 4: The RR-interval along with annotations of the P- and T-waves and the QRS complex, [10].

By studying somebody's heart rate variability one can gain insight into how the individual copes with sudden physical stress, for example exercise. If a person's heart rate variability is very low, the consequence would be a lacking increase in heart rate when e.g. climbing stairs, which in turn leads to dizziness and fatigue. At a resting heart rate, there is more time between heart beats than during exercise. This also means that there is more room for variation of time between the heart beats, whereas during a high heart rate the time between beats is shortened and thus the variability is lowered in time accordingly.

When studying heart rate variability one examines different frequency bands, which correspond to different physiological origins. The frequency bands describe the changes in RR interval time, over time. They are as follows (in Hz):

$$\text{very low frequency (VLF): } 0.0033 \leq f < 0.04 \tag{1}$$
$$\text{low frequency (LF): } 0.04 \leq f < 0.15 \tag{2}$$
$$\text{high frequency (HF): } 0.15 \leq f < 0.40. \tag{3}$$

The very low frequency band's correspondence to physiological mechanisms is not well defined, but it has been shown to correlate with low testosterone, and is also associated with arrhythmic death, [11], and post-traumatic stress disorder, [4]. The lower limit of 0.0033 Hz corresponds to slightly over five minutes long periods of variability. As this thesis is concerned with exercise bike tests with a duration of about seven minutes, too short for VLF analysis, this band is of no interest and will not be explored further.

The low frequency band (7-25 second periods) is known as the baroreceptor range, where the spectral content corresponds to baroreceptor activity[12]. As both sympathetic and parasympathetic activity have an effect on the power in the LF band, together with the rate of breathing during low respiratory rates, one can expect changes to occur here during exercise tests. The high frequency band (2.5-7 second periods) is often referred to as the respiratory band, as it corresponds to the heart rate variations induced by respiration. The effect of breathing on heart rate is known as respiratory sinus arrhythmia, [2].

## 2.6 Heart Rate Variability and Disease

There have been many studies connecting HRV with heart disease, as well as other diseases, where a reduced heart rate variability can act as a predictor of increased lifetime risk of developing heart disease. A recent study by Kubota et al, [13], found that several characteristics of HRV were independently associated with an increased risk of developing heart disease. However one needs to bear in mind that there are several other risk factors for developing heart disease that need to be taken into account as well, such as smoking status and BMI, [14].

The study of heart rate variability and its associations with heart disease and mortality is not a new phenomenon. Kleiger et al, [3], studied patients who had suffered from acute myocardial infarction and looked at the standard deviation of their RR-intervals in relation to the mean RR-interval for each patient as a measure of HRV. It could be shown that for patients with a heart rate variability below 50ms, the chance of survival 3 years after the infarction was 50% while the group with a HRV above 100ms had an 85% survival chance.

Reduction or increase in HRV is also linked to non-cardiac disease, such as 30-50% of diabetics having a lowered amount of heart rate variability both in the lower and higher

frequency bands. Another case is sufferers of irritable bowel syndrome (IBS) having increased levels of heart rate variability, [15, p. 242-249].

## 2.7 Ectopic Beats

With increasing age comes increasing prevelance of exercise-induced ventricular premature beats, but these themselves have no link to morbidity and thus are no indication of the need for therapy, [16]. However, an excessive amount of ectopic beats is linked to increased risk for heart complications such as atrial fibrillation, [17].

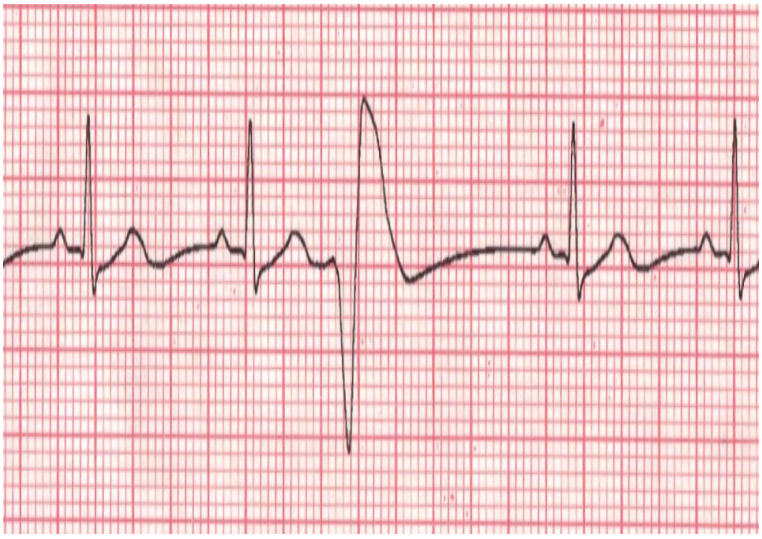


Figure 5: Example of a PVC registration with its deviating appearance, [18].

Ectopia within medicine is a term usually used to describe when something is out of place, for example a hair or a gene expression. Similarly an ectopic beat is a heart beat that occurs out of place, either due to a premature ventricular contraction, or due to a premature atrial contraction. The premature ventricular contraction (PVC) gives rise to an early registered contraction followed by a compensatory delayed one. This can be seen in Figure 5. There are different reasons for a PVC occurring, where it can either be due to cells spontaneously depolarizing, action potentials not being conducted or premature triggering by action potentials, [19].

Premature atrial contractions (PAC) occur when a region of the atria depolarizes before the SA node. The depolarization spreads through the bundle of His like an SA node depolarization would, resulting in a normal looking ECG measurement. Like the PVC, a premature atrial contraction leads to a delayed following heart beat. The spontaneous depolarization is spread through the bundle of His and leads to a premature contraction, but it also reaches the SA node and resets it, [20]. This means that the

time between the PAC and its following contraction is the time for a normal heart beat plus the time it takes for the spontaneous impulse to travel to the SA node, as can be seen in Figure 6. The rate of occurrence of PACs and PVCs increases with age, and are


Figure 6: Example of PAC occurrences with delays following beats, [21].

also more frequent when an individual is performing a physically straining task.

## 2.8 The Fourier Transform

When looking at a measured signal, discerning the existing frequencies in said signal can be difficult. A solution to this problem is the *Fourier transform*, which is described in (4). In a real-world situation, where a signal is not continuously measured but instead measured in discrete time, the discrete time Fourier transform (DTFT) is used as described in (5). See Figure 7 for an illustration of the change between time-dependent and frequency-dependent functions, in this case the step between a sinusoidal function with frequency $1Hz$, [22].

$$\mathscr{F}\{g(t)\} = G(f) = \int_{-\infty}^{\infty} g(t)e^{-2i\pi ft}dt \tag{4}$$

$$\mathscr{F}\{g(n)\} = G(f) = \sum_{n=-\infty}^{\infty} g(n)e^{-2i\pi fn} \tag{5}$$

As can be seen, the transform takes a function (signal) which is time-dependent, and transforms it into a function of frequency. It is then possible to go back to the original time dependant function by applying the *inverse Fourier transform* shown in (6), [24].

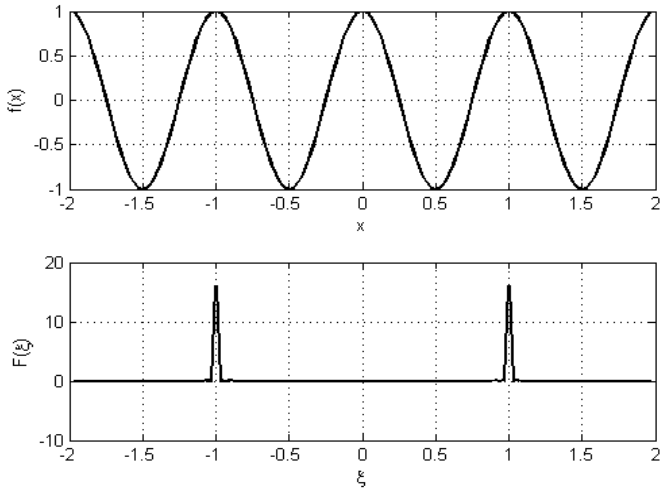



Figure 7: The Fourier transform of a 1Hz sinusiodal, [23].

$$\mathscr{F}^{-1}\{G(f)\} = g(t) = \int_{-\infty}^{\infty} G(f)e^{2i\pi ft}df \tag{6}$$

However, when looking at Figure 7 one can see that the connection between frequency and location is not clear. In the illustrated example, the sinusoidal signal is stationary (does not change over time), making frequency localization easier. When looking at signals from heart measurements, this assumption of stationarity can not be made.

The Fourier transform yields a good indication regarding existing frequencies, however the localization of said frequencies in time is not possible. When studying RR-intervals it may be of interest to know when frequencies occur as well, which is not possible when using the Fourier transform for frequency analysis. To accommodate this need for time localization are several different transforms, and the most natural progression from the Fourier transform would be the short time Fourier transform (STFT).

## 2.9 Time-Frequency Analysis

The localization of frequencies in time is important for understanding of processes occurring in the system that is being measured. It also helps give a clearer picture of what is going on in the system at certain time points. If time-domain measures (the original signal) signifies when a sound is made, and the Fourier transform helps reveal which sounds are made, then the time-frequency representation would be the equivalent of the notes for a song.

It would be appealing to find a measure of instantaneous frequency, i.e., which frequencies that are present in every time segment. In this section, methods for finding the instantaneous frequencies are described, as well as the limitations of this measure.

### 2.9.1 Instantaneous Frequency

When describing the concept of instantaneous frequency, let us begin by examining a monochromatic, i.e. never changing in its expression, signal as described in (7).

$$x(t) = a\cos 2\pi f_0 t \tag{7}$$

With a constant amplitude $a$ and constant frequency $f_0$, x(t) is stationary in the wide sense. This means that if we measure over an even amount of periods or if time goes to infinity, the average and variance of x(t) will remain constant. It is possible, and perhaps tempting, to describe x(t) as a function of a time varying amplitude and a time varying frequency:

$$x(t) = a(t)\cos\phi(t). \tag{8}$$

However, after looking at (8), one realizes that there are an infinite amount of $(a(t), \phi(t))$ pairs, since a change in one could be compensated by a change in the other to yield the same values. This problem of infinite solutions can be solved by first instead representing (7) as

$$x(t) = a\cos 2\pi f_0 t = Re\{ae^{i2\pi f_0 t}\}. \tag{9}$$

The amplitude of x(t) is the modulus,

$$|\, x(t)\,| = |\, Re\{ae^{i2\pi f_0 t}\}\,| = \sqrt{a^2(\cos(2\pi f_0 t)^2 + \sin(2\pi f_0 t)^2)} = \sqrt{a^2} = a, \tag{10}$$

and $2\pi f$ is the argument. It is known that the imaginary part, $\sin 2\pi f_0 t$, can be derived from the real part by shifting its phase by $\frac{\pi}{2}$ radians. As $\frac{\pi}{2}$ radians is the same as a quarter's rotation in the complex domain, we say that the real and imaginary parts are in quadrature. This shift of a real-valued signal by the phase of $\frac{\pi}{2}$ is also called the Hilbert transform. The Hilbert transform of a signal x(t) is defined as the convolution of x(t) and $\frac{1}{\pi t}$:

$$H[x(t)] = \frac{1}{\pi} pv \int_{-\infty}^{+\infty} \frac{x(s)}{t-s} ds \tag{11}$$

where pv is the Cauchy principal value. The Cauchy principal value will not be extensively described in this thesis, but is essentially a method for guaranteeing that an integral exists. In the case of (11), the principal value helps us avoid the case where s=t by changing the single integral into two integrals:

$$\int_{-\infty}^{+\infty} \frac{x(s)}{t-s} ds = \lim_{\varepsilon\to 0}\left(\int_{-\infty}^{t-\varepsilon} \frac{x(s)}{t-s} ds + \int_{t+\varepsilon}^{+\infty} \frac{x(s)}{t-s} ds\right). \tag{12}$$

If we wish to create a "complexified" signal from our original signal, we can use the transform in (11) to create the complex part:

$$z_x(t) = x(t) + iH[x(t)]. \tag{13}$$

It is now possible to define a measure of the instantaneous amplitude and frequency from 13, namely

$$a_x(t) \equiv \|z_x(t)\| \tag{14}$$

$$f_x(t) \equiv \frac{1}{2\pi}\frac{d\arg(z_x(t))}{dt}. \tag{15}$$

It lends itself natural to find an equivalent relationship between time and how it depends on frequency,

$$t_x(f) \equiv -\frac{1}{2\pi}\frac{d\arg(Z_x(f))}{df}, \tag{16}$$

which can be visualized as an expression for the appearance of the time for a certain frequency's behavior. This can be thought of as the output of several different frequency channels over time, where each channel only listens to a certain range of frequencies. This is known as the *group delay*.

The complexified signal $z_x(t)$ is also called the *analytical signal*. This analytical signal's interpretation in the frequency domain is

$$Z_x(f) = X(f) + i(-i\,\mathrm{sgn}(f))X(f) = 2U(f)X(f), \tag{17}$$

where sgn is the sign function of f,

$$\mathrm{sgn}(f) = \begin{cases} 1, & \text{if } f > 0 \\ 0, & \text{if } f = 0, \\ -1, & \text{if } f < 0 \end{cases} \tag{18}$$

and U(f) is the Heaviside step function:

$$U(f) = \begin{cases} 1, & \text{if } f \geq 1 \\ 0, & \text{if } f < 0. \end{cases} \tag{19}$$

This simplification becomes apparent when examining what happens when the frequency is negative: the sign function returns -1, which means that

$$Z_x(f_{negative}) = X(f_{negative}) + i(-i*(-1))X(f_{negative}) =$$
$$= X(f_{negative}) + i^2X(f_{negative}) = X(f_{negative}) - X(f_{negative}) = 0.$$

In other words, any negative frequencies in the real signal's spectrum are removed by this complexification. One might think that essentially halving the frequency content of the signal would reduce the information that is in it, but in reality because we now compute both a real and a complex part for each positive frequency, the amount of measure points remains the same, [25, 26].

The instantaneous frequency and group delay gives us representations of how the signal behaves in both the time and frequency planes, which is more detailed than simply using the Fourier transform on a signal. However, we still have not achieved a truly joint time-frequency distribution, where instead of measuring the same thing in two dimensions, we measure in the two dimensions simultaneously.

### 2.9.2 The Short Time Fourier Transform (STFT)

After the second world war, different approaches to solving the time-frequency problem emerged with the invention of the sonagraph. The sonagraph's output was similar to the metaphor suggested as an illustration of the group delay, and was used in radar detection of objects. Several different band pass filters were employed, and as a signal was interpreted by the sonagraph, the power in these filters were sketched out in black on paper. One approach was to decompose signals into a collection of elementary signals, just like how the Fourier transform decomposes a signal into complex exponentials. Another approach, which will be described later, was the use of pseudo-densities to describe the energy behaviour depending on both time and frequency. So the first natural step in this decomposition would be a time localized Fourier transform, [26, p. 50-52].

Realizing the deficiencies of using the Fourier transform, the idea might spring to mind to do a Fourier transform at shorter intervals, rather than the whole signal. By splitting a measured signal into several intervals and performing the Fourier transform for each of the intervals, the localization of frequencies in time can more easily be performed. This interval, or *window*, that is analyzed may be overlapping with neighbouring windows, or it may be completely separate.

After splitting the signal up into a series of windows, each window is processed with the discrete-time fourier transform (DTFT) described in (5). This means that the frequencies found through the transform can be localized in the time domain with increased precision. The continuous STFT in mathematical form can be seen in (20) with the discrete STFT shown in (21). In (20), $\omega(t-\tau)$ is the window function centered around $\tau$ with zero-values outside of a certain width. In (21) the window is instead centered around m. By varying the value of $\tau$ or m, the STFT is performed on the different time locations in the signal, [27].

$$G(\tau, f) = \int_{-\infty}^{\infty} g(t)\omega(t-\tau)e^{-2i\pi ft}dt \tag{20}$$

$$G(m, f) = \sum_{n=-\infty}^{\infty} g(n)\omega(n-m)e^{-2i\pi fn} \tag{21}$$

While the time precision increases using this method, the reduction of window sizes leads to a reduction of possible wave frequencies that can be found. Low-frequency waves will be excluded, and the cutoff frequency is set by the width of the window. If a window is 1 second wide in the time domain, then the longest period, $T_s$, that can be detected would also be 1 second, meaning that the lowest frequency detectable is $\frac{1}{T_s} = \frac{1}{1s} = 1\text{Hz}$. This rigidity means that when applying the STFT one needs to trade time resolution for frequency resolution in an antagonistic tug of war. There exist methods for improvement of the short-time Fourier transform, but they will not be discussed in this thesis as other methods are of greater interest.

### 2.9.3 The Wigner-Ville Distribution

Instead of decomposition, one can choose to look at the energy of the signal of interest. The energy of a signal, $E_x$, can be described by both the original signal and its Fourier transform,

$$E_x = \int_{-\infty}^{\infty} |\, x(t)\,|^2\, dt = \int_{-\infty}^{\infty} |\, X(f)\,|^2\, df. \tag{22}$$

If the energy is the integration of $|\, x(t)\,|^2$ and $|\, X(f)\,|^2$, they can be described as temporal densities for the original signal or spectral densities for its Fourier transform. Instead of having a separate density for time and another for frequency, a joint time-frequency density would be a natural thing to search for. This time-frequency density, $\rho_x(t,f)$, would then be related to the signal's energy by a double integral,

$$E_x = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \rho_x(t,f)\,dt\,df. \tag{23}$$

A further requirement of this representation is that it contains the unidomain energy densities as its marginals, i.e.

$$\int_{-\infty}^{\infty} \rho(t,f)\,dt = |\, X(f)\,|^2, \tag{24}$$

and

$$\int_{-\infty}^{\infty} \rho(t,f)\,df = |\, x(t)\,|^2\,. \tag{25}$$

A tool in the search of this joint density or "instantaneous spectrum", is the Wigner-Ville distribution. The Wigner-Ville distribution (WVD), originally proposed by Eugene Wigner in 1932 for use in quantum mechanics, and repurposed for signal processing by Jean Ville in 1948, [28], is based on the non-stationary autocorrelation function called the instantaneous autocorrelation function or ambiguity function, [29], described in (26).

$$R_{ss}(t,\tau) = s(t+\frac{\tau}{2})s^*(t-\frac{\tau}{2}), \tag{26}$$

where * denotes the complex conjugate of the signal x. The WVD is then the Fourier

Transform of this instantaneous autocorrelation function :

$$W_x(t,f) = \frac{1}{2\pi}\int_{-\infty}^{\infty} R_{zz}(t,\tau)e^{-2i\pi f\tau}d\tau = \frac{1}{2\pi}\int_{-\infty}^{\infty} z(t+\frac{\tau}{2})z^*(t-\frac{\tau}{2})e^{-2i\pi f\tau}d\tau, \quad (27)$$

where z is the analytical signal previously mentioned,

$$z(t) = s(t) + iH[s(t)], \quad (28)$$

and H[] is the Hilbert transform.

From the WVD one can then find the local frequency $f_x(t)$ by computing the "local momentum",

$$f_x(t) = \frac{\int_{-\infty}^{\infty} fW_x(t,f)df}{\int_{-\infty}^{\infty} W_x(t,f)df}. \quad (29)$$

As can be seen in (26), this multiplication of $z(t+\frac{\tau}{2})z^*(t-\frac{\tau}{2})$ yields cross terms if the signal contains several frequency components. These cross terms also mean that the Wigner Ville distribution can have negative values, which disqualifies it from being a probability density function. It is rare that a real-world signal only contains one clean frequency, and as such it is important to take these cross-terms into account. Consider the signal

$$z_{x+y}(t) = x(t) + y(t), \quad (30)$$

where x(t) and y(t) are two analytic signal components of different frequencies. Performing the multiplication in (26) results in

$$z_{x+y}(t+\frac{\tau}{2})z^*_{x+y}(t-\frac{\tau}{2}) = [x(t+\frac{\tau}{2}) + y(t+\frac{\tau}{2})][x(t-\frac{\tau}{2}) + y(t-\frac{\tau}{2})]^* = \quad (31)$$

$$x(t+\frac{\tau}{2})x^*(t-\frac{\tau}{2}) + y(t+\frac{\tau}{2})y^*(t-\frac{\tau}{2}) + x(t+\frac{\tau}{2})y^*(t-\frac{\tau}{2}) + x^*(t-\frac{\tau}{2})y(t+\frac{\tau}{2}), \quad (32)$$

where the final two terms are the resulting cross-terms. Also note the instantaneous auto-correlation functions for both x(t) and y(t) in the first two terms. This yields the Wigner Ville distribution

$$W_{x+y}(t,f) = W_x(t,f) + W_y(t,f) + 2Re[W_{xy}(t,f)], \quad (33)$$

and as such shows a flaw of the use of this method for time-frequency localization, [26, p. 50-59]. As the number of signal components increases, the number of introduced cross-terms will also increase.

## 2.10 The Hamming Window

When looking at a subset of a signal, one uses windowing. Windowing is the process of assigning amplitude to the signal subset of interest, in an effort to reduce spectral leakage. Because the Fourier transform assumes periodicity in the signal being transformed, any signal without this continuity will yield ”false” frequency contents, also referred to as spectral leakage. This leakage also leads to a reduction in the registered amplitude of the frequencies as illustrated in Figure 8.

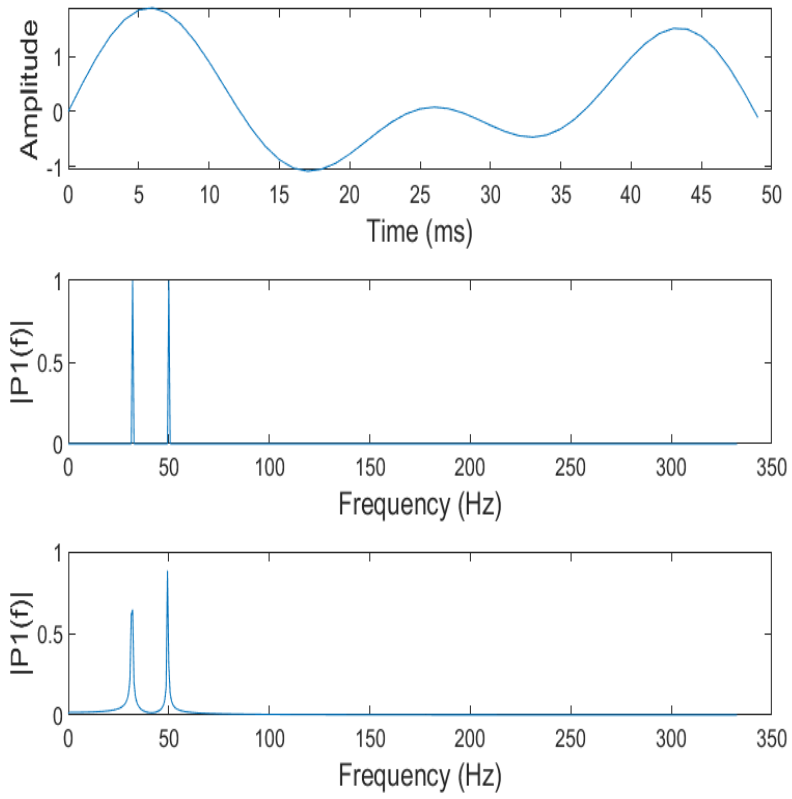



Figure 8: Spectral leakage for a non-periodic signal consisting of two sinusoidals with amplitude one and frequencies 32 and 50Hz, [30]. The upper plot shows the original signal, while the middle plot shows the Fourier transform of an even number of periods. The bottom plot shows the Fourier transform for the same signal with some samples removed to induce irregularity.

For the middle plot, the frequency contents have been computed for a signal with its

length being even periods, i.e. the first sample is the continuation of the last sample. For the bottom plot, however, the signal has had its last samples removed. This creates a discontinuity between the last sample of the series and the first, yielding a less precise estimation of existing frequencies and their amplitudes.

This means that when studying non-stationary signals, such as the signal registered in an ECG-recording during an exercise bike test, the discontinuity of the signal will yield false frequency contents. By using a windowing function where the values at the edges are set to zero (or close to zero), this leakage effect can be reduced, [31].

The Hamming Window, proposed by Richard W. Hamming, is part of the window family known as Cosine-sum windows [31]. They are defined as can be seen in (34),

$$w(n) = a_0 - (1 - a_0)cos(\frac{2\pi n}{N-1}), \quad 0 \leq n \leq N-1, \tag{34}$$

where N is the length of the window. The Hamming window is constructed as in (34), and $a_0$ is set to $\frac{25}{46} \sim 0.54$. As can be seen by looking at the equation and also Figure 9, the edges are not reduced to zero. This is of course unfortunate due to the rise of spectral leakage, but in turn the shape of the Hamming window means a reduction in immediate side lobe amplitude. This makes frequency localization more precise.

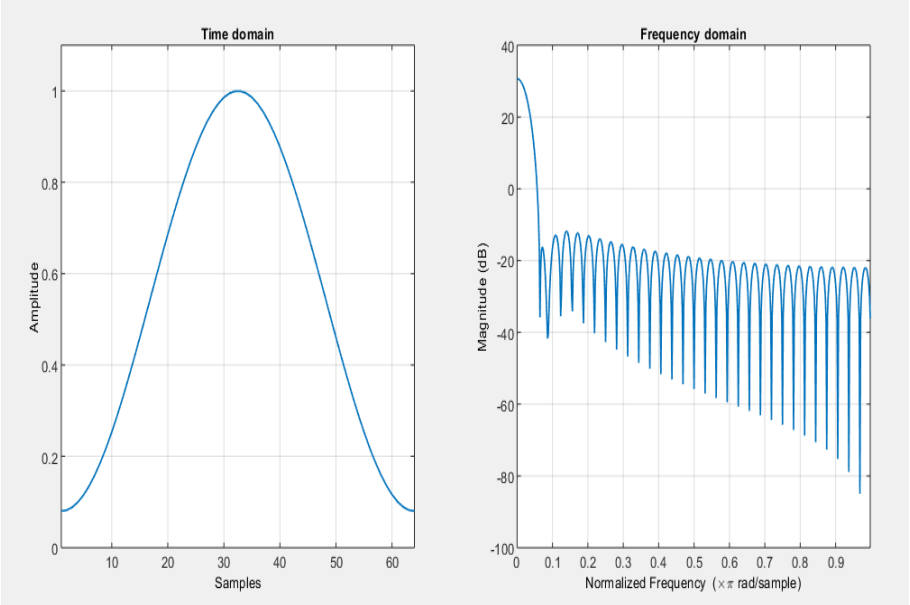



Figure 9: Hamming window of width 64 samples in the time and frequency domain, [32].

## 2.11 Correlation

### 2.11.1 Pearson Correlation

The Pearson correlation is the quotient, ratio, of the covariance of two variables, referred to as x and y in (35), and their respective standard deviations. The value of $\rho$ is an indication of the strength of the correlation, and how an increase in one variable affects the other.

$$\rho_{xy} = \frac{\text{Cov}(x,y)}{\sqrt{\text{var}(x)}\sqrt{\text{var}(y)}} \tag{35}$$

This computation requires assumptions of independent variables, where the value of a certain sample does not influence the value of other samples for that variable, or the values of other variables. It is also assumed that both x and y are normally distributed. The Pearson correlation is a measure of linearity, and as such it can not detect non-linear relationships between variables, [33].

### 2.11.2 Spearman Correlation

Whereas the Pearson measure of correlation is linear, the Spearman rank-order correlation is non-parametric and thus has the ability to give an indication of non-linear relationship between variables. If the assumptions made for the Pearson correlation are violated, one might rather use Spearman correlation. The assumptions made in this case are instead that there is a monotonic relationship between the variables, where an increase in one leads to an increase or decrease in the other. The difference between a linear relationship and a monotonic one is that the rate of change in y when x changes is not constant, as illustrated in Figure 10, [34].

Like in the Pearson example, the Spearman correlation is the quotient of covariance and standard deviation, but in this case one looks at the covariance and standard deviation of the **rank** of the variables' values. The values for the two variables are ranked in descending order, where rank 1 is given to the highest value, and the highest rank is given to the lowest value. This is done individually for both of the variables, and in the case where two (or more) values are the same, they will receive a fractional rank value instead. So if for example the value 10 appears twice and thus takes up both rank 6 and 7, this leads to both samples being ranked as $\frac{6+7}{2} = 6.5$.

After ranking the values for both variables, a familiar division of covariance and standard deviation is constructed and can be seen in (36). $r_x$ and $r_y$ are the respective rank values for variables x and y, [36].

$$\rho_{r_x r_y} = \frac{\text{Cov}(r_x, r_y)}{\sqrt{\text{var}(r_x)}\sqrt{\text{var}(r_y)}} \tag{36}$$

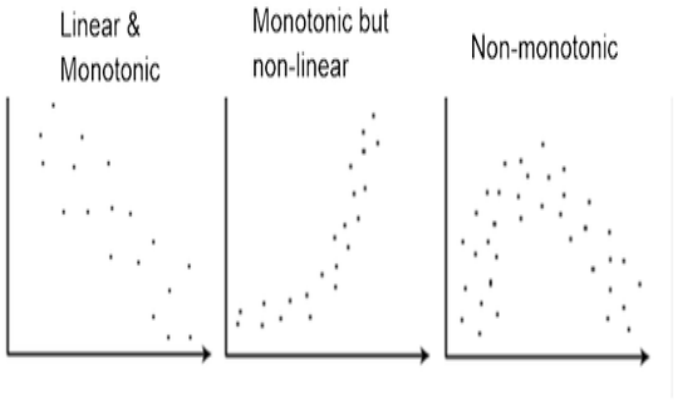



Figure 10: Illustration of linear monotonic, non-linear monotonic and non-monotonic relationships, [35].

### 2.11.3 Intraclass Correlation

Both the Pearson and Spearman correlation coefficients are meant to measure the correlation between two bivariate variables, i.e., two different items. This is useful when investigating whether there is a correlation between, e.g., height and weight, or IQ and school grades. Instead, the intraclass correlation (ICC) is a method for determining correlation for the same item, observed by different raters. Like in the case of Pearson correlation, an assumption of linearity is required for the analysis. The aim is to determine how consistent or reproducible measurements are when measured by different raters. The ICC is also commonly used in studies investigating the reliability of HRV indices [37–40], where the raters in this case would be different times of measurement. The ICC can be used to determine two coefficients. The first of them is consistency which is an indication of whether or not there is a systematic difference between the measurements assuming that there is **no** systematic difference between raters. The second one is absolute agreement, which assumes that there **is** a systematic difference between raters, [41].

In addition to this, there are several different forms of doing the analysis. The different models will not be further described in this thesis, except for the two-way random-effects model. Using this model under the assumption that the measurements do not

interact, the relationship

$$x_{ij} = \mu + r_i + c_i + e_{ij} \tag{37}$$

arises, where $\mu$ is the average for all observations, $r_i$ is difference between raters, $c_i$ is the difference between the columns and $e_i$ is the residual value.

The two-way random-effects model is the best approach if one wants to generalize the reliability results acquired. As an example, if one was to use the two-way random-effects model, with additional assumptions that there is no interaction between the measurements, the ICC consistency coefficient would be

$$\rho = \frac{\sigma_r^2}{\sigma_r^2 + \sigma_e^2}, \tag{38}$$

where $\sigma_r^2$ is the variance of the difference between the raters, and $\sigma_e^2$ is the variance of the residuals, [42].

## 2.12 Laplace Distribution

The Laplace distribution is a probability distribution, with probability density function

$$f(x;\mu,s) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}, \tag{39}$$

where $-\infty < x < \infty$, $\mu \in (-\infty, \infty)$, and $b > 0$. The variable $\mu$ in this case is the location (mean) parameter and b is a scaling parameter that decides the slope of the distribution. An example of the Laplace distribution for different values of b can be observed in Figure 11.

## 2.13 Receiver Operating Characteristic Analysis

Say that a method/algorithm for classification of something has been developed, in the space of biomedical signal processing perhaps to try to detect ectopic beats, described in section 2.7. A thresholding value where it is determined that if over three heartbeats one can see the succession normal-short-long with regards to time between the beats, the latter two beats will be classified as the result of an ectopic beat. How would one go about evaluating this algorithm objectively? What thresholds should be set for the relation between the normal, short and long registered R-R intervals? It is also important not only to detect the presence of ectopic beats, but also to keep the amount of misclassifications of normal heart beats as ectopic to a minimum. To do this in an objective fashion without spending hours manually reviewing the performance of the algorithm, one can use the receiver operating characteristic (ROC) curve.

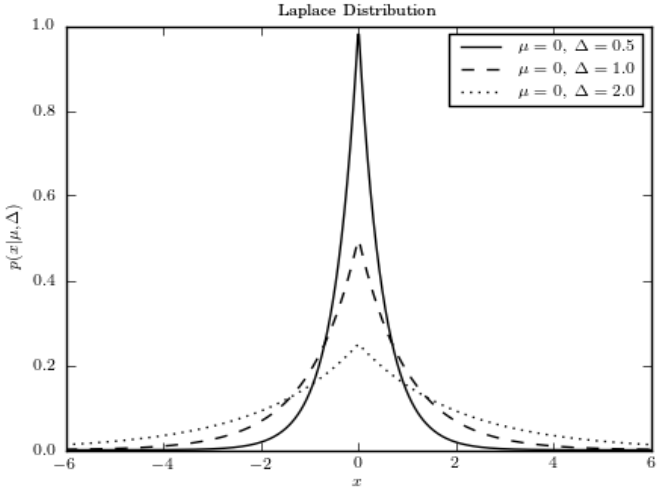



Figure 11: Laplace distribution for different scaling values, b being denoted as Δ in this example, [43].

The ROC curve is a way of evaluating classification problems, in the case mentioned above a binary classification of ”Ectopic beat” or ”Not ectopic beat”. Working under the assumption that the two distributions of normal and ectopic beats overlap, i.e., that no perfect threshold value for differentiation between the two exists, the use of an ROC curve can help decide which threshold to select, and in general describe the classifying power of our algorithm. This method is a supervised endeavour, i.e., in order to find the optimal thresholding value we need to know a priori where the ectopic beats are in the R-R series. Either a real-world signal can be manually annotated for indices of ectopic beats, or ectopic beats can be simulated with their indices registered.

Then the decision on what different threshold values to look at, perhaps it is desirable to try out 50 different values. For each threshold value, the sensitivity and specificity are calculated. Sensitivity can also be described as true positive rate (TPR), while specificity is true negative rate (TNR) or 1-(false positive rate). A true positive (TP) is the case where the algorithm detects an actual ectopic beat, whereas a true negative (TN) is when a normal heartbeat is classified as such. If a normal heart beat is classified as an ectopic beat, it would be called a false positive (FP). In a similar fashion, an ectopic beat being registered as a normal beat is a false negative (FN). These are the values of interest for ROC-analysis, in order to maximize sensitivty and specificity simultaneously. For each threshold value, the specificity and sensitivity is calculated. This is done by comparing the algorithm’s output of ectopic beat indices with a priori known

indices. Then,

$$TPR = Sensitivity = \frac{\#TP}{\#TP + \#FN} \tag{40}$$

$$TNR = Specificity = \frac{\#TN}{\#TN + \#FP}. \tag{41}$$

From these measures it is possible to construct the ROC curve, by plotting the ratio

$$q = \frac{TPR}{1 - TNR} = \frac{TPR}{FPR}. \tag{42}$$

An illustration of an ROC curve can be seen in Figure 12.

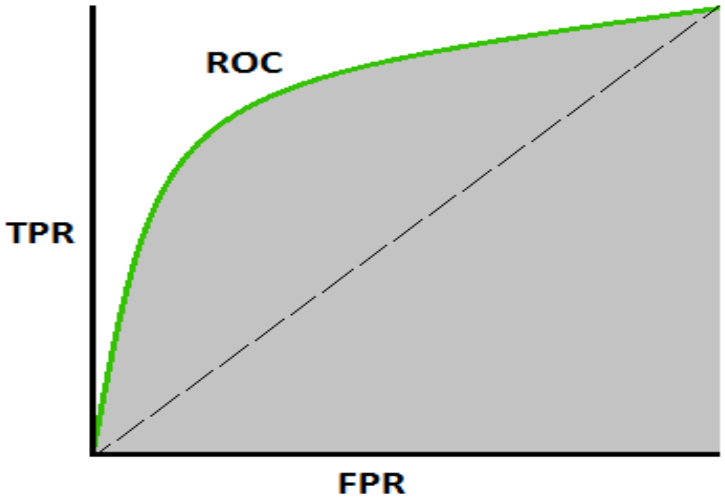



Figure 12: ROC curve with TPR and FPR ranging from 0 to 1. [44]

The green line represents the quotient q for the different thresholding values, and is a good illustration of the trade-off between true positive and false positive rates. The grey area is referred to as the area under the curve (AUC) and describes the predictive capability of our algorithm.

### 2.13.1 Area Under the Curve

The AUC can be used as an over all estimator of predictive capability. As the green line in Figure 12 of the ROC curve approaches the top left corner (1,0), i.e., a 100% TPR with a 0% FPR, the greater the area will be. For a perfect classifier, the AUC

would be 1. For a worthless classifier, the AUC would be 0.5 where the classifier essentially randomly decides between a positive or negative output. For values below 0.5, the predictive capability of the classifier is worse than chance, but by simply using the classifier's worse-than-chance output and classifying all positives as negatives and vice versa, one increases the AUC again. For example, in the (extremely unlikely) case of an AUC value of 0, one can simply reverse all classifications to get an AUC of 1, which is a perfect value.

By selecting the threshold which gives the point on the ROC curve closest to (1,0), one can find the value where the true positive rate is maximized in relation to the false positive rate, [45].

## 2.14 Survival Analysis

Something that is of value in medicine is the ability to classify patients into risk groups. A method for risk classification can be survival analysis, where one uses measured values to estimate time until an event occurs. It can be a literal survival analysis, i.e. predicting time until death, or it can be time until another event such as a heart attack or stroke. In short, a survival analysis model's outcome is a length of time. In order to gain insight into how measured values predict into the future, sufficient follow-up is required during the study, [46].

Patients dropping out of studies or simply not experiencing the event of interest during the study period is called *censoring*. Because there will be incomplete information regarding the subjects' survival time, simply doing linear regression modelling will not suffice to correctly predict it. If the censoring is rare, and random, in the study population, it may be fair to simply remove censored data. This approach, however, does not work if a significant portion of participants are censored[47], and may lead to biased results. Fortunately there are several methods for dealing with this missing data such as Cox regression, which will be described in further detail later in this thesis.

Despite the name of the analysis being survival analysis, an alternative name could just as well be time to event analysis. Instead of modelling time until death or similar, one could apply this reasoning to time until an autonomous car is involved in a car accident or time until someone slips on a patch of ice. In addition to the predictors one chooses to study, it is important to take relevant factors into account such as age, weight, sex so that no bias exists in the analysis, [48].

### 2.14.1 The Survival Function

The survival function is a description of likelihood of survival for different time locations. Let time until event occurrence be denoted by the variable T. We can then consider its probability density function f(t) and cumulative distribution function F(t) = Pr(T $<$ t). Using the cumulative distribution function of F(t) we can describe the

likelihood of survival past a certain time, or the *survival function* S(t), as

$$S(t) = Pr(T \geq t) = 1 - Pr(T \leq t) = 1 - F(t) = \int_t^\infty f(x)dx. \tag{43}$$

There exist several methods for estimation of this survival function, one of which will be described later in this section.

### 2.14.2 Hazard Rates and Ratios

In order to correctly estimate how long time will pass until the event of interest occurs, one needs to have a measure of the hazard at a given time. Again, the name hazard could be replaced by probability of the event occurring at the given time point when applied to studies not involving negative outcomes e.g. probability of winning the lottery, being cured of a disease or finding a mobile phone in between the couch pillows.

Let us use the act of smoking as a potential hazard. It is widely known that smoking is linked to a shortened lifespan, and as such it may be a good predictor to use in survival analysis of individuals. For example, the cumulative *hazard rate* of dying due to bronchitis and emphysema is 17 times higher for male smokers compared to male non-smokers and 12 times higher for female smokers compared to non-smoking females, [14]. We can in this case say that the *hazard ratio* of smoking for men is 17, and 12 for women with respect to dying from bronchitis of emphysema. It is also apparent that the increase in hazard rate is greater for men than women, so in a survival analysis one may need to take sex into consideration as well.

The hazard rate is calculated for a certain index of time, for example perhaps $\frac{10}{1000} = 1\%$ participants die in the first month, and $\frac{15}{990} = 1.5\%$ the second month would correspond to a hazard rate of 0.01 for the first month and 0.015 for the second. By reducing the time interval for which we evaluate hazard, the instantaneous hazard rate or *hazard function* emerges,

$$h(t) = \lim_{dt \to 0} \frac{Pr(t \leq T < t + dt | T \geq t)}{dt}, \tag{44}$$

where T is the variable corresponding to time until the event happens. The probability given in (44) can be described as the probability that the time until event is somewhere between t and t + dt, given that T$\geq$t. As dt approaches zero we achieve a continuous hazard function.[49]

The joint probability that T is within the interval of [t,t+dt) can be expressed using the survival function. For infinitesimally valued dt, the numerator in (44) translates to (roughly) Pr(T = t — T $\geq$t). The probability that T = t is given by the probability density function times the time difference, f(t)dt, and the probability that T $\geq$ t is S(t). Through Bayes' theorem,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \tag{45}$$

we find that h(t) can be rewritten as

$$h(t) = \frac{f(t)}{S(t)}. \tag{46}$$

From (43), we find that the derivative of S(t) is -f(t) and as such we can rewrite (46) as

$$h(t) = -\frac{d}{dt} ln(S(t)). \tag{47}$$

The hazard rate may change over time, for example the hazard of dying due to smoking-related illness increases with age, but it is often assumed that the hazard ratio, i.e. 17 for men as listed above, remains constant. This assumption is only reasonable in certain situations, and as such the need for an approach that can take changing hazard ratios into consideration arises, [50].

## 2.15 The UK Biobank Dataset

The UK Biobank dataset was, and continues to be, constructed with the goal of improving prevention, diagnosis and treatment of numerous diseases in the population. It follows half a million individuals' health development over time. The participants are between the ages 40 and 69 years old, with a roughly even gender distribution. Not all of these participants took part in the exercise bike test(s), [51].

A subset of the half a million participants, roughly 100,000, took part in exercise bike tests combined with recording of electrocardiograms. Before partaking in physical exertion, a maximum workload, i.e. maximal allowed heart rate, was calculated using factors such as e.g. sex, age and weight. If the participant exceeded this workload limit the physical test was to be terminated immediately. In addition to this termination criteria, in the situation where a participant experienced sudden chest pain, felt faint or ill in some other way, the test would be immediately terminated.

Depending on their risk group classification (dependent on participant anamnesis among other things), the ECG recording could take on three different shapes. If everything was in order and the individual was classified as a low-risk participant, the test consisted of four phases. Initially, the participant was to be seated and remain at rest on the exercise bike for 15 seconds. Then the participant was instructed to start pedalling and to keep a pace of 60 rotations per minute (rpm) for two minutes. After the two minutes of constant workload, the workload would begin to gradually increase over the next four minutes. The participant was to remain pedalling at a pace of 60 rpm throughout this workload increase. Finally, after the four minutes of increasing workload, the participant was instructed to stop pedalling immediately, coming to a full stop and remaining silent and still on the bike for an additional minute to record recovery data.

If an individual was deemed to be in the ”medium risk” category, they would perform the exercise bike test for the full duration of 7 minutes and 15 seconds, but instead of increasing workload after two minutes and fifteen seconds they would simply continue at a constant workload for another four minutes. In the case that the individual was deemed to be at a high risk level, they would simply be asked to sit in a chair for two minutes while their resting ECG was taken. This test was repeated after $34.2 \pm 2.8$ months for 1187 out of the original participants, [52].

# 3 Method

This thesis workflow was split up into two stages: data processing, feature engineering and an analysis of intra-individual correlation of features between exercise sessions was performed first. This was followed by a second stage consisting of variable selection via machine learning techniques and finally a Cox proportional hazards survival regression. Overall workflow is illustrated in Figure 13.

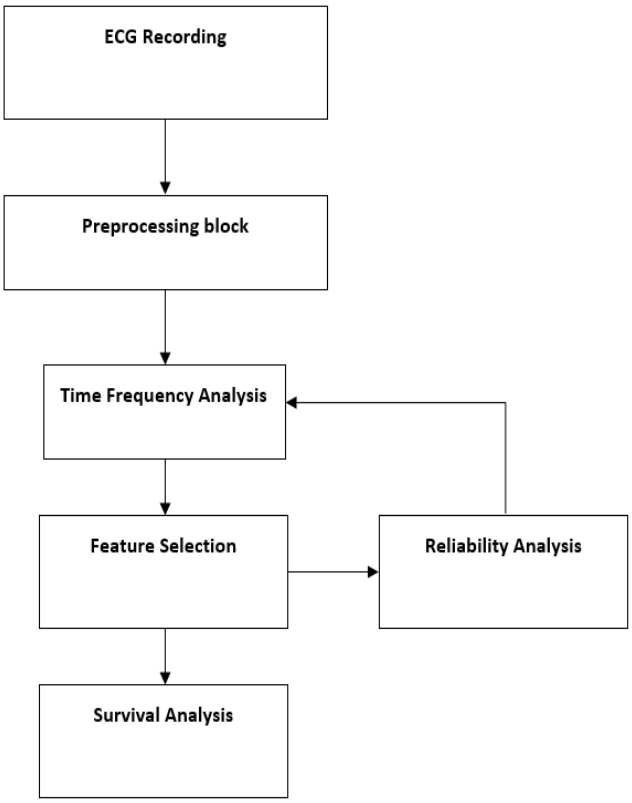



Figure 13: Thesis workflow

Within the different sections of the workflow are several challenges that needed to be addressed. Among these were:

- Transforming raw data to HRV signal
- Removal of artefacts and noise

- Extraction of features
- Evaluation of feature reliability over time
- Feature selection for use in survival analysis
- Survival analysis,

which will all be addressed in this section.

Before any analysis is performed on the recording, recordings that were deemed too short (<5 minutes) were excluded, as they likely contained incomplete data due to for example premature terminations. The preprocessing block first takes the raw ECG recording and translates it into a cumulative RR-series. Then follows a transformation of the initial RR-interval data from a cumulative form to a series of RR-intervals. In the signal processing block, the data is checked for outliers and transformed to a continuous time-varying signal for analysis. In the time-frequency block, the transformed signal is analyzed using a variant of the Wigner-Ville distribution, combined with a kernel operation, and features are selected. For a smaller subset of participants, reliability of said features is analyzed, informing the final time frequency feature selection. Lastly, a survival analysis is performed on the bigger data set after using a random survival forest to perform variable selection.

## 3.1 Pre-processing Block

Initially, an exploratory project looking at repeated exercise bike tests and the time-frequency feature repeatability was undertaken. The reasoning for this approach was two-fold: First reason being that it is overwhelming to start analyzing a dataset of 100,000 individuals without first formulating a step by step plan for analysis. Secondly it is important to know if the features being selected for survival analysis are reliable, meaning that the features do not differ a lot in between exercise bike tests (EBTs). As can be seen in Figure 13, analysis of the EBT-data consists of several stages, where the signal is processed and analyzed with different methods. For this section, repeated EBTs with roughly three years in between were examined, [52].

The thesis work was based on already processed recordings, i.e. the ECG had already been analyzed. After being translated from ECG-recordings to an RR-series, the signal initially consists of cumulative time elapsed (in milliseconds) for each detected R-peak with respect to the start of the EBT. In order to get the signal in RR-interval series format, the series is differentiated,

$$RR_{diff}(i) = RR(i) - RR(i-1). \tag{48}$$

This takes the series from the upper to the lower plot illustrated in Figure 23, shown in the results section.

The signal at this stage is described as time difference between heartbeat$_i$ and heartbeat$_{i-1}$. In order to perform time-frequency analysis, we would rather have a signal spread out

over time. It also is not possible to extract frequency behavior from this signal, as the time between each sample varies over the course of the EBT. In order for the signal to spread out over time instead of samples, the cumulative time in the original RR series is upsampled at 4Hz, so that we have four samples per second. Our signal series x(n) is then linearly interpolated with the upsampled time, $t_{int}$, to yield $x_{int}$ with estimated heart rate every 0.25 seconds. Now the relationship between heart rate and time, illustrated in Figure 14, is more intuitive to look at. The interpolation also means that we now have equidistant samples, which means we can start extracting frequency information from the series.

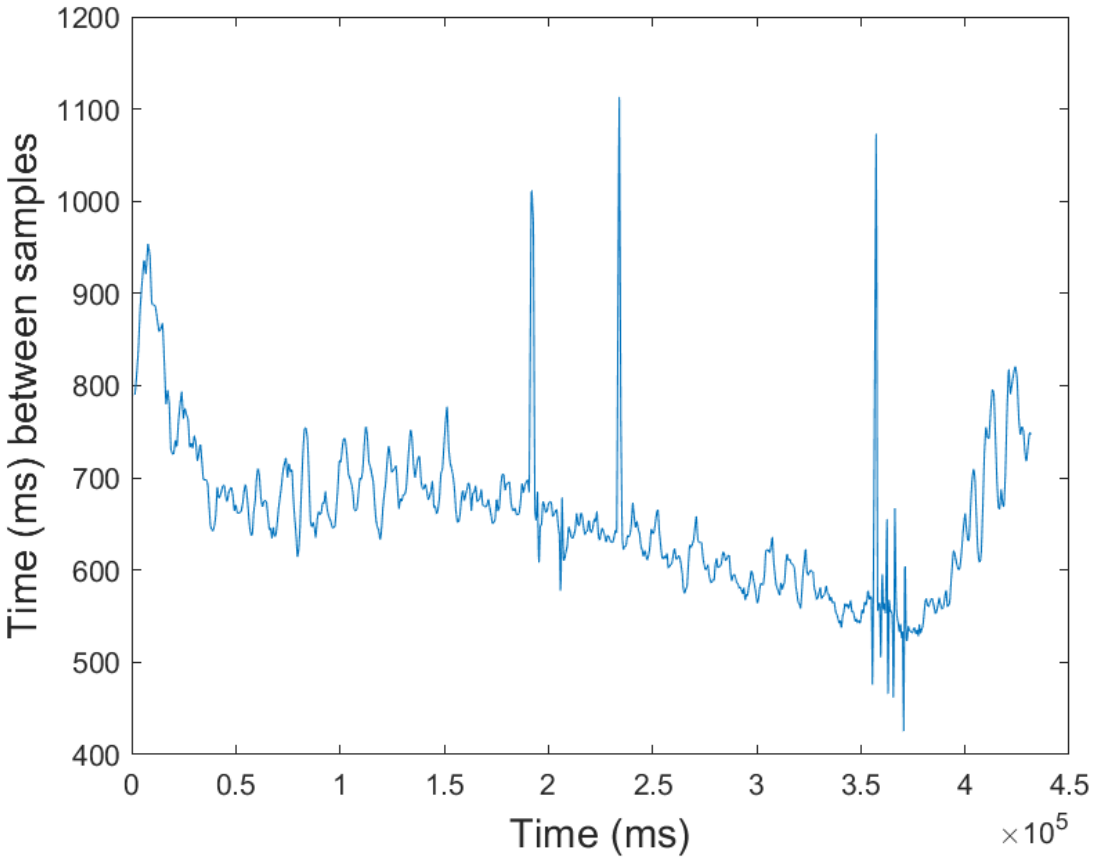



Figure 14: Interpolated RR interval series

As can be seen in Figure 14, there are still very clear trends in the data. If one is to measure changes in heart rate variability for an individual, the existence of a trend

would skew the data. Of course, during an EBT the individual's heart rate will increase over time as the workload increases, but the changes in heart rate are not of interest but instead the changes in heart rate variability. In order to estimate this trend, a low-pass Hamming window is used on the signal. This yields the smooth red curve seen in Figure 15, plotted together with the signal in blue. The trend is then removed by subtraction,

$$x_{varability} = x_{int} - H[x_{int}], \tag{49}$$

and the result can be seen in the lower part of Figure 15. The reason why the smoothed average is removed is to remove the influence of long-term trends on instantaneous measures. This means that some information may be lost as a result of trend removal, but is outweighed by the reduction in trend influence.

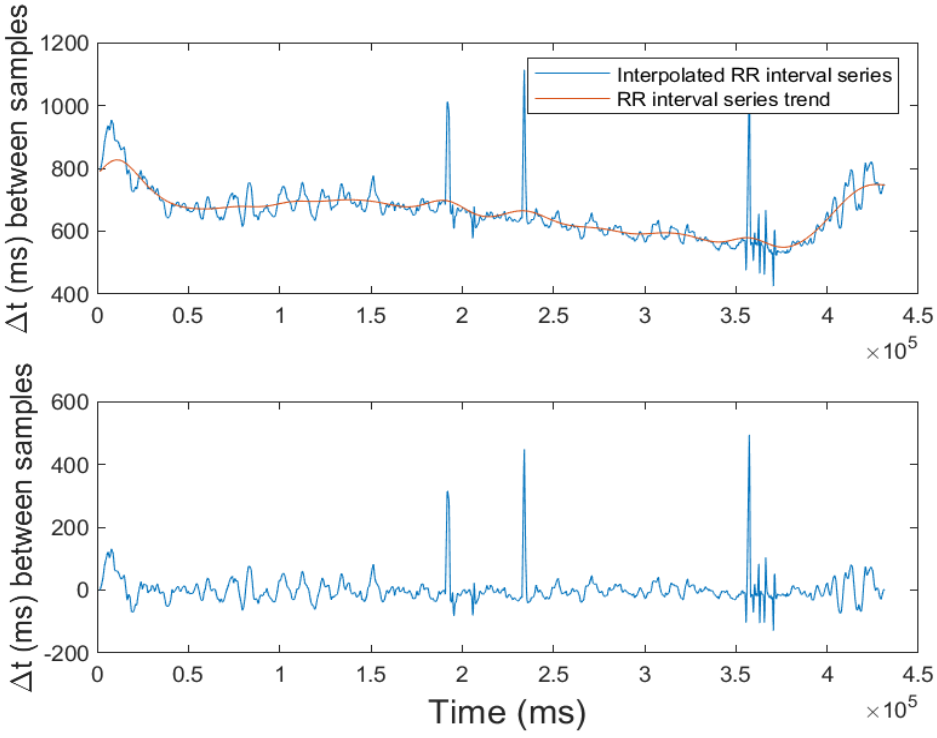



Figure 15: Trend removal of RR interval series

### 3.1.1 Time Domain Noise Reduction

It may happen that the R-R intervals contain skewed information. For example, if the function that detects R-peaks in the ECG data misses one or even several peaks, the time between the first detected and the delayed second detection will be double, or a larger multiple, that of the previous one. In a plot of R-R intervals this would express itself as a sudden, sharp peak, with the interval series then returning to normal afterwards. In a similar fashion, false detections of R-peaks e.g. due to noise in the recording, will result in a sharp and sudden dip. This time domain noise reduction was performed prior to time interpolation.

These peaks and dips can occur with varying results in time-domain changes. It is also important to not be too harsh when searching a signal for outliers, as some may have biological origin. Therefore, when developing a threshold algorithm for detection of these artifacts, extensive ocular inspection of detected outliers was conducted to verify their validity.

In addition to this, a normally occuring biological ”artifact” is that of ectopic beats, see Figure 16. Similar to premature and delayed detections this will give rise to a sudden spike and then normal signal following. The difference between ectopic beats and premature/delayed detections is that the morphology will be a dip followed by a peak, opposed to a single peak or dip. The short and long intervals should together amount to roughly twice that of the normal heart beat time. With this in mind, a thresholding algorithm was implemented to detect and interpolate ectopic beats.

For this purpose, a median filtered R-R interval value was calculated with the median filter being of width 12 samples, 12 being chosen empirically, centered around the interval of interest, $RR_i$. $RR_i$ is then compared to the moving median value for its index. If it is 90% or lower than the median value while its following R-R interval is 110% or higher than the median for i+1, $RR_i$ is deemed to be a premature beat. The threshold limits were decided through empirical evaluation. In order to rid the signal of this noise, a spline interpolation is performed. A similar but opposite morphology was also consistently observed as a ”reversed ectopic” beat, see Figure 16, with an initial long followed by a short beat. This shape arises from one delayed (i.e. missed) detection, followed by a correct detection. The same thinking was applied as for ectopic detection, in a reversed fashion.

The ”reverse ectopic” is just one mis-detection artifact among many, and as such a general purpose algorithm was created to handle all other time-domain noise reduction. Much like in the ectopic beat case, beats deviating by in this case 10% from the median value in any direction are deemed to be abnormal. The aim of this algorithm is to catch any other discrepancies that may have arisen due to poor ECG-processing, such as missed beats or double registrations. After having checked all beats to decide whether or not they are abnormal, the resulting classification is checked to see whether or not intervals of several abnormal beats in a row exist. These intervals may for ex-

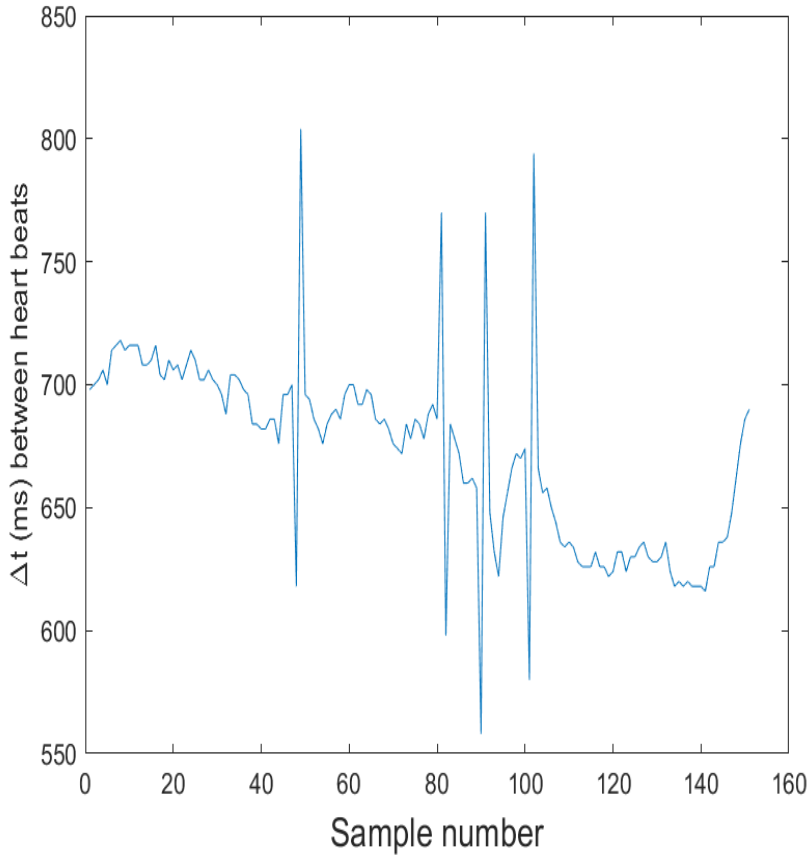



Figure 16: Signal containing different kinds of noise (ectopic beats and ”reverse ectopic” morphology), addressed by the time-domain noise reduction algorithms.

ample consist of two short R-R intervals, resulting from e.g. classifying a T-wave as an R-peak and then detecting the real R-peak.

The total time that these intervals last for is summarized, to see if the remainder after dividing the sum with the median interval value is within 10% of the median R-R interval. If it is, then, similar to the ectopic case, the time is split into evenly sized intervals and a spline interpolation is performed. This may result in an increase or decrease in the total number of registered heart beats in the R-R interval series. This time-domain noise correction is performed on the raw RR-intervals.

## 3.2 Time-Frequency Analysis

The original ECG recording has gone from its original form to a HRV signal that is ready for time-frequency analysis. This is performed by using the Wigner-Ville distribution described in section 2.9.3, with some modifications. The analytical signal is first computed by performing the Hilbert transform on the HRV signal. The analytical HRV signal is analyzed using the tfrwv command from Time-Frequency Toolbox, developed by François Auger and others, which computed the Wigner-Ville distribution of the signal. This yields our initial time-frequency distribution, but containing cross-terms which make localization in both time and frequency difficult. Luckily, there exist methods for reducing these cross-terms.

### 3.2.1 Feature Extraction

First the ambiguity function is computed from the Wigner-Ville transformed signal. Then an exponential kernel, [53], is applied with parameters

$$\beta = 1 \tag{50}$$
$$\gamma = 1 \tag{51}$$
$$\alpha = 0 \tag{52}$$
$$r = 0, \tag{53}$$

in order to reduce the influence of the WVD crossterms. This is done by convolution of the kernel and the time-frequency representation, where values are averaged out resulting in removal of weak terms and enhancement of strong ones. From the resulting time-frequency distribution, instantaneous power for the duration of the EBT is extracted.

The extracted instantaneous power spectra are divided into five time intervals. The first interval is 15 seconds long, and encapsulates the rest time before the exercise bike test begins. The features extracted from the resting interval are Rest_med_low, Rest_med_high and Rest_med_tot, signifying the **med**ian instantaneous power in the low, high and total frequency bands during rest. The next interval stretches from t = 15s to t= 135s, where the cyclist is subjected to a constant workload (WLc). After this, the third interval stretches from 135s to an estimated end of exertion and corresponds to a section of increasing workload (WLi). The estimated end is computed as the time at which the trend of the interpolated RR series, $x_{int}$ above, reaches its maximum. The need for this computation arises as the exercise does not end after exactly four minutes of increasing workload. The fourth interval is defined as the remaining time after exertion ends, during which the individual recovers (Rec). Median power for the low, high and total frequency bands was also extracted for the entire EBT length (All), constituting the fifth interval.

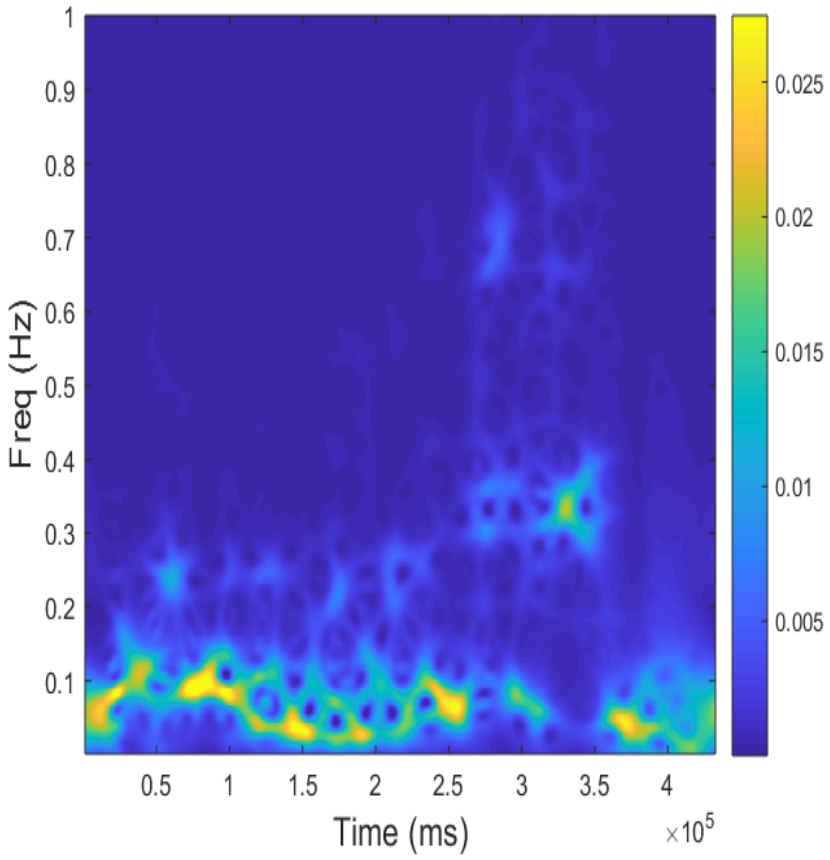



Figure 17: Time-frequency distribution for the trend removed RR interval series

After splitting the power spectrum into four interval, six different features were extracted for each interval: median and mean instantaneous power in the low, high and total frequency bands.

### 3.2.2 Time-Frequency Domain Noise Rejection

Time-frequency analysis of heart rate variability is very sensitive to noise, and as such a need for noise detection arose. The working assumption was that noise induced by for example electrodes detaching and reattaching could be illustrated as a sudden impulse in the signal. An impulse consists of all frequencies and therefore it is expected to yield broadband noise in the time-frequency domain. The ambition was to devise an algorithm for detection of this noise and rejection of segments that contained noise.

In order to develop a noise detection algorithm under controlled circumstances, the

process was divided into several steps. Initially, about 500 ECG recordings were analyzed and checked for noise manually, and separated into two groups. The signals that were deemed to be noise free, i.e., not containing sudden impulses or HRV changes that were deemed to have causes other than biological, were used for construction of the noise detection algorithm. Essentially the threshold for being classified as not noisy was that no sudden large changes in amplitude occurred. The ambition was to add noise to the ”clean” signals at controlled intervals with ground truth available for algorithm development and testing.

Since the noise was generated artificially, measures to avoid creating a method that only detects artificial noise were taken. Firstly, the number of occurrences of noise were varied between each generated signal. This was done by generating a vector of eight equally distributed random values between 0 and 1, and rounding to the nearest integer. If the value for a certain index was one, noise would be added to a corresponding location in the signal. Secondly, the width of the noise was varied. This was done in a similar fashion, where the width of the intervals would be a random integer between 25 and 50 samples.

The noise was generated by drawing values from a Laplace distribution, described in section 2.12, with parameters $\mu = 0$ and b set to twice the standard deviation of the signal that was being noisified. The interval was then replaced with random values sampled from the Laplace distribution. This was done in order to simulate the erratic behavior of the noise that was to be detected. In addition to adding noise to the signal, a ground truth vector was also constructed providing indications of where noise existed, to be used for training and validation.

For the noise generation, 274 signals deemed to be without or with minor amounts of noise were used. For each one of those, 5 noisy signals were created. In total this resulted in a data set containing 1385 noise signals accompanied by ground truth indicators of noise location. This noise generation was repeated for other b values in the Laplace distribution, being 0.25, 0.5, 1 and 1.5 times the standard deviation. Because one would expect there to be near-zero power in frequency bands far above 0.4Hz for purely biological HRV signals, or at least the higher frequencies are uninteresting, it would lend itself natural to look at frequencies over this limit and classifying them as noise.

During exercise it is possible that the respiratory rate goes above this limit for some participants. For non-biological noise sections one would expect the instantaneous power in the frequency band $0.4Hz \leq f < 1Hz$ to increase significantly. With the potential increase in respiratory rate during exercise, it was decided to define the noise band as a dynamic frequency band as opposed to a set lower limit. However, the upper frequency limit was kept at 1Hz.

The reason why the upper limit of the very high frequency band was set to 1Hz and not higher is that there are sampling limitations to the RR-intervals that lay the foundation for this analysis. Because a measurement is only taken from the ECG-recording when

an R peak is detected, the inherent sampling frequency of the signal is $\frac{1}{f_i}$ where $f_i$ is the inverse of the time between two consecutive heart beats. If an individual has a heart rate of 60 beats per minute then the sampling rate would be 1Hz, and in accordance to the Nyquist criterion the maximal frequency that can be detected would be 0.5Hz. Instead of having an estimated sliding sampling frequency, the frequency contents in the before mentioned frequency mentioned were deemed adequate for noise classification. The reason for why this problem arose to begin with was that noise was added directly to the HRV signal instead of to the ECG-recording.

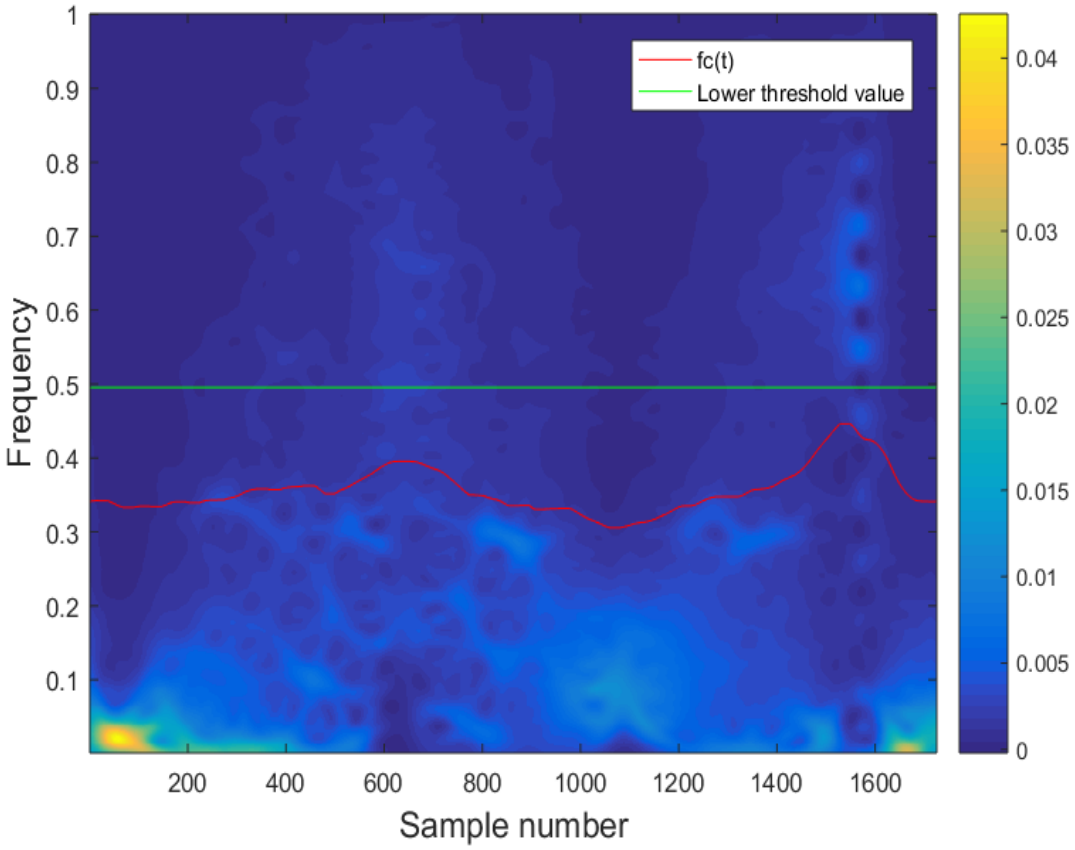



Figure 18: Center of mass and minimum threshold value for the very high frequency band.

Since there was a possibility of respiratory influence on HRV reaching frequencies above 0.4Hz, a movable, varying approach was adopted. High frequency noise was defined as the contents in the frequency bands $0.15 < f \leq 0.50$ as opposed to the upper

limit of 0.4Hz. A second frequency band used for noise band construction was defined as $0.2 < f_{cband} \leq 0.6$. For each sample, the centre of mass in this band is computed,

$$f_{c_i} = \frac{\sum_{t=1}^{n} f_{cband} W(t, f_{cband})}{\sum_{t=1}^{n} W(t, f_{cband})} \tag{54}$$

and a moving median window of size 40 is applied to smooth out the center of mass function. Finally, the lower limit for the very high frequency band is computed as the maximum between 0.45 and the 90:th percentile of $f_{c_i}(t)$ plus 0.1. The motivation for not choosing $\max(f_{c_i}(t))$ is to reduce the impact of sudden high peak values. Choosing a maximum value between 0.45 and the 90:th percentile plus 0.1 ensures that the very high frequency bands remain above the high frequency one. An example of the center of mass and decided lower threshold value can be seen in Figure 18, where the lower threshold for the VHF band is set to slightly below 0.5. Also note how the centre of mass of the HF band changes when sections that can be regarded as noise (high frequency contents in frequencies above 0.5) appear.

After finding the lower bound for the VHF-band, the average instantaneous power is computed. The ratio of the average VHF power and the total TF power was used as a noise indicator for very high frequency noise. To find the optimal thresholding value, an ROC optimization was applied. The data used for training was the one with a noise amplitude of twice the signal standard deviation. In addition to simply thresholding the ratio value and classifying noise sample-by-sample, a dilation technique was trialed.

The idea for applying a dilation function, illustrated in Figure 19, to classified noise indices was that the optimal value would simply detect peaks and classify an area around the peaks. The hypothesis was that this would lead to a reduction in false positives, due to the thresholding limit being increased during optimization.
As can be seen in Figure 20, for intervals where noise is added there is a significant change in instantaneous power for the very high frequency interval. This feature was then used to complete ROC optimization for six classification functions: a simple value cutoff with no dilation and five with dilation varying from one up to ten times the time resolution of the TFD.

After computing the optimal classification function, it was used to classify noise in signals before feature selection. In the instances where noise was said to exist, the LF, HF and TF value(s) for the index(es) were removed.

## 3.3 Reliability Analysis

For each of the extracted features, Spearman's rank correlation and the intraclass correlation coefficient were computed between session one, the first EBT, and session two, the second EBT. In order to measure spurious correlation, all exercise bike tests from

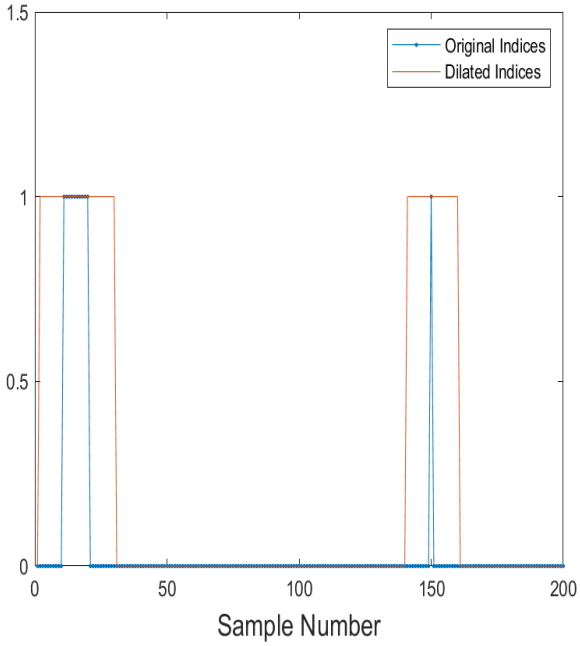



Figure 19: Example of the resulting indices after applying a dilation function of width 20.

session one were compared with randomly selected sessions from session two.

## 3.4 Feature Selection

Because of how some of the features are constructed, one would expect certain interaction between them. For example, the total frequency band is completely dependent on the low and high frequency bands. This could pose a problem when performing parametric regression modelling, as the features in this case do not change independently. This means that the weights assigned by the regression model will not be entirely adequate, as , e.g., total frequency power during rest can not change without either low or high frequency power also increasing. A cross-correlation matrix was constructed for the time-frequency features in order to get a sense of the scope of the problem, see Figure 21.

Another method for the analysis of dependence and interaction of variables is through the use of a random survival forest algorithm, which we now will take a closer look at. Random survival forests are a continuation of the ensemble machine learning algorithm

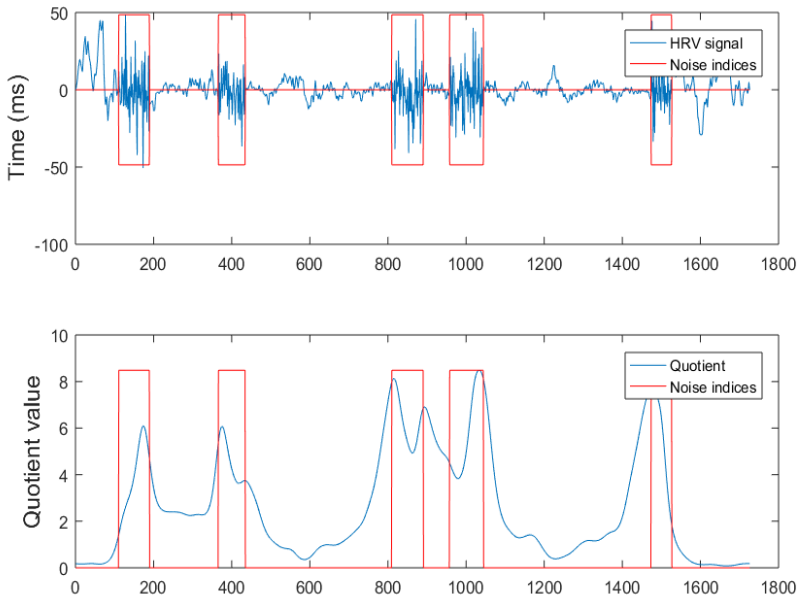



Figure 20: VHF feature behavior during presence of generated noise.

known as random forests. Random forests are used for classification and regression problems, and as such need adjusting to be able to handle (right) censored data, that is data where no event occurs before the end of the study. Before describing the random survival forest algorithm, let us take a step back and look at the single survival tree.

During training, a bootstrap sample of data is used containing on average 63% of the total dataset. Survival trees are binary in nature, i.e., a split will yield two new nodes and is illustrated in Figure 22. When performing a split, the features used are evaluated, and the feature that results in the greatest survival difference between the resulting daughter nodes is used as the split criterion. This process is repeated until the tree reaches a terminal node, where the next split would result in one daughter node with zero deaths/events. There are different measures of change in survival between groups, but the one described here will be the logrank criterion.

The logrank is the null hypothesis test of survival analysis, and is used to determine whether or not there is a statistically significant difference between survival groups. For each incidence of the event, the probability of the event occurring is computed for each group separately. An example given in [54] is two survival groups, consisting of 20 and 31 patients respectively. After six weeks, the first death occurs in group one. The joint probability of death is $\frac{1}{51}$, for group one the expected number of deaths is $20\frac{1}{51}$ and for group two $31\frac{1}{51}$ at week six. When the next death occurs the joint


Figure 21: Colormat showing the relationship between extracted features.

probability would be $\frac{1}{50}$, expected number of deaths $19\frac{1}{50}$ for group one, as one group member died earlier, and $31\frac{1}{50}$ for group two.

Patients who survive until the end of the study (right censored) are considered to have a risk of death during the final week, and none thereafter. For each group the total number of expected deaths is summarized, and compared with the actual observed number of deaths in each group. The difference between the expected and observed values lays the foundation for the logrank test score:

$$LR = \frac{(O-E)^2}{E}, \tag{55}$$

where O and E denote the observed and expected number of deaths in the group. The LR-value for both groups is added together, and the sum is checked against the $\chi^2$ distribution to obtain a P-value. The aim of the logrank split criterion is to maximize the value of this test statistic.

After the survival tree is fully formed, the cumulative hazard function (CHF) for each terminal node, h, is computed,

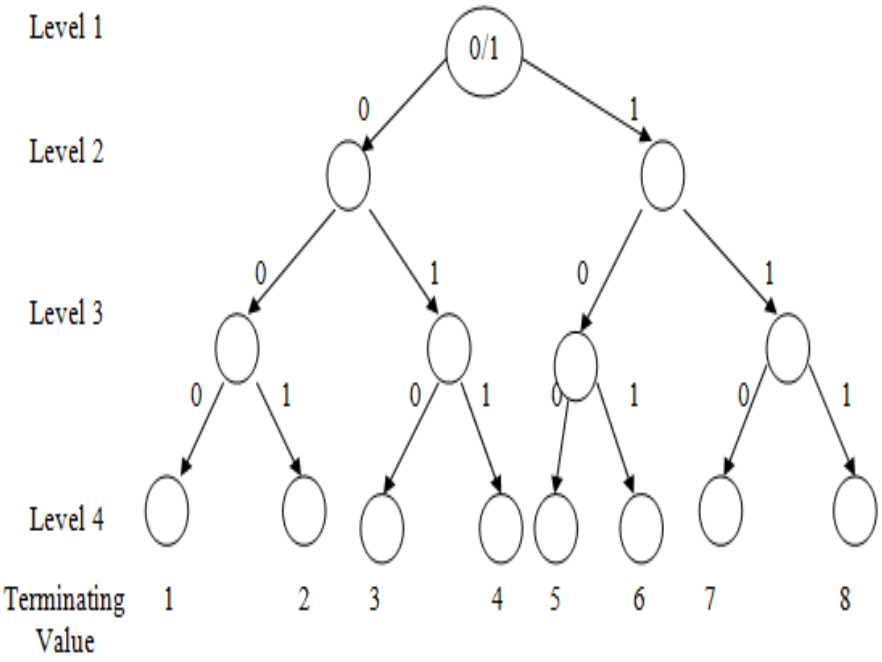



Figure 22: Example of a binary decision tree

$$\hat{H}_h(t) = \sum_{t_{l,h} \leq t} \frac{d_{l,h}}{Y_{l,h}}, \tag{56}$$

where d is the number of deaths at the particular time t and Y is the number of susceptible individuals in the node. After having computed the CHF for all terminal nodes, we can use the survival tree to estimate the cumulative hazard for an individual, and thus the chance of survival. So for one individual we have the set of features $\mathbf{x}_i$, with resulting cumulative hazard function

$$H(t \mid \mathbf{x}_i) = \hat{H}_h(t), \text{ if } \mathbf{x}_i \in \text{h}. \tag{57}$$

In order to evaluate the predictive capabilities of the survial tree, the remaining 37% of the available data that was not used for training, also known as out of bag (OOB) data, is used. By repeating this growing of survival trees, one can use the ensemble of trees or *forest* to create an ensemble average CHF value for each set of features over B trees:

$$H_e^{**}(t \mid \mathbf{x}_i) = \frac{\sum_{b=1}^{B} I_{i,b} H_b^*(t \mid \mathbf{x}_i)}{\sum_{b=1}^{B} I_{i,b}}, \tag{58}$$

where $I_{i,b}$ is 1 if sample $i$ is in the OOB dataset, and 0 otherwise. Using the ensemble average CHF value, we can construct the ensemble mortality for an OOB sample:

$$\hat{M}^{**}_{e,i} = \sum_{j=1}^{n} H^{**}_{e}(T_j \mid x_i) \tag{59}$$

, i.e., the mortality is the summation of the ensemble average CHF over time $T_j$.

Now that we have a way of estimating the mortality of an individual coupled with ground truth data, it is possible to get an estimate for the prediction error of the survival tree model(s). This is done by calculating Harrell's concordance index (C-index), which is computed in four steps:

1. Pair all the individuals in the dataset.
2. Omit pairs where the individual with the shorter survival time is censored. Also where the survival time is the same, unless at least one of the survival times is time to death.
3. For each pair where the survival times are not identical, count 1 if the predicted mortality is higher for the one with a shorter survial time. If the predicted outcomes are tied, count 0.5. Where times are identical and both are deaths, count 1 if the predicted outcomes are the same for both, and 0.5 otherwise. If the times are identical but only one is a death, count 1 if the death has worse predicted outcome, and 0.5 otherwise.
4. The sum of all the counts divided by the number of pairs is the concordance index.

Using the C-index, we can get a sense of the importance of each feature used. For a certain feature x, we first compute the C-index with the dataset intact. Then, to get a measure of x's predictive power, all x values are randomly assigned to individuals. Randomly assigning feature values means that the predictive power of the feature is, on average, nullified as its value now essentially is noise. By computing the C-index again we get a sense of the decrease in predictive power. If the C-index *increases*, we can draw the conclusion that noise is a better predictor than our feature, [55].

Before using the random survival forest algorithm for evaluation of features, some pre-processing had to be done. Individuals who had a cardiac event before the study began were excluded from further analysis. In addition to this, the features were dichotomized so that the top 25% values were classified as 1, and the bottom 75% as 0. The reasoning for this was that the disparity between the median and the maximum value for individuals was very great, and would have an effect on especially the standard deviation of the feature. Roughly 10-15% of each feature were significantly larger than the median feature value, and so 25% was used as a catch-all threshold.

After this, a matching algorithm (matchit, [56]) was used to find samples with similar covariate distributions. The feature selected as "treatment" was the binarized feature representing median power in the high frequency band during increasing workload, and the covariates being matched were age, sex, diabetes status and systolic blood pressure. In essence the reason for this was to select individuals with similar demographical characteristics in order to reduce bias.

After removing unwanted individuals with prior cardiac events and selecting a matched subset of individuals from the dataset, 29.846 measures remained. These were used to train a random survival forest, in order to discriminate between important and unimportant features. Due to hardware constraints, subsampling was employed during training. The forest consisted of 1000 trees.

## 3.5 Survival Analysis

After eliminating variables of low importance, a multivariate Cox regression was applied to the remainders. This process was initially performed on the top four time-frequency features, with iterative removal of non-significant features to follow. The Cox proportional hazards model is the most commonly used model for survival analysis in medical research and was first suggested by D. R. Cox in 1972, [57]. The hazard function h(t) is defined as

$$h(t) = h_0(t)e^{(b_1x_1+b_2x_2+...+b_px_p)}, \tag{60}$$

Where $h_0(t)$ is referred to as the baseline hazard, i.e. the hazard at time t if all the predictors $x_1...x_p$ were equal to zero, and $b_1...b_p$ are their corresponding coefficients. By rewriting (60) as

$$ln(h(t)) - ln(h_0(t)) = b_1x_1 + b_2x_2 + ... + b_px_p, \tag{61}$$

we can view the predictors and coefficients as the result of a multiple linear regression describing the behavior of the logarithm of the ratio of $h(t)$ and $h_0(t)$ with respect to the predictors. The variable $h_0$ can be viewed as the missing $b_0$ term that we otherwise would find in a linear regression. Using the relationship described in (60), we can via integration from 0 to t get the cumulative hazard function:

$$H(t) = H_0e^{(b_1x_1+b_2x_2+...+b_px_p)} \tag{62}$$

From the cumulative hazard function, we can via exponentiation and changing signs obtain the survivor function:

$$S(t) = e^{(H_0)e^{(b_1x_1+b_2x_2+\ldots+b_pxp)}} = S_0(t)^{e^{(b_1x_1+b_2x_2+\ldots+b_pxp)}}[49] \tag{63}$$

# 4 Results

## 4.1 Pre-processing Block

After pre-processing the dataset, we are left with a series of R-R interval times over time. An example of this transformation can be seen in Figure 23. We can see in this instance that the interval series contains several interval times that differ significantly from their neighboring values. These are removed during the noise reduction section of this thesis workflow.

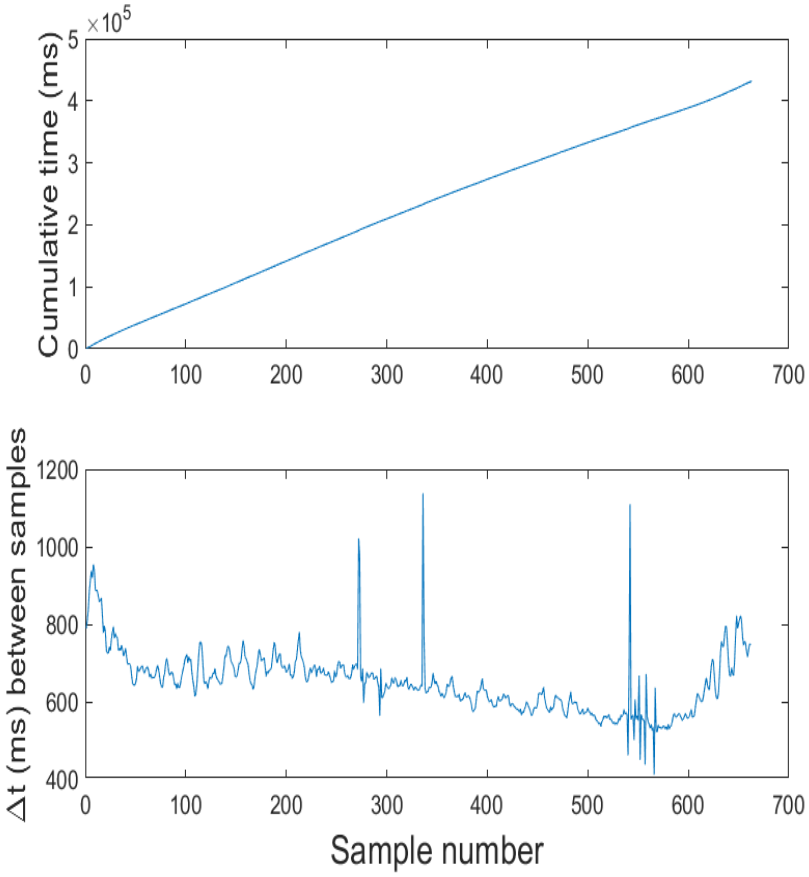



Figure 23: Initial RR spike data, shown in the upper plot, together with the differentiated RR-interval series, shown below.

### 4.1.1 Time Domain Noise Reduction

As can be seen in Figure 24, post-processing RR interval series seemingly contain no trace of the original noise. The original signal, also shown in Figure 16, contained three ectopic beats and one ”reverse ectopic” beat. The resulting spline interpolated signal instead has smoothed out sections where the original spikes were located.

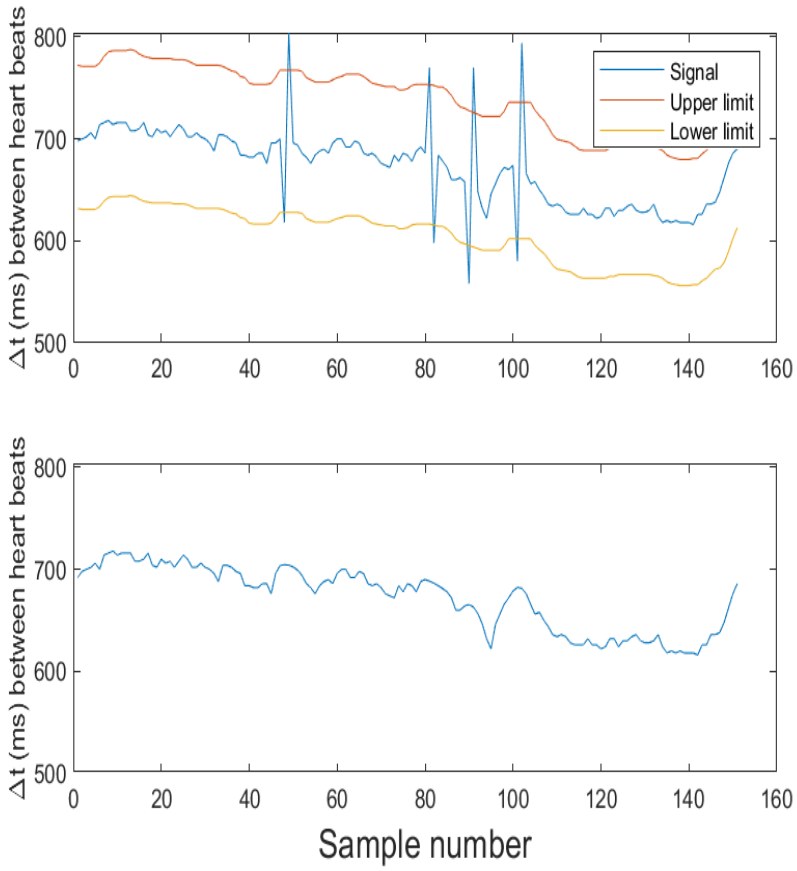



Figure 24: Results of time-domain noise removal. The upper plot consists of a signal with four abnormal beats, along with the upper and lower limits for noise classification. The bottom plot contains the corresponding interpolated signal post-noise removal.

No editing is applied to the sections that are not close to the anomalies. This result also appears similar for cases where a missed detection leads to one high peak of roughly twice the time value of its surrounding R-R intervals. In the case of ectopic or ”reverse ectopic” beats, the total amount of heart beats in the series remains the same, while the

case of a missed heart beat leads to an additional heart beat being interpolated into the series.

## 4.2 Time-Frequency Analysis

### 4.2.1 Feature Extraction

The behavior of the instantaneous power over the duration of the EBT can be seen in Figure 25. As can be observed, the relation between the low and high frequency components vary over the duration of the EBT.

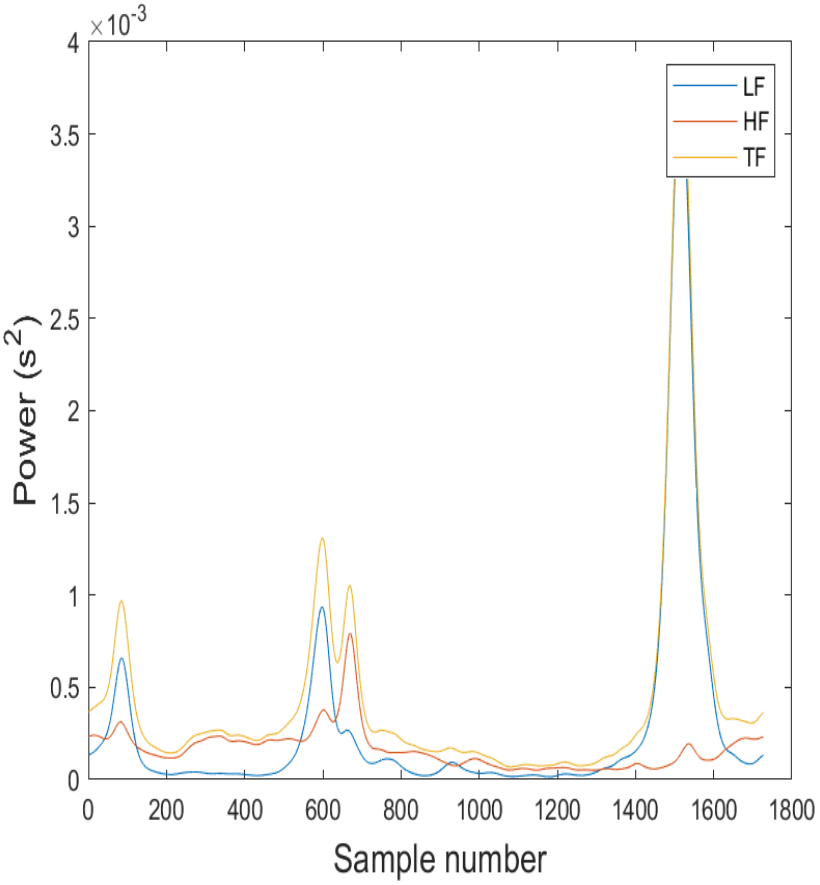



Figure 25: Instantaneous power for the three frequency bands

### 4.2.2 Time-Frequency Domain Noise Rejection

After performing the ROC-curve analysis, see Figure 26, it was found that the dilation of noise classification indices did not outperform classifying individual samples as noise, one at a time. This result was also verified by running the noise classification algorithm on the test dataset, where not using dilation performed the best. The results from running the algorithm on the test dataset can be seen in table 1.

Not using any image dilation results in a higher true $\frac{truepositive}{falsepositive}$ quotient consistently, and reaches a 100% true positive rate before any other alternative. At a false positive rate of roughly 0.25 and higher, using no image dilation performs at the same level as using an image dilation of 1x the time time frequency distribution time resolution.

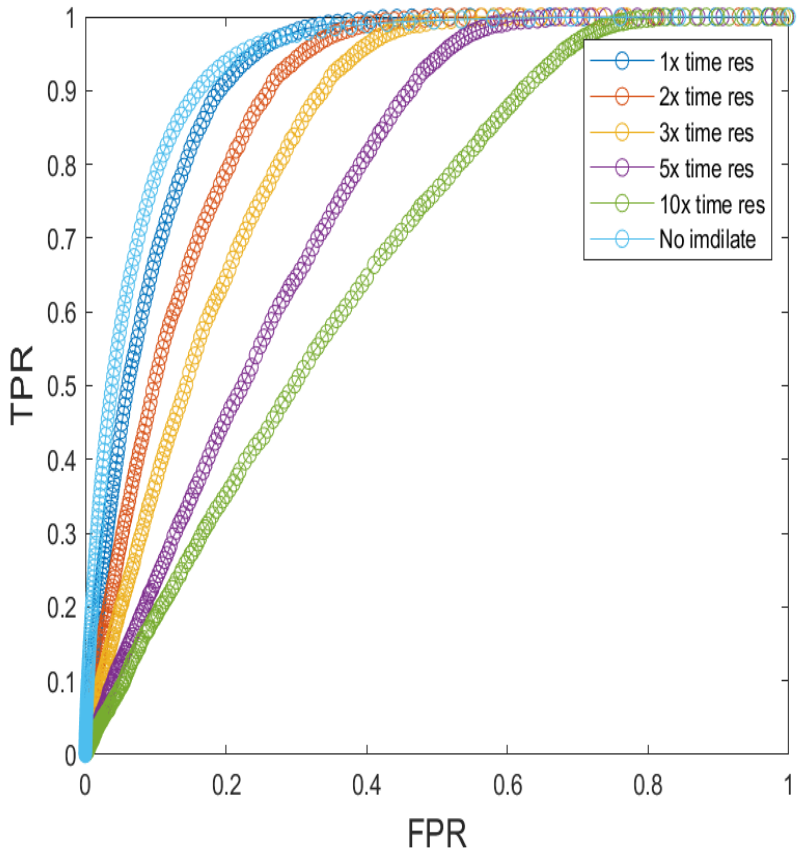



Figure 26: ROC-curves for different time dilations

Table 1: Performance of optimal thresholding limits for different dilation values on the test data set

| Dilation Resolution | Sensisivity | Specificity |
|---|---|---|
| No Imdilate | 0.89 | 0.85 |
| 1x Imdilate | 0.88 | 0.82 |
| 2x Imdilate | 0.86 | 0.77 |
| 3x Imdilate | 0.84 | 0.71 |
| 5x Imdilate | 0.77 | 0.63 |
| 10x Imdilate | 0.65 | 0.60 |

Using 2x the time resolution consistently performs at levels in-between 1x and 3x the time resolution. At a false positive rate of 0.4 and onward, the 2x time resolution classifier intersects with the 1x time dilation curve and maintains similar performance for higher rates of false positivity. In a similar fashion, using 3x time resolution performs at levels in between 2x and 5x, 5x performs at levels in between 3x and 10x, and 10x performs worst of all the classifiers. In short, the wider the dilation applied to the noise classification, the worse the algorithm performs.

In Figure 27 we can see the performance of the VHF noise rejection algorithm on a signal. The signal, upper plot, contains a section with high frequency noise (mainly several ectopic beats), which affects the high frequency band seen in the middle plot. In the bottom plot, we can see the green line indicating where values have been removed due to being deemed too noisy. Because no dilation function has been used, the part between the two high frequency segments remains untouched. Also observe that the power in the high frequency bands increases together with the increase in the very high frequency.

After running the time-frequency domain noise rejection algorithm on the data used for reliability analysis, it was found that 15.4% of the samples contained very high frequency noise. Out of these sessions, the average percentage of noise classified was 21.3%. This value is reached after performing time-domain noise rejection prior to time-frequency noise analysis. This because it was preferred to have data edited in the time-domain filtering part rather than having it discarded.

## 4.3 Reliability Analysis

Reliability analysis was performed for both median and mean features, and the resulting correlation values can be seen in tables 2 and 3. Measures of both Spearman, $\rho$, correlation and intra class correlation, ICC, both before and after noise reduction seem to indicate no significant change or a reduction in repeatability for both sets of features. We also see a constant difference between Spearman's $\rho$ and the intraclass correlation coefficient, which remains after the noise removal.

Table 2: Reliability measures for log(features) using median and mean power, before noise removal

| **Feature** | $\rho_{Median}$ | $\rho_{Mean}$ | $ICC_{Median}$ | $ICC_{Mean}$ |
|---|---|---|---|---|
| **Low Frequency** | | | | |
| Rest | 0.384 | 0.394 | 0.560 | 0.566 |
| Const. Workload | 0.567 | 0.517 | 0.686 | 0.635 |
| Inc. Workload | 0.659 | 0.599 | 0.757 | 0.690 |
| Recovery | 0.495 | 0.497 | 0.649 | 0.647 |
| Total | 0.669 | 0.552 | 0.777 | 0.659 |
| **High Frequency** | | | | |
| Rest | 0.468 | 0.465 | 0.622 | 0.618 |
| Const. Workload | 0.560 | 0.493 | 0.681 | 0.625 |
| Inc. Workload | 0.618 | 0.512 | 0.717 | 0.658 |
| Recovery | 0.527 | 0.512 | 0.693 | 0.679 |
| Total | 0.629 | 0.506 | 0.747 | 0.674 |
| **Total Frequency** | | | | |
| Rest | 0.427 | 0.432 | 0.597 | 0.599 |
| Const. Workload | 0.555 | 0.495 | 0.674 | 0.627 |
| Inc. Workload | 0.633 | 0.556 | 0.731 | 0.673 |
| Recovery | 0.533 | 0.520 | 0.684 | 0.672 |
| Total | 0.649 | 0.536 | 0.758 | 0.669 |

*Note: all coefficients are significant with p-values <0.01.* $\rho$ = correlation coefficient, $ICC$ = consistency coefficient

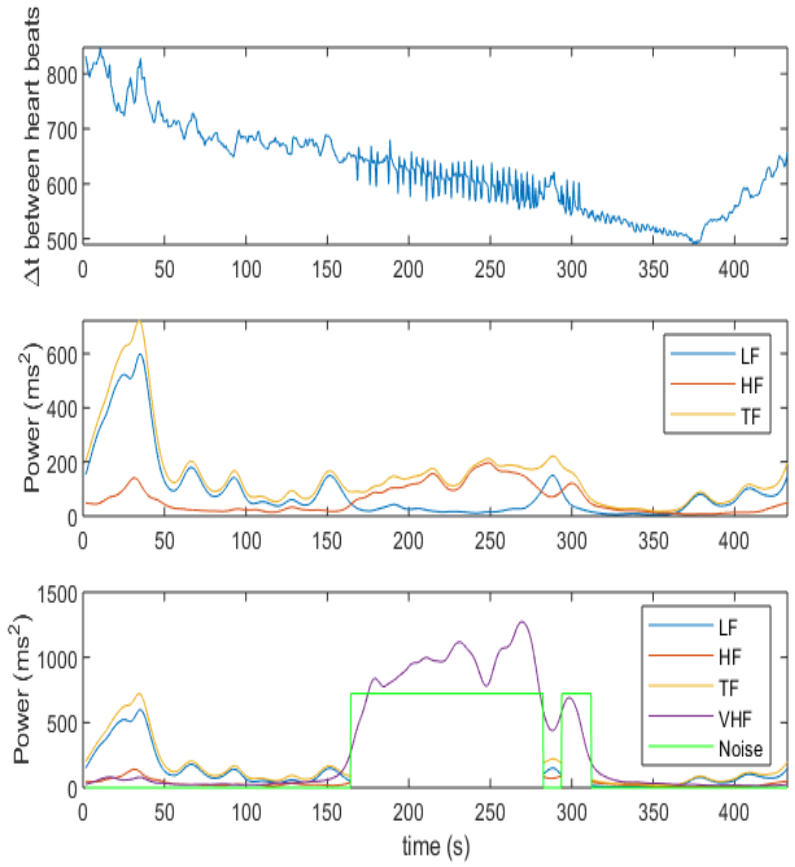



Figure 27: Results of time frequency-domain noise removal. The top picture shows the original signal. Original power in low, high and total frequency bands is illustrated in the middle plot, while the effects of VHF rejection are illustrated in the bottom plot.

Low frequency feature correlation is higher for median value across the entire exercise bike test, however the relationship between the low and high frequency correlation values varies for the different segments. The total frequency feature correlation often finds itself in between the values of the low and high frequency features. Again, this relationship remains after applying the time and time-frequency domain noise rejection algorithms. There were no indications of spurious correlation, as all correlation coefficients were $<0.1$ and statistically insignificant with p-values $>0.05$.

## 4.4 Feature Selection

The results from fitting a random survival forest to the 15 median value features, in combination with demographic features regarding bmi, smoking status, diabetes, alco-

Table 3: Reliability measures for log(features) using median and mean power, **after** noise removal

| **Feature** | $\rho_{Median}$ | $\rho_{Mean}$ | $ICC_{Median}$ | $ICC_{Mean}$ |
|---|---|---|---|---|
| **Low Frequency** | | | | |
| Rest | 0.389 | 0.399 | 0.563 | 0.569 |
| Const. Workload | 0.525 | 0.498 | 0.650 | 0.609 |
| Inc. Workload | 0.623 | 0.586 | 0.722 | 0.676 |
| Recovery | 0.504 | 0.499 | 0.655 | 0.651 |
| Total | 0.623 | 0.529 | 0.733 | 0.641 |
| **High Frequency** | | | | |
| Rest | 0.470 | 0.466 | 0.626 | 0.620 |
| Const. Workload | 0.557 | 0.506 | 0.663 | 0.625 |
| Inc. Workload | 0.631 | 0.566 | 0.712 | 0.675 |
| Recovery | 0.537 | 0.532 | 0.702 | 0.693 |
| Total | 0.627 | 0.529 | 0.732 | 0.684 |
| **Total Frequency** | | | | |
| Rest | 0.430 | 0.434 | 0.601 | 0.602 |
| Const. Workload | 0.536 | 0.487 | 0.648 | 0.614 |
| Inc. Workload | 0.627 | 0.578 | 0.714 | 0.678 |
| Recovery | 0.540 | 0.527 | 0.692 | 0.680 |
| Total | 0.631 | 0.535 | 0.729 | 0.667 |

*Note: all coefficients are significant with p-values <0.01.* $\rho$ = correlation coefficient, $ICC$ = consistency coefficient

hol consumption frequency, systolic blood pressure and sex, can be seen in Figure 28.

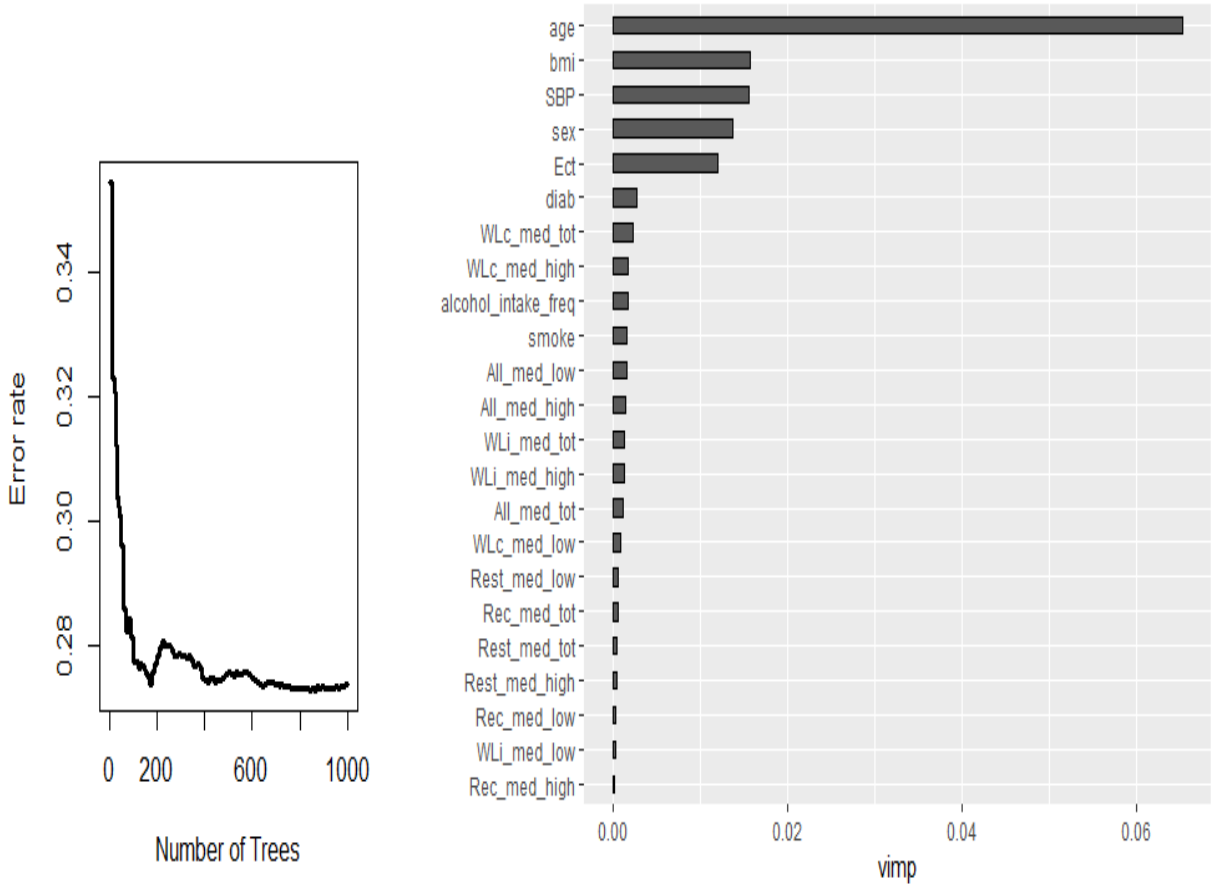



Figure 28: Prediction error rate and variable importance (vimp) for the trained random survival forest.

On the left side we can see how the error rate reduces as the number of trees increases, while the right side of Figure 28 contains the variable importance for each variable included. Variable importance is a measure of how the random forest evaluates its split criterions, where a higher variable importance means a greater difference between groups post-split. The top four time-frequency features are the median power in the high and total frequency bands during constant workload, and median power in the low and high frequency bands across the entire exercise bike test. We can see a trend towards the low frequency features scoring the lowest variable importance.

The error rate bottoms out already after less than two hundred trees, but then goes back up again and later converges after roughly 800 trees. Age, BMI, systolic blood

pressure, sex, ectopic beat categorization all rank as having a higher variable importance than any of the HRV indices. Alcohol intake frequency and smoking status rank below the two top rated HRV indices for variable importance.

## 4.5 Survival Analysis

The hazard ratios for the final Cox model with all demographic and the most significant time-frequency-predictor can be seen in table 4. As we can see, the most significant TF-predictor, diabetic status and alcohol intake frequency are all statistically insignificant.

In order to investigate whether or not to keep the signals where major editing had been performed, another analysis was performed on data where any signal that contained 10% or more edited samples was discarded. The results from this analysis can be viewed in table 5. The major change in predictor significance is that when more aggressive exclusion criteria are applied, all the three variables that in table 4 were insignificant now are significant. The highest hazard ratio by far belongs to the ”sex” category, where being male increases hazard with respect to cardiovascular events by a factor of 2.37 or by 137% in comparison to being female. Do note that the age, SBP, BMI and alcohol intake frequency predictors are continuous, while the others are categorical.

Being in the high ectopic beat count group indicates a 52% increased risk of experiencing a cardiac event within the following six years of taking an exercise bike test. Having diabetes means that an individual runs a roughly 36% increase in cardiovascular event risk. An individual who is 75 years old, has a risk increase by 8.9% compared to if they are 74. This is a multiplicative compounding factor, so aging ten years increases risk by 134% for example.

BMI also is a compounding continuous variable, where each point increase in ones BMI leads to a 6.7% increase in risk for cardiovascular events. By comparison, being in the top 25% group with respect to the median power in the high frequency band over the entire duration of the exercise bike test leads to a 20% increase in risk.

Table 4: Hazard ratios with 95% confidence interval values for the cox regression model when keeping abnormal signals

| **Variable** | **95% CI Lower** | **Hazard Ratio** | **95% CI Upper** | **p-value** |
|---|---|---|---|---|
| All_med_high | 0.9762358 | 1.127366 | 1.301892 | 0.103 |
| Age | 1.0764128 | 1.087109 | 1.097912 | <2e-16 |
| SBP | 1.0051501 | 1.009253 | 1.013373 | 9.4e-06 |
| Sex | 1.7482597 | 2.020955 | 2.336185 | <2e-16 |
| BMI | 1.0494601 | 1.066113 | 1.083029 | 1.6e-15 |
| Smoking Status | 1.1542736 | 1.427062 | 1.764319 | 0.001 |
| Alcohol Intake Frequency | 0.9993597 | 1.044857 | 1.092426 | 0.053 |
| Diabetes | 0.8018364 | 1.069178 | 1.425653 | 0.649 |
| Ectopic Beats | 1.5591148 | 1.889079 | 2.288876 | 8.3e-11 |

Table 5: Hazard ratios with 95% confidence interval after removing excessively edited signals

| **Variable** | **95% CI Lower** | **Hazard Ratio** | **95% CI Upper** | **p-value** |
|---|---|---|---|---|
| All_med_high | 1.042239 | 1.203862 | 1.390548 | 0.0117 |
| Age | 1.066685 | 1.077542 | 1.088510 | <2e-16 |
| SBP | 1.007150 | 1.011591 | 1.016051 | 2.8e-07 |
| Sex | 1.732468 | 2.024948 | 2.366804 | <2e-16 |
| BMI | 1.048258 | 1.066633 | 1.085331 | 3.5e-13 |
| Smoking Status | 1.102979 | 1.393681 | 1.761001 | 0.0054 |
| Alcohol Intake Frequency | 1.025498 | 1.075873 | 1.128723 | 0.0028 |
| Diabetes | 1.006134 | 1.356261 | 1.828229 | 0.0455 |
| Ectopic Beats | 1.082319 | 1.521640 | 2.139285 | 0.0157 |

# 5 Discussion

## 5.1 Pre-processing Block

### 5.1.1 Time Domain Noise Reduction

As previously mentioned, the removal of noise may result in the addition of new forms of noise. Because an interpolation operation is performed on the data, the occurrence of e.g. an ectopic beat will have an effect not only on the two samples that are the premature beat and then the following delayed beat, but also the surrounding beats. The same holds true for other abnormalities found in the signals. It would be beneficial in future studies to consider rejection of signals where an excessive amount of interpolation has been performed, to prevent training models on too artificial data. In this thesis a limit of ten percent was chosen, but a finer limit may be better.

A manual review of about 500 signals was performed during the development of the time domain noise reduction algorithms, only satisfactory results where deviating intervals were correctly interpolated were produced. However, during the creation of plots for this section I found an anomaly in the results after noise reduction where some RR interval values were $> 5$ seconds. It seemingly is a rare occurrence as it did not show up until this final stage, and so the assumption is that it did not significantly effect the analysis at a general level. However, in future applications this should be looked at and fixed to improve performance at an individual level. It is possible that the recalculation of RR-interval times becomes skewed when the abnormality lies at the beginning of the analyzed signal. Perhaps the use of padding the beginning of signals or running some complimentary analysis of results post-processing would help reduce the impact of this issue.

## 5.2 Time-Frequency Analysis

### 5.2.1 Feature Extraction

There may be a value to revisiting how the time intervals are defined when extracting features, as the length of intervals seem to affect correlation values, further discussed in section 5.3. An alternative to redefining intervals may be to extend the intervals when performing new measurements in the future.

### 5.2.2 Time-Frequency Domain Noise Rejection

Some of the signals initially deemed as ”clean” have later been discovered to contain noise elements. These elements have potentially lead to a corruption of the ROC optimization. Perhaps someone with a finer eye for what is noisy and what is not should have done the estimation, instead of an engineering student. An alternative to generation of data would be to carefully manually annotate real signals to decide where noise indices of interest are, but this was deemed too time consuming and uncertain.

The approach taken to generating noise and its application may need to be reviewed. The use of two times the standard deviation as amplitude when generating the Laplacian noise is perhaps too generous for detection, or maybe too conservative. As the noise added was based on empirical evaluation rather than for example generation through use of manually annotated noisy sections, it is possible that the amplitude used did not properly represent the noisy sections contained in the real signals.

The optimized cutoff threshold selected in section 4.2.2 still only achieves a sensitivity of 0.89 and specificity of 0.85. This means that 11% of all very high noisy frequency samples are kept, while 15% of all non-noise segments are discarded due to being classified as noise. Since it is assumed that a minority of recordings will contain very high frequency noise post-processing with time domain based noise reduction algorithms, a 15% rejection rate for non-noisy segments means a high loss in information that could otherwise be used for analysis.

When applying the very high frequency noise rejection algorithm, some biologically based but noisy looking segments, such as atrial fibrillation with its irregular and heightened heart rate, will likely be rejected. For example different kinds of tachycardia such as atrial fibrillation will be classified as very high frequency noise. For other studies of heart beat behavior this will be a significant loss of information, but in the study of heart rate variability it is preferable that tachycardia sections are removed due to the distortion they cause to the heart rate variability measures, as shown previously in for example section 4.2.2.

The fact that 15.4% of signals were classified as having very high frequency noise contents, and each of them having on average 21.3% of their signal contents removed means that some more thorough analysis of noise rejection may need to be performed prior to deciding whether or not to include the signal for survival analysis. As it stands now, there is still a possibility that individuals with significantly reduced signal information are included in the following survival analysis, where as it may be better to discard them instead. If no time-domain noise interpolation had been performed, it is not impossible to assume that some segments would be completely removed after very high noise frequency analysis and rejection.

## 5.3 Reliability analysis

As can be seen in table 2, the correlation between test sessions is consistently lower for the mean power feature. This was also expected, as a short burst of noise can have a great impact upon the mean value without necessary affecting the median value. We can see from these tables that consistency is lower for almost all mean features compared to their median counterparts. As noise most likely affects mean power values more adversely than median power values, this was expected and we would expect to see the greatest improvement in repeatability values for the mean feature after noise removal.

The lower correlation values for certain segments can perhaps be explained by means other than noise. Because the resting period before exercise begins is only 15 seconds, it would not take a lot of noise or simply variation in daily form to greatly impact the similarity between the two sessions. This pattern actually repeats itself in both median and mean power features: the constant workload, being two minutes, consistently scores a higher correlation value than the resting section. In turn, the increasing workload section, being even longer than the constant workload time of two minutes and the longest out of the four intervals, scores the highest. The recovery section also follows this ”rule” for the median features but not for the mean ones. We can also see that the correlation of the total exercise bike test results consistently is the greatest for median features, but not for mean where the highest value varies in interval location. Perhaps it would be better, for the sake of this analysis at least, that the exercise bike tests contain equally long segments of rest, constant workload, increasing workload and recovery.

Interestingly the changes in correlation between test sessions are negative or almost negligable after applying the noise removal algorithms, as can be seen in table 3. The repeatability values do not improve for the mean features either as expected. Perhaps the noise itself was well correlated between exercise bike tests, or the noise reduction algorithms instead introduced new forms of noise. The pattern of longer segments resulting in higher correlation values still holds true after noise reduction.

## 5.4 Feature Selection

While the dataset that was being worked on originally contained 100.000 individuals, after exclusion due to different criteria and the matching algorithm the final dataset consisted of 29.870 exercise bike test instances. Out of these, only 890 had a cardiovascular event during the follow up period after taking the test. This is because the study population was a general one, and in the general population the incidence of events is low.

The random survival forest algorithm proved to be very computationally expensive on the Lenovo Ideapad 310 that was used, which is why the need for subsampling during training arose. Finally, the variable importance values given by the fully trained model did not seem to correspond to the level of significance that the variables held in a Cox proportional hazards model. Random forests focus on creating a model that maximizes predictive power, with no consideration given to for example p-values. This means that in order to maximize model performance, all variables should be kept. However, if one wishes to look at causes for certain occurences, in this case cardiac events, feature deletion is necessary. This is why a combination of random survival forest with traditional Cox regression yields the most information about links between features, prediction and causation.

## 5.5 Survival Analysis

Diabetes status originally proved to be an insignificant variable when predicting cardiovascular events with a Cox proportional hazards regression model, while keeping signals where abnormalities made up $<$50% of the signal. This contrasts conventional knowledge regarding the link between diabetes and cardiovascular disease, [58–61]. It is not likely that this lack of significance is a break-through in research regarding the link between diabetes and cardiovascular disease, as the results in table 5 indicate that somehow the process of editing signals distorts the relationship between diabetes and cardiovascular events.

When removing abnormal signals with more aggressive criteria, diabetic status proves to be significant again. However, one needs to keep in mind that changing the data to fit the theory is not something to be done for all circumstances. In this case, keeping a signal where 50% of its content is artificially generated through noise editing does not seem to give a fair representation of the underlying process being investigated.

It is also possible that the connection between the significant time-frequency predictor and cardiovascular events is caused by the noise editing algorithm employed during time domain noise reduction. In the final analysis, any signal with $>$10% edited samples is rejected, however further analysis in the localization of these edits in the future will help shed further light on the results of editing. Perhaps certain sections are edited at larger rates than others, and the rejection criteria may need to be adjusted to reject on a section-by-section basis instead of only looking at the overall editing rate. However, a chi-squared test reveals that variables Ect, signifying whether or not an individual has a high amount of ectopic beats, and All_med_high, representing the median power in the high frequency band over the entire EBT, are independent from each other.

Since this study was conducted on a general population sample, its results could, after further investigation, potentially be used as a screening method. For example exercise bikes in gyms often have electrodes that measure your pulse, so it would be possible to implement the exercise bike test done for the UK Biobank on a commercial exercise bike as well.

# 6 Conclusion

Using time-frequency characteristics as a means of noise reduction in RR-interval series proved to be a useful tool in detecting sections of very high frequency noise. However, the removal of very high frequency noise in combination with time domain noise reduction not only did not increase but rather decreased the reliability of exercise bike test recordings. While a random survival forest approach to variable selection helped reduce the number of model fitting iterations, the variable importance measure does not translate to significance level. Median power in the high frequency band across the entire exercise bike test proved to be the only significant time-frequency feature extracted during this thesis.

It may be a bit discouraging that only one out of fifteen features are significant in a multivariate Cox regression, however as part of the thesis problem formulation was extraction and evaluation of said features this result is still positive. In addition to the time-frequency feature, using information about an individual's proportion of ectopic beats also provides significant predictive power for a survival analysis. While obvious conclusions with respect to HRV changes during exercise and their link to cardiac events have not necessarily been discovered through this thesis, new and interesting questions have arisen that need further investigation.

The project resulted, as desired, in a pipeline from raw data to survival analysis.

# References


[1] Time-frequency analysis for the dynamic quantification of the interactions between signals related to the cardiovascular system. PhD Dissertation by Michele Orini.

[2] Yasuma et al. Respiratory sinus arrhythmia: Why does the heartbeat synchronize with respiratory rhythm? *Chest*, 125(2):683–690, 2004. URL https://doi.org/10.1378/chest.125.2.683.

[3] Kleiger et al. Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. *The American Journal of Cardiology*, 59(4):256–262, 1987. URL https://doi.org/10.1016/0002-9149(87)90795-8.

[4] Shah et al. Posttraumatic stress disorder and impaired autonomic modulation in male twins. *Biological Psychiatry*, 73(11):1103–1110, 2013. URL https://doi.org/10.1016/j.biopsych.2013.01.019.

[5] Wikipedia. -https://upload.wikimedia.org/wikipedia/commons/c/cb/Action_potential_ventr_myocyte.gif, . Accessed: 2018-10-04.

[6] Bartholomew Martini. *Essentials of Anatomy  Physiology Sixth Edition*. Pearson, 2013. ISBN 978-0-321-79862-6.

[7] Cardiac conduction system. -http://mdnxs.com/wpcontent/uploads/2015/11/CARDIACCONDUCTION.png, . Accessed: 2018-10-04.

[8] Ecg normal waveforms. -http://www.insetfair.com/admin/images/water_91908712_Normal\%20ecg\%20recording\%20for\%20one\%20cardiac\%20cycle_m.jpg. Accessed: 2019-03-15.

[9] Fisher et al. Central sympathetic overactivity: Maladies and mechanisms. *Autonomic Neuroscience*, 148(1-2):5–15, 2009. URL https://doi.org/10.1016/j.autneu.2009.02.003.

[10] The rr interval. -https://9097bca1a75e8bb2e831365d-hrvfitltd.netdna-ssl.com/wp-content/uploads/2017/02/R-R-interval-trace.png. Accessed: 2018-10-04.

[11] Lampert et al. Decreased heart rate variability is associated with higher levels of inflammation in middle-aged men. *American Heart Journal*, 156(4):759.e1–759.e7, 2008. URL https://doi.org/10.1016/j.ahj.2008.07.009.

[12] Rahman et al. Low frequency power of heart rate variability reflects baroreflex function, not cardiac sympathetic innervation. *Clinical Autonomic Research*, 21(3):133–141, 2011. URL https://doi.org/10.1007/s10286-010-0098-y.

[13] Kubota et al. Heart rate variability and lifetime risk of cardiovascular disease: the atherosclerosis risk in communities study. *Annals of Epidemiology*, 27(10): 619–625, 2017. URL https://doi-org.ludwig.lub.lu.se/10.1016/j.annepidem.2017.08.024.

[14] Centers for disease control and prevention: Tobacco-related mortality. https://www.cdc.gov/tobacco/data_statistics/fact_sheets/health_effects/tobacco_related_mortality/index.htm. Accessed: 2019-01-30.

[15] Sörnmo Pahlm. *Elektrokardiologi, Klinik och teknik*. Narayana Press, 2006. ISBN 978-91-44-00615-4.

[16] Busby et al. Prevalence and long-term significance of exercise-induced frequent or repetitive ventricular ectopic beats in apparently healthy volunteers. *Journal of the American College of Cardiology*, 14(7):1659–1665, 1989. URL http://www.onlinejacc.org/content/14/7/1659.

[17] Le Nguyen et al. Supraventricular ectopic activity: When excessive it is not all benign! *Journal of Atrial Fibrillation*, 3(2):307, 2010. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4956355/.

[18] Pvc example. -http://medlibes.com/uploads/Screen%20shot%202010-07-06%20at%209.47.50%20PM.png, . Accessed: 2018-10-04.

[19] Premature ventricular conctractions. -https://ecgwaves.com/premature-ventricular-contractions-complex-beats-ecg/, . Accessed: 2018-10-04.

[20] Premature atrial conctraction. -https://ecgwaves.com/premature-atrial-contraction-beat-complex/, . Accessed: 2018-10-04.

[21] Premature atrial conctraction image. -https://afibemory.files.wordpress.com/2017/09/premature-atrial-contraction-pac1.jpg, . Accessed: 2018-10-04.

[22] P. Singh. Time-frequency analysis via the fourier representation. 2016. URL https://arxiv.org/pdf/1604.04992.pdf.

[23] Fourier transform of 1hz. -http://www.johnloomis.org/eop513/notes/fourier_transform/html/test_fourier_04.png, . Accessed: 2018-09-30.

[24] The fourier transform webpage. -http://www.thefouriertransform.com, . Accessed: 2018-09-25.

[25] Wavemetrics.com. https://www.wavemetrics.com/products/igorpro/dataanalysis/signalprocessing/hilberttransform. Accessed: 2018-10-07.

[26] translated to English by Joachin Stöckler Patrick Flandrin. *Time-Frequency/Time-Scale Analysis, vol 10*. Academic Press, 1999. ISBN 978-0-12-259870-8.

[27] Short-time fourier transform, mathematical definition. -https://ccrma.stanford.edu/~jos/sasp/Mathematical_Definition_STFT.html, . Accessed: 2018-12-05.

[28] Boualem Boashash. Note on the use of the wigner distribution for time-frequency signal analysis. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(9):1518–1524, 1988. URL https://doi.org/10.1109/29.90380.

[29] Aalborg university, lecture by josé biurrun manresa. https://person.hst.aau.dk/enk/ST8/Lecture2_Slides.pdf. Accessed: 2018-10-23.

[30] Matlab's website. https://se.mathworks.com/help/matlab/ref/fft.html. Accessed: 2018-10-22.

[31] Fredric J. Harris. On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, 66(1):51–83, 1978. URL http://web.mit.edu/xiphmont/Public/windows.pdf.

[32] Mathworks: hamming window. https://se.mathworks.com/help/signal/ref/hamming.html. Accessed: 2018-10-23.

[33] Pearson correlation. -https://libguides.library.kent.edu/SPSS/PearsonCorr. Accessed: 2018-11-27.

[34] minitab.com. -https://support.minitab.com/en-us/minitab-express/1/help-and-how-to/modeling-statistics/regression/supporting-topics/basics/linear-nonlinear-and-monotonic-relationships/, . Accessed: 2018-11-27.

[35] quoracdn.net. -https://qph.fs.quoracdn.net/main-qimg-ea97a91276f3d8d711434cc99130286a, . Accessed: 2018-11-27.

[36] Spearman correlation. -https://statistics.laerd.com/statistical-guides/spearmans-rank-order-correlation-statistical-guide.php. Accessed: 2018-11-27.

[37] Sima et al. The reliability of short-term measurement of heart rate variability during spontaneous breathing in people with chronic obstructive pulmonary disease. *Revista Portuguesa de Pneumologia*, 23(6), 2017. URL 10.1016/j.rppnen.2017.06.001.

[38] Cellini et al. Heart rate variability during daytime naps in healthy adults: Autonomic profile and short-term reliability. *PSYCHOPHYSIOLOGY*, 53(4), 2016. URL 10.1111/psyp.12595.

[39] Weiner et al. Test-retest reliability of pediatric heart rate variability: A meta-analysis. *Journal of Psychophysiology*, 31(1), 2017. URL 10.1027/0269-8803/a000161.

[40] Hernando et al. Validation of the apple watch for heart rate variability measurements during relax and mental stress in healthy subjects. *Plos One*, 18(8), 2018. URL 10.3390/s18082619.

[41] Koo et al. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2):155–163, 2016. URL https://doi.org/10.1016/j.jcm.2016.02.012.

[42] McGraw et al. Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1):30–46, 1996. URL http://doi.apa.org/getdoi.cfm?doi=10.1037/1082-989X.1.1.30.

[43] Laplace distribution image. -https://upload.wikimedia.org/wikipedia/commons/thumb/0/0a/Laplace_pdf_mod.svg/325px-Laplace_pdf_mod.svg.png. Accessed: 2018-12-07.

[44] Roc curve. https://cdn-images-1.medium.com/max/1600/1*pk05QGzoWhCgRiiFbz-oKQ.png. Accessed: 2019-01-21.

[45] Hajian-Tilaki et al. Receiver operating characteristic (roc) curve analysis for medical diagnostic test evaluation. *Caspian Journal of Internal Medicine*, 4 (2):627–635, 2013. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3755824/.

[46] Timothy L. Lash Kenneth J. Rothman, Sander Greenland. *Modern Epidemiology, 3rd edition*. 2008. ISBN 978-1-4511-9005-2.

[47] Cornell university: Censored data. http://www.cscu.cornell.edu/news/statnews/stnews67.pdf, . Accessed: 2019-01-24.

[48] Cornell university: What is survival analysis? https://www.cscu.cornell.edu/news/statnews/stnews78.pdf, . Accessed: 2019-01-24.

[49] Princeton: Survival models. https://data.princeton.edu/wws509/notes/c7.pdf. Accessed: 2019-01-31.

[50] Richard Kay. An explanation of the hazard ratio. *Pharmaceutical Statistics*, 3(4): 295–297, 2004. URL https://doi.org/10.1002/pst.119.

[51] Uk biobank assessment data. http://biobank.ctsu.ox.ac.uk/~bbdatan/Repeat_assessment_doc_v1.0.pdf. Accessed: 2019-01-22.

[52] Orini et al. Long-term intra-individual reproducibility of heart rate dynamics during exercise and recovery in the uk biobank cohort. *Plos One*, 13(2), 2017. URL https://doi.org/10.1371/journal.pone.0193039.

[53] Costa et al. Design of time-frequency representations using a multiform, tiltable exponential kernel. *IEEE Transactions on Signal Processing*, 43(10), 1995.

[54] Bland et al. The logrank test. *the British Medical Journal*, 328(7447), 2004. URL https://dx.doi.org/10.1136%2Fbmj.328.7447.1073.

[55] Ishwaran et al. Random survival forests. *The Annals of Applied Statistics*, 2(3): 841–860, 2008. URL http://dx.doi.org/10.1214/08-AOAS169.

[56] Matchit package description. https://cran.rproject.org/web/packages/MatchIt/index.html, . Accessed: 2019-02-26.

[57] D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220, 1972. URL https://www.jstor.org/stable/2985181.

[58] Aune et al. Diabetes mellitus and the risk of sudden cardiac death: A systematic review and meta-analysis of prospective studies. *Nutrition, Metabolism and Cardiovascular Disease*, 28(6):543–556, 2018. URL https://www.ncbi.nlm.nih.gov/pubmed/29730085.

[59] Aune et al. Diabetes mellitus, blood glucose and the risk of heart failure: A systematic review and meta-analysis of prospective studies. *Nutrition, Metabolism and Cardiovascular Disease*, 28(11):1081–1091, 2015. URL https://www.ncbi.nlm.nih.gov/pubmed/30318112.

[60] Thomas M Maddox Benjamin M Leon. Diabetes and cardiovascular disease: Epidemiology, biological mechanisms, treatment recommendations and future research. *World Journal of Diabetes*, 6(13):1246–1258, 2015. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4600176/.

[61] Diabetes and heart disease. https://www.diabetes.org.uk/guide-to-diabetes/complications/cardiovascular_disease. Accessed: 2019-03-04.