Universitetet
i Stavanger

FACULTY OF SCIENCE AND TECHNOLOGY

MASTER'S THESIS

Study program/specialization: Information Technology - Automation and Signal Processing	Spring semester, 2019 Open / ~~Confidential~~
Author: Vebjørn Kaldahl Bottenvik	……………………………………… (Signature of author)
Supervisor(s): Hein Meling, Trygve C. Eftestøl, Ståle Freyer.	
Title of master's thesis: Biometric Authentication from ECG Signals on Wearable Devices. Norwegian title: Biometrisk autentisering fra EKG-signal på kroppsnære enheter.	
Credits: 30	
Keywords: Biometrics, Authentication, Electrocardiogram (ECG), Machine Learning	Number of pages: 69 + supplemental material/other: 21 + zipped file Stavanger, 14th of June 2019 date/year

University of Stavanger

Department of Electrical Engineering and Computer Science

Master Thesis in Automation and Signal Processing

Biometric Authentication from ECG Signals on Wearable Devices

---

Author:

Vebjørn Kaldahl Bottenvik

Supervisor(s):

Hein Meling

Trygve C. Eftestøl

Ståle Freyer

June 14, 2019

University of Stavanger

Abstract

Biometric authentication is currently being used for numerous devices; such as mobile phones, computers, etc. However, for now, the only authentication methods for wearable devices are those of passwords and pin codes. The newest instance of the Apple Watch series 4. has an integrated Electrocardiogram (**ECG**) recording possibility that could be used for biometric authentication. Having the possibility for biometric authentication on wearable devices could potentially provide seamless authentication applications as the wearable device is always on standby.

The objective of this thesis was to test biometric authentication based on ECG signals recorded on wearable/mobile devices. By collecting data from a set of volunteers with recordings performed under different circumstances such as; resting heart rate, increased heart rate after exercise, and noisy signals while in motion. By performing denoising and feature extraction, various machine learning models were trained and evaluated to provide a classification model that performed well on the variety of ECG signals. The classification model was further used to present a biometric authentication system.

The biometric authentication system presented in this thesis was tested on three different sets of acquired ECG data. Biometric authentication based on ECG signals recorded with resting heart rates correctly authenticated 17/19 subjects, resulting in an acceptance rate of 89.5%. For the recordings after physical activity and in motion, the authentication system correctly authenticated 52.6% (10/19) and 31.6% (6/19) of the subjects. An additional subject that had been excluded from the system did not get authenticated for either of the different recordings. Overall, no subjects were misclassified as other subjects.

Acknowledgements

List of Figures

List of Tables

Glossary

interindividual Interindividuality refers to variability between people. 12

intraindividual Intraindividuality refers to variability in one person. 12

lead Imaginary line between two electrodes. 7

Acronyms

ACC Accuracy. 15

AUC Area Under the Curve. 33

AV Atrioventricular. 6

ECG Electrocardiogram. i, 1

EER Equal Error Rate. 33

FN False Negative. 15

FP False Positive. 15

HRV-ECG Heart Rate Variety ECG. 19

HRV Heart Rate Variability. 10

ICA Independent Component Analysis. 13

M-ECG Moving ECG. 19

ML Machine Learning. 13

OVR One-vs-Rest. 32

PCA Principal Component Analysis. 13

QTc Corrected QT interval. 31

R-ECG Resting ECG. 19

ROC Receiver Operating Characteristics. 33

SA Sinoatrial. 5

TN True Negative. 15

TP True Positive. 15

bpm beats per minute. 10

Contents

Chapter 1

Introduction

In recent years the use of biometric data for authentication has become popular and is increasingly replaced the need for pin codes and passwords in many consumer devices. Biometric data, such as fingerprints and face morphology, can be acquired by most of the newest devices that are currently being launched. Despite increasingly taking over for pin codes on mobile phones, this trend has yet to reach the domain of wearable devices. Wearable devices do not contain the hardware required to acquire most of the common biometric data used for authentication purposes. However, the latest wearable devices such as the Apple Watch series 4 comes with the possibility to record Electrocardiogram (**ECG**) signals. The availability of biometric authentication applications for wearable devices could take over for some of today's mobile authentication applications, as a wearable device is more accessable than mobile devices.

1.1 Previous Work

The use of ECG for biometric identification and authentication of individuals has been the motivation for a variety of publications over the last 20 years, with an increase of publications for the last few years. One of the first to investigate the possibility of ECG biometrics was Biel et al. [10]. They used the medical grade ECG recorder and classical Machine Learning algorithms. Following their work, publications using different methods for pre-processing, ECG acquisition and classification methods have been published. Most of the research has been performed on ECG signals from large databases such as; MITDB [11] and the ECG-ID [12]. These databases contain ECG signals recorded on medical grade equipment and provide higher resolution than those recorded on mobile devices.

A 2016 study by Choi et al. explored the possibility of using noisy signals from a mobile ECG sensor [13]. They found that biometric identification could be accomplished with results comparable to those of medical grade ECG recorders. However, the signal obtained by the mobile sensor was noisier than the data found in MITDB; thus, additional work on pre-

processing had to be accomplished. The result lead to an identification accuracy of 95.99%. A more recent and less explored classification method for ECG biometrics is performed by Neural Networks. Salloum and Jay Kuo proposed using this method to identify individuals from both the MITDB and ECG-ID datasets and ended up with a maximum 100% classification score [14]. Lugovaya [15] used a combination of Neural Networks and simple template matching to achieve 100% accuracy as well.

Although publications have accomplished accurate identification rates based on signals acquired with a resting heart rate, there are a few that has investigated the effect increased heart rates had for identification. A 2014 study by Pathoumvahn et al. [16] tested the robustness of biometric identification with increased heart rate and found that the accuracy decreased with up to 20% for a 20% increase in heart rate and up to 70% for heart rate increased by 40%.

Although the ECG signals for biometric identification and authentication purposes have shown promising results in most studies, more methods and generalizations have not yet been fully explored. Pinto et al. published a paper June 2018 [17] where they evaluated a variety of published work on the field. They proposed that some additional factors that should be considered to give a realistic estimate of the overall ECG identification accuracy. Some of these considerations have been used as motivation for the proposed method in Chapter 3.

1.2 Problem Description

The goal of this thesis is to investigate the possibility to perform biometric authentication from ECG signals acquired under different conditions from wearable or mobile devices. The data should be acquired by a set of volunteers and should be recorded under different conditions to provide data with increased heart rates, and increased noise. The project can be decomposed into three main steps. The first step is to develop a robust pre-processing method for noisy ECG signals. The second step is to find a suitable classification method that provides accurate identification for the different ECG signals acquired. The third step is to use this classifier to develop a biometric authentication system. The biometric system should be tested on ECG signals acquired under the different conditions to evaluate how changes in heart rate and noise affects a biometric authentication system.

The process of this thesis can be listed as follows:

1. Collect data from a set of volunteers.

2. Implement a robust pre-processing algorithm based on the research of noisy ECG signals.

3. Fiducial feature extraction.

4. Classifier evaluation.

5. Test biometric authentication based on the best performing classification model.
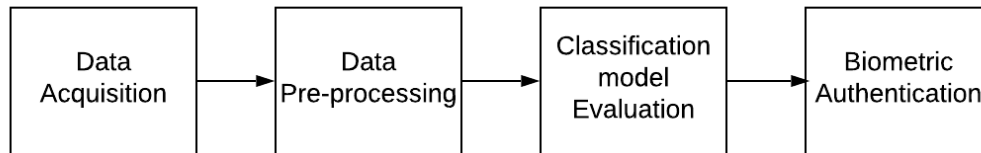
Figure 1.1: Block chart presenting a simplified overall progression of this thesis.

1.3   Thesis Outline

Chapter 2: Theory

This chapter provides relevant background information of Electrocardiogram, Biometrics, Machine learning and finally some usefull ECG signal processing techniques.

Chapter 3: Method

This chapter goes through the pre-processing system. The system includes adaptive denoising, feature extraction based on fiducial marks, classifier evaluation and an experimental setup.

Chapter 4: Results

This chapter presents the results obtained by the pre-processing, classifier evaluation and the final biometric authentication experiments.

Chapter 5: Discussion

This chapter discusses results and method, presents some threats to validity, some concluding remarks and finally some directions for future work.

Chapter 2

Theory

This chapter contains the required background information needed for this thesis. The theory consists of background information about the electrocardiogram, biometrics, machine learning, and some beneficial electrocardiogram signal processing methods.

2.1 Electrocardiogram

The ECG is a measurement of the electrical activity of the heart and describes the voltage variation of the different cardiac cells that build up the heart [18]. Today electrocardiography is an essential part of an initial evaluation of cardiac complaints and offers a non-invasive and cost-effective tool to evaluate different heart diseases and arrhythmias [19].

2.1.1 The Heart

The heart is a muscle that provides oxygen-rich blood to the body. The heart can be divided into four chambers, consisting of two sides. The two "mirrored" sides of the heart are connected to different circular systems. However, they both pump in a rhythmic and synchronized manner. The chambers are; the right and left atrium, which is where the blood enters, and the left and right ventricles, where the blood is forced out through the body for blood circulation [18]. Figure 2.1 illustrates how the left and the right side are connected to different vessels, where the right side circulate blood to the lungs, and the left side circulates blood to the rest of the body.

The heart consists of cardiac cells, which, in their resting state, are electrically polarized. What this essentially means is that the outside of the cell is positively charged, while the inside is negatively charged. This charge is maintained by ions being pumped into the cells through individual ion channels in the cell membranes. The cardiac cells can lose their internal charge in a depolarization process. Depolarization is the process in which creates the heartbeat, and for some of the cardiac cells, this is a process that happens spontaneously,
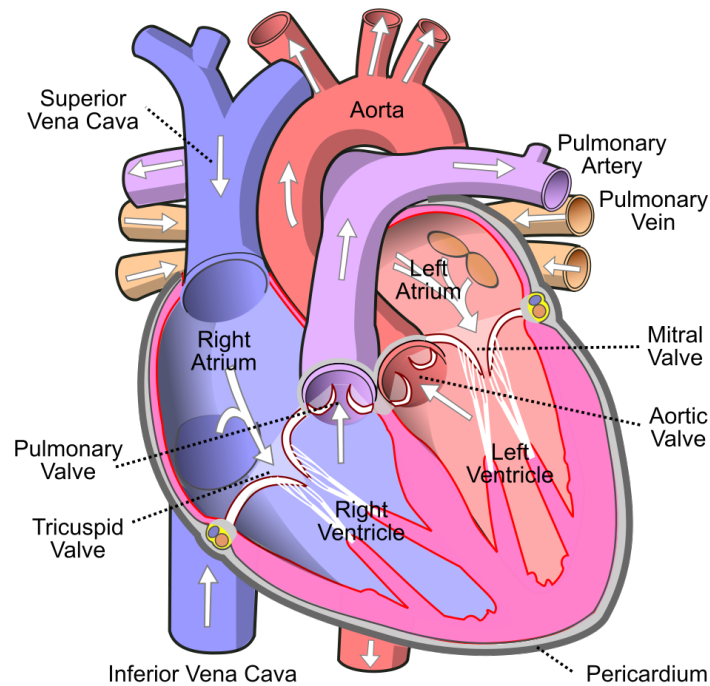
Figure 2.1: Labeled illustration of the human heart [1][1]. This figure illustrates the four chambers of the heart and their connections to the rest of the body.

while some cells must be "forced" into this state. After the depolarization process is complete, the cells go back to their original state through a process called repolarization. These two states of the cardiac cells are what makes the heart beat continuously, as will be explained in chapter 2.1.2. The cardiac cells can be divided into three different types [9], as shown in table 2.1.

2.1.2 The Cardiac Cycle

The heartbeat is the action that makes the heart pump the blood through the body. Each heartbeat consists of a series of events in which the cells are being depolarized. These events start in a dominant group of pacemaker cells called the Sinoatrial (**SA**) node. For simplicity, the heartbeat can be divided into three different events that generate the characteristic waveforms of the ECG, containing the P-wave, the T-wave, and the QRS complex [9].

1. **Atrial Depolarization**:
   The SA node starts its periodical depolarization; this will spread across the electrical conducting cells and to the myocardial cells. The atrial depolarization results in a small "burst" of electrical activity in the heart and is noticeable on the ECG reading as the P-wave.

2. **Ventricular Depolarization**:

---

[1]CC BY-SA 3.0 https://creativecommons.org/licenses/by-sa/3.0/legalcode

When the atrium is fully depolarized, it will activate the Atrioventricular (**AV**) node. The AV node function as the coupling between the atrium and the ventricles and will put the electrical activity to rest. This results in the PR-segment seen in Figure 2.2 and lasts for a small fraction of a second before beginning the ventricular depolarization. The ventricular consists of two sets of bundles. The left and the right bundle are both divided into multiple small branches called Purkinje fibers. The left side is larger than the right side, and the contractions of these results in the iconic QRS-complex. Because the ventricles consist of more muscle mass than the atrium, the QRS-complex results in a larger burst of electrical energy than the atrial depolarization, hence the QRS-complex has a larger amplitude than the P-wave. This also explains the fact that it is impossible to see atrial repolarization because this is happening simultaneously with the ventricular depolarization.

3. **Ventricular Repolarization**:
   As the ventricular depolarization has been completed, there is a small pause in the electrical activity of the heart resulting in the ST-segment as seen in Figure 2.2, before the ventricles are repolarized. The repolarization of the ventricles is a slow process relative to the depolarization. Hence the T-wave generated from this event is wider than the QRS-complex.

The events can be seen in Figure 2.2.

Cell type	Function
Pacemaker cells	The pacemaker cells are cells that are in a continuously depolarization/repolarization process. Normally this happens at a rate of 60 to 100 times per minute. However, this rate changes depending on the activity of the autonomic nervous system.
Electrical conducting cells	The "wiring" of the heart. These cells will lead the depolarization from the pacemaker cells to the myocardial cells. Both the atrium and the ventricles have a "conducting system" made up of these cells.
Myocardial cells.	The contracting cells of the heart. These are the dominant cells in the heart. It is the process of depolarization and repolarization of these cells that result in a heartbeat.

Table 2.1: The different types of cell found in the heart with descriptions [9].

2.1.3 Recording the ECG

As explained, the ECG is the measurement of electrical activity in the heart. The electrical activity can pinpoint the current activity of the heart and is most commonly used for medical analysis purposes. The recording device itself consists of electrodes and simple electronic
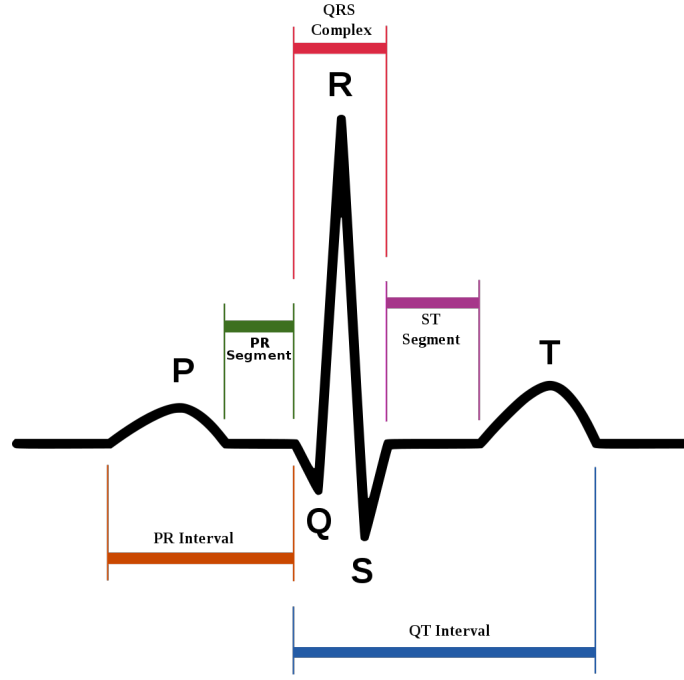
Figure 2.2: Illustration of a typical P-QRS-T complex for a single heartbeat with annotations of waves and intervals [2].

components to compare the voltage potential in the electrodes. Different devices use two, three, or ten electrodes to measure electrical activity, where additional electrodes provide additional views of the heart. A typical recording device found in hospitals has ten electrodes covering all the limbs; left arm, right arm, left leg and right arm, six electrodes placed on the patient's chest and one placed on the right leg as a reference. This gives a full 360°view of the heart, in both vertical and horizontal direction as seen in Figure 2.3.

Based on the number of electrodes used, and the placement of the electrodes, different leads can be found. A lead is an imaginary line between two electrodes, used to illustrate the electrical activity for the different views of the heart [20]. ECG recording can be separated into three different categories where there are six limb leads and six precordial leads. Each lead provides a new "view" of the heart and might include information that cannot be seen in the other leads. In total, all these leads add up to 12 different views of the heart [9].

1. **The bipolar limb leads**
   The leads denoted as I, II, and III in Figure 2.3. These can be recorded using two electrodes and depends on placement. The electrodes can be connected to; left arm (LA), right arm (RA), or the left leg (LL). The leads are then calculated as follows:

$$I = V_{LA} - V_{RA} \tag{2.1}$$

$$II = V_{LL} - V_{RA} \tag{2.2}$$
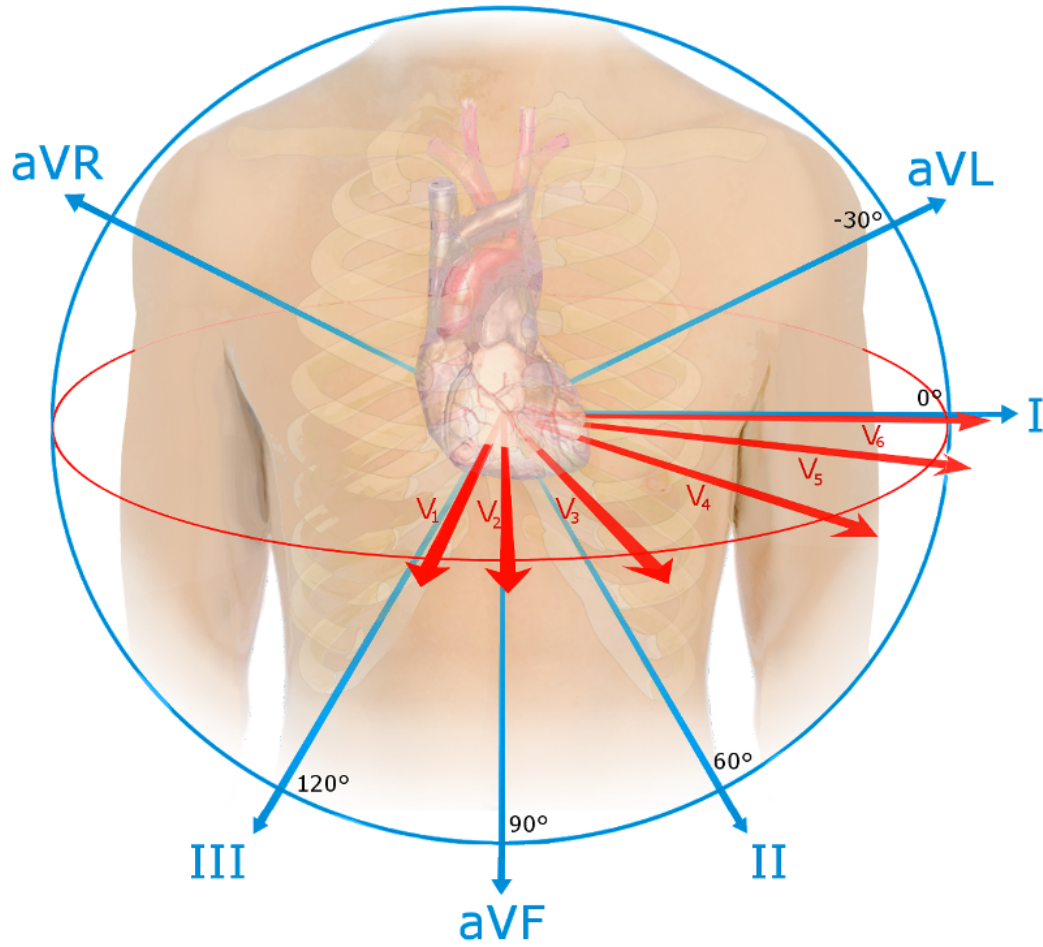
Figure 2.3: Illustration of the different angles that are being monitored in a 12 lead ECG. The blue arrows describe the bipolar and augmented limb leads, and the red arrows illustrates the precordial leads[2][3].

$$III = V_{LL} - V_{LA} = II - I \tag{2.3}$$

As seen in Figure 2.3 lead I, described in Equation 2.1 provides a 0°view of the heart, where lead 2 and 3, described in Equation 2.2 and 2.3 describes the electrical activity with 60°and 120°orientations.

2. **The augmented unipolar limb leads**

   The leads are denoted as aVR, aVL, and aVF. These are found by using three electrodes, where one is "exploring" while the average of the other two serves as a reference [18, p. 420].

$$aVR = V_{RA} - \frac{V_{LA} + V_{LL}}{2} \tag{2.4}$$

$$aVL = V_{LA} - \frac{V_{RA} + V_{LL}}{2} \tag{2.5}$$

$$aVF = V_{LL} - \frac{V_{LA} + V_{RA}}{2} \tag{2.6}$$

3. **The precordial leads**

   The leads are denoted V1 through to V6. These are found using six leads placed on the front left side of the chest and are used for a more detailed view of the heart.

Equations 2.1 - 2.6 from [18].

Various ECG devices provide different possibilities when it comes to recording the ECG. Professionals are using a full 12-lead ECG recorder to provide high-resolution ECG with 12 different views of the heart. Medical grade ECG can be used to discover numerous diseases related to the heart. Newer consumer devices provide two electrodes, thus resulting in a single lead ECG recording. This has been implemented in the newest instance of the Apple Watch, shown in Figure 2.4. The use of single-lead ECG can be used for real-time analysis and provides analytic tools that can discover arrhythmia [21].

Figure 2.4: The Apple watch series 4 with one electrode on the back, and one electrode at the crown[3][4].

2.1.4 Noise

A common problem with recording and monitoring of weak electrical signals is handling noisy signals. The common factors for noise originate from muscle movement, power line noise, and high-frequency noise [22]. A mobile device that is running on batteries will not experience the same level of power line noise. However, there is no escaping the noise generated from muscle movement and the high-frequency noise. Most ECG equipment today have built-in filters to handle most of the different types of noise. Denoising of a noisy signal is always a

---

[3]CC BY-SA 4.0

compromise. The "cleaner" the signal, the more information will get lost. For the problem of identification, one can allow some loss of information. However, excessive denoising might erase the differences that make identification impossible. It is therefore crucial that as much as possible of the characteristics and information of the signals stay unfiltered. The denoised signal gives a more accurate feature extraction and provides a more precise identification result.

2.1.5 Heart Rate Variability

Heart Rate Variability (**HRV**) is a phenomenon in which the intervals between the heartbeats do not come at a predictable interval [18]. If a subject has a heart rate at 60 beats per minute (bpm), this means that there might not be one heartbeat each second. However, these might vary. If the intervals at a certain heart rate are nearly constant, the HRV value is low, and if the variety between heartbeat intervals changes from one heartbeat to another, with an approximate constant heart rate, the HRV value is high.

2.2 Biometrics

Biometrics is the collection of biological patterns that characterizes a person based on physical- or behavioral characteristics. Biometric data represent physical attributes and is, therefore, unique. The property of being unique allows biometric data to be used for methods of identification and authentication. Biometric identification is the process of finding the identity of an individual based on some biometric data. The biometric data is then compared against others, to find whom the biometric data belongs to. Biometric authentication is the process of verifying an individual's identity based on biometric data. Implementation of biometric authentication can be the use of fingerprints as a replacement for pin codes or passwords. Biometric authentication systems require protocols for identity proofing and encryption protocols, while biometric identification provides simple matching [23]. For the biometric data to be valid for either identification or authentication, some requirements listed in Table 2.2 has must be fulfilled [24].
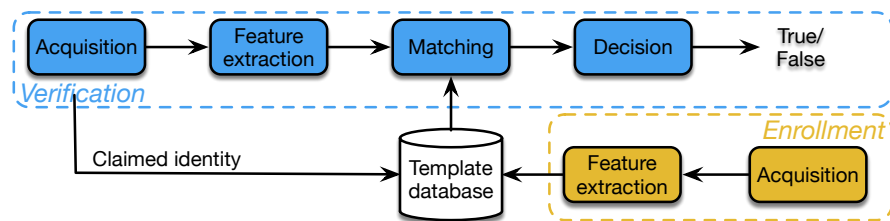
Figure 2.5: Biometric process with enrollment and matching. Figure is an adaption from [5].

Characteristic	Description
Universal	The signal can be found for all individuals.
Unique	The signal is different from one individual to another.
Permanent	The signal is permanent. However, allowing changes over time.
Recordable	The signal must be voluntarily recordable and cannot be recorded without the individual's consent.
Measurable	The signal can easily be measured from some sort of recording device.
Forgery-proof	The signal cannot be forged.

Table 2.2: Table over the different aspects that makes a signal valid for biometric systems.

2.2.1 Common Biometric Standards

Today biometrics to verify identity are broadly used in multiple applications. Biometric methods can be divided into two categories, static and dynamic. Biometrics based on static features use scanners or images to identify individuals, where some of the methods are listed in Table 2.3. Biometrics based on dynamic features uses recordings over time in order to identify individuals. This requires recordings over time and adds a layer of security based on the "aliveness factor" to be identified. Some methods have been listed in Table 2.4. Both the tables contain information from a 2019 biometric review by Rui and Yan [25].

Method	About
Fingerprint	Fingerprints have, for a long time, been the most common biometric standard for identification and authentication. The method has been proven to be stable, and fingerprints have proven to be unique for individuals.
Face ID	Facial recognition software has proven to be capable of using facial structures to identify individuals due to different features such as spacing between eyes etc. A problem with face recognition is that some structures of the face are similar between individuals. Another problem is different camera angles and illumination, and it can cause a high rejection rate.
Iris Recognition	The iris has been proven to be unique for each person. However, the hardware required for iris scanners is more than those of face-ID. Therefore, this is not a broadly used biometric standard for mobile devices.

Table 2.3: Example of some biometric standards with static features.

Method	About
ECG	Biomedical signals such as the ECG has proven to be unique enough to be used for identification and provides "aliveness detection".
Voice Recognition	Using a person's voice to identify a person has proved to be a simple and efficient method. However, it has been proven that these recordings can be fooled without difficulties.
Keystroke and Touch Dynamics	Using a person's keystroke pressure and dynamics to identify the individual.

Table 2.4: Example of some biometric standards with dynamic features.

2.2.2 Biometric Data Encryption

Authentication to gain access to locations, devices, and web services is something that happens daily for most people. Access cards and passwords are perhaps the most common method for authentication for everyday use. Every day people lose their passwords etc. due to hacking and theft, however, this is easily fixed by changing passwords or gaining new access cards. Biometric data, however, does not provide a simple solution if lost. Therefore, it is important that biometric data gets encrypted such that data on individuals does not get lost. For mobile devices, local encryption has been used to avoid third-party access to biometric data. The mobile device thus only gives the authentication result instead of the biometric template itself [26]. Biometric data can also contain sensitive information, such as diseases from ECG signals and iris scanning. It is therefore important that the data does not get in the wrong hands but stay encrypted.

2.2.3 ECG as Biometrics

From table 2.4 it is stated that the ECG signal is a dynamic biometric feature. For ECG to be a valid metric for biometrics, it must fulfill all the standards in Chapter 2.2. The uniqueness of ECG signals comes from age, sex, height, weight, body mass index, ethnicity; and much more [27]. These factors provide interindividual variability [28], however giving that time-varying factors such as increased heart rate and stress also affects the ECG, thus providing intraindividual variability. Most of the biometrics being used for everyday applications is adaptive and is always updating the collected template[4] to compensate for the change over time that is expected for biological recordings. However, the additional short time effects e.g. due to increased heart rate, and stress makes ECG biometrics difficult under certain circumstances.

---

[4]A template is a collection of extracted features.

2.3   Machine Learning

Machine Learning (**ML**) is perhaps one of the fastest growing areas of computer science, with numerous of various applications. ML can be used to perform advanced and complex computations [29]. By having access to large sets of data, ML uses the provided data to learn from the past and based on the presented data make predictions on future data. The data consist of two or more classes and is trained based on feature vectors. Each feature in the feature vectors provide one dimension, thus having feature vectors with length n creates a n-dimensional feature space. To train a classification model, the learning method depends on if the feature vectors have been labeled or not and can be divided into supervised and unsupervised learning.

2.3.1   Supervised Learning

Supervised learning is performed when the actual state, also known as the label, for a set of data is known. Consequently, the data is used to train a classifier by dividing the different sets of data based on their actual state. When new data that is not labeled is presented to the classifier, it will, based on different metrics provide the state which has the highest score. The metrics used to estimate which state the new data belongs vary from model to model [29]

2.3.2   Unsupervised Learning

Unsupervised learning is the method where only the data is known, hence the actual state, or label, of each of the classes, are unknown. Unsupervised learning is done by clustering methods or associations, where the likelihood of the samples is used to group the samples into classes [7]

Component Analysis

Component Analysis is an unsupervised method of finding directions in the feature space and use these directions to generate new features. The algorithms have different goals and can be used to find features that describe different aspects of the feature space and use features that do not provide relevant information to lower the dimension of feature space. Principal Component Analysis (**PCA**) projects the data into dimensions that describe the variance of the features, thus can be used to reduce features that are correlated to each other. Independent Component Analysis (**ICA**) is used to find the directions of the feature space that show the independence of the signals [30].

2.3.3   Feature Scaling

For most ML models the need for feature scaling provides more accurate classifiers. Hence, feature scaling is a common step in the pre-processing algorithm. Feature scaling can be accomplished with different methods, where feature normalization and feature standardization are two of them [31]. Feature standardization is performed by estimating the mean and variance for each of the features in the feature vector. Then, for each feature, calculate the scaling according to Equation 2.7.

$$\hat{x} = \frac{x - \mu}{\sigma} \tag{2.7}$$

Another known method for feature scaling is by performing feature normalization. Feature normalization scales the features based on the maximum and minimum values of each feature. This can be calculated according to Equation 2.8. This method is more sensitive to outliers and can provide un-even scaling.

$$\hat{x} = \frac{x - \mu}{max(x) - min(x)} \tag{2.8}$$

2.3.4   Training and Validation

Training the classifier is done in order to make the classifier learn the patterns of the given data, either supervised or unsupervised. To find the best classification for the provided data, multiple different classification models should be evaluated. The "No Free Lunch" theorem states that there are no context-independent nor usage-independent reason to favor one classifier over another [30]. Validation of a a machine learning model is done by testing some data that was not introduced under training are tested on the model.

Figure 2.6: Confusion matrix showing the relationship between predicted class and actual class.

Validation of the classification model is done by performing prediction on the validation data. Validation data is data that is labeled and used to verify that the classifier predicts

the true label. For a binary classification problem[5] the predictions will be one of two classes. If the first class is positive, and the second class is negative, the different predictions are as illustrated in Figure 2.6. The predictions consist of True Positive (**TP**) and True Negative (**TN**) if the correct classes have been predicted, and False Positive (**FP**) and False Negative (**FN**) if the class is predicted as the wrong class [32]. From this the Accuracy (**ACC**) can be calculated as shown in Equation 2.9.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.9}$$

2.3.5 Tuning

Different classifiers require different parameters to generate the decision boundaries for the classifier. By setting up a grid with different parameters, the model can be trained with different parameters. The parameters that provide the best ACC for the classification model is then used. Though hyperparameter estimation with grid search is a time-consuming process, it can provide better classification models. Random search, like grid search, tests different parameters. However, does provide random parameters. Both the methods are being used with cross-validation to give a generalized result [33].

2.4 ECG Signal Processing

The ECG signal has for many years been used for medical analysis, resulting in numerous toolboxes and libraries in both Matlab and Python. Some of the most important and validated methods for even-detection[6]. In Python, multiple open source libraries are available, where Biosppy [34] is one example that can be used for ECG filtering and QRS detection.

2.4.1 QRS Detection

Analysis of ECG signals starts by detecting the QRS complex. This is a crucial task, as false detection of R-peaks can cause problems further down the ECG analysis. Since noise and some physiological origins can cause suppressed R-peaks and increased T- and P-waves, robust QRS detection should therefore be implemented, and should include a decision rule [18]. In this thesis, the QRS detection algorithm due to Pan and Tompkins has been used [35].

2.4.2 Wave Delineation

Wave delineation is the method of determining the boundaries for each of the different waves in the PQRST complex. This can further be used to calculate the duration of the different

---

[5]A binary classification model only handles two classes.
[6]The detection of certain events in the ECG. This includes peaks and waves.

waves and events in the signal. Hence it is of interest to have a robust wave delineation method. The standard definition of wave delineation is where a wave reaches a certain threshold. Though, in noisy ECG signals, this required threshold may never be satisfied. A suggested approach is to use the changing slope of the differentiated signal to determine a threshold [18], then use the acquired threshold to find the delineation. The method is sensitive to high-frequency noise, and it is therefore suggested to use a low pass filtered differentiated signal.

2.4.3 Heart Rate Estimation

Heart rate estimating is usually accomplished by finding the average RR interval within a given period, then based on the sampling frequency find the estimated heart rate. In the case where ECG is recorded under or after exercise, the heart rate is changing thus averaging the RR intervals does not result in an accurate heart rate estimate. On the other hand, using short time windows gives HRV and ectopic beats[7] more leverage regarding the heart rate estimation [30].
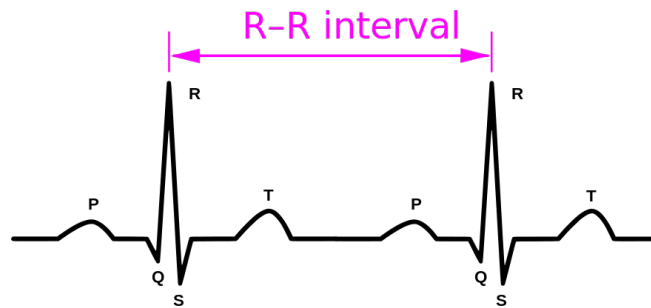
Figure 2.7: RR interval between two heartbeats in the ECG signal [6].

---

[7] An ectopic beat is a disturbance in the cardiac cycle, and is heartbeats that are generated from a focus other than the SA node [9].

Chapter 3

Method

The Method used for this project can be divided into 3 main categories:

1. **Data acquisition**

2. **Data pre-processing**

3. **Machine Learning/Classification**

4. **Biometric authentication**

A flow chart of the method is as shown in Figure 3.1. The pre-processing step consists of de-noising and feature extraction, and the classification step consists of training and evaluation of machine learning models.
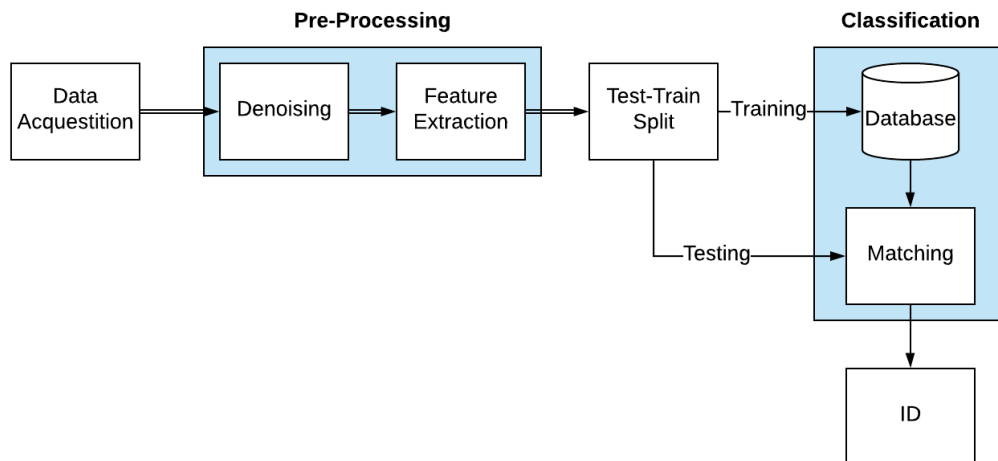
Figure 3.1: Flow chart for the proposed method for this thesis, containing data acquisition, pre-processing and classification.

3.1 Data

This section contains information about the recording device and explains the data acquisition and anonymization process. The data used for this thesis was collected from a set of 20 volunteers, where each volunteer was recorded under three different circumstances. The different circumstances consisted of a one-minute recording where the volunteer was relaxed and had a resting heart rate, a two-minute recording after physical activity, and a 20-second recording where the volunteer was in motion.

3.1.1 Recording device

The recording device used for this project was the Alive Bluetooth Heart and Activity Monitor as shown in Figure 3.2. This is a wireless health and fitness monitor that can be used for identification of atrial fibrillation, heart failure, etc. The device provided a single lead recording, and has a sampling rate at 300 Hz, 8-bit resolution, and a dynamic range of 5.3 mV p-p[1] [36].

Figure 3.2: Alive Bluetooth Heart and Activity Monitor with connections for two electrodes.

From this information the time step Δt and quantification level Δx could be calculated from Equations 3.2 and 3.1.

$$\Delta x = \frac{5.3mV}{2^8 - 1} = 0.0208mV \tag{3.1}$$

$$\Delta t = \frac{1}{300}s = 30e^{-3}s = 30ms \tag{3.2}$$

---

[1]Peak to peak.

3.1.2 Data Acquisition

Data acquisition was planned accordingly to the flow chart shown in Figure 3.3. As introduced earlier, the test consisted of three separate recordings for each of the volunteers. The first recording provided the Resting ECG (**R-ECG**) dataset, which provides the basis for this project, and was a one-minute recording with a resting heart rate. The second recording was done after a brief physical exercise and provided the Heart Rate Variety ECG (**HRV-ECG**)[2] dataset. This dataset was recorded over two minutes and provided ECG signals with a continually decreasing heart rate. The physical activity used for this recording was running in starts, as it provides a fast increase in heart rate. The third recording was performed while the volunteer was in motion and provided the Moving ECG (**M-ECG**) dataset. This was done by making the volunteer walk approximately 15 meters at a slow pace. This recording provided ECG signals with extreme baseline drift and muscle noise. The three different recordings for one of the volunteers has been shown in Figure 3.4. Further, these datasets have been denoted as the R, HRV and M datasets.
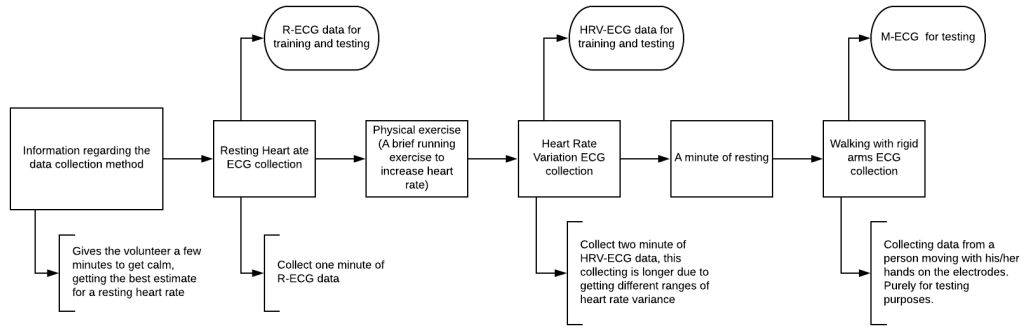
Figure 3.3: Proposed collection protocol. The complete data collection protocol has been placed in Appendix C.

In total, this data acquisition method provided approximately 270 heartbeats for each of the volunteers. These are divided into training and testing data, where 20% of the data is used for testing. The M dataset was used exclusively for testing.

ECG Dataset	Heartbeats Training	Heartbeats Testing
R-ECG	40	10
HRV-ECG	150	50
M-ECG	-	20

Table 3.1: Heartbeats for each data set for the average individual. The training data can further be divided into training and validation or be used for k-fold validation.

The data collected from the recording device is unfiltered, thus a general filtering method that can handle the different noise components for the different data sets needed to be

---

[2]Not to be confused with Heart Rate Variability.

implemented. This is described in Section 3.2.1.
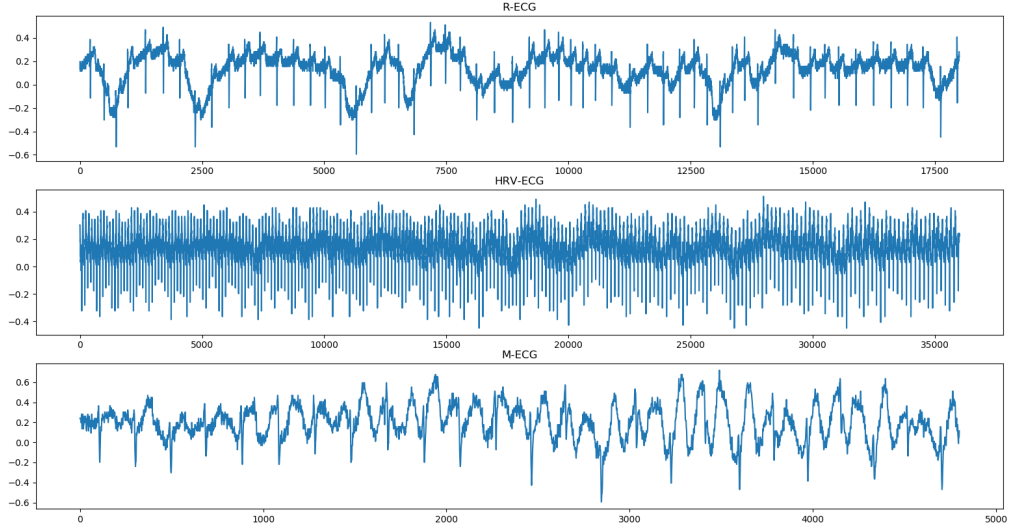
Figure 3.4: Collected data from an arbitrary volunteer. The three plots show the R, HRV and the M datasets.

3.1.3 Anonymization

Biometric data is regarded as sensitive and protected data by the EU and GDPR. Data subject rights [37] states that the subject always should have the authority to access or delete their data from the collection and has strict guidelines when it comes to the collection and storing of said sensitive data. ECG data is bio-medical data that can be used to analyze the healthiness of a person's heart and might reveal different cardiac deceases, which raises moral questions on how to manage this information.

Biometric identification does not require any prior information about age, sex, height, weight, etc. and it is therefore of no interest in keeping personal information for any of the individuals. This gave the possibility to use anonymous data for the project, thus avoiding the strict guidelines from the GDPR [38].

The anonymization process used in this project consisted of a simple script that gave each of the subjects a possibility to choose their subject number. As the number had been decided, a new folder was generated, which contained the subject's ECG data. When the data acquisition process had been finished the folders were renamed from 01-20, as shown in Figure 3.5.
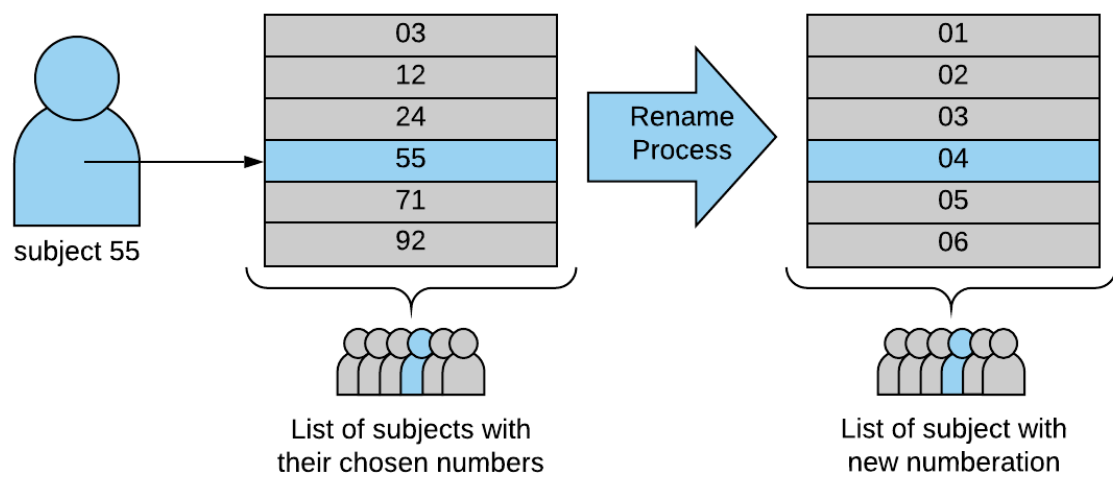
Figure 3.5: Anonymization procedure used for this project. This example illustrates how one subject based on the chosen number get shuffled into the dataset based on the complete list of subjects.

3.2 Pre-processing

This chapter presents the pre-processing methods used for this project. The pre-processing can be decomposed into three main steps:

1. **Denoising**

2. **Feature extraction**

3. **Heart rate normalization and feature space reduction**

To generate data for each of the heartbeats, it was crucial to find and implement denoising that maximized the amount of valuable data that could be used for further feature extraction. Feature extraction methods based on fiducial features in the ECG signal are susceptible to noisy signals. Thus, ECG signals acquired from a mobile device requires robust denoising methods.

Event	Frequency region [Hz]
Heart rate	0.67-3.33
P wave	0.67-5
QRS-complex	10-50
T wave	1-7
Baseline drift	0-1.5
Muscle noise	5-50
Power line noise	50-60
High frequency noise	100-500

Table 3.2: Frequency regions for typical ECG signals

Table 3.2 Contains the different frequency regions expected for the waves and the noise components in the ECG signal [22]. The frequency domains some noise components and signal components overlap, which makes filtering only the noise alone a difficult task.

3.2.1 Denoising

Denoising has an essential function in pre-processing for all types of data, included biomedical signals. One problem with denoising ECG signals is that by definitionthat, the distinct waves and the QRS complex are considered to be noise for the filters. By incorrectly determining cut-off frequencies, these components would be removed from the filtered signal. Therefore, it is essential to study the different frequency regions of a typical ECG signal before designing filters. It is also recommended using linear- or zero phase filters to avoid phase distortion; which can be accomplished by using either FIR filters or backward-forward IIR filtering [18]. In this project, the need for real-time filtering was not important; therefore, the additional delay caused by the FIR filter did not provide further disadvantages.

The different noise components found in ECG signals from Chapter 2.1.4 can be compensated for by implementing a bandpass filter followed by a smoothing filter. The process has been illustrated with a flow chart in Figure 3.6.

Denoising
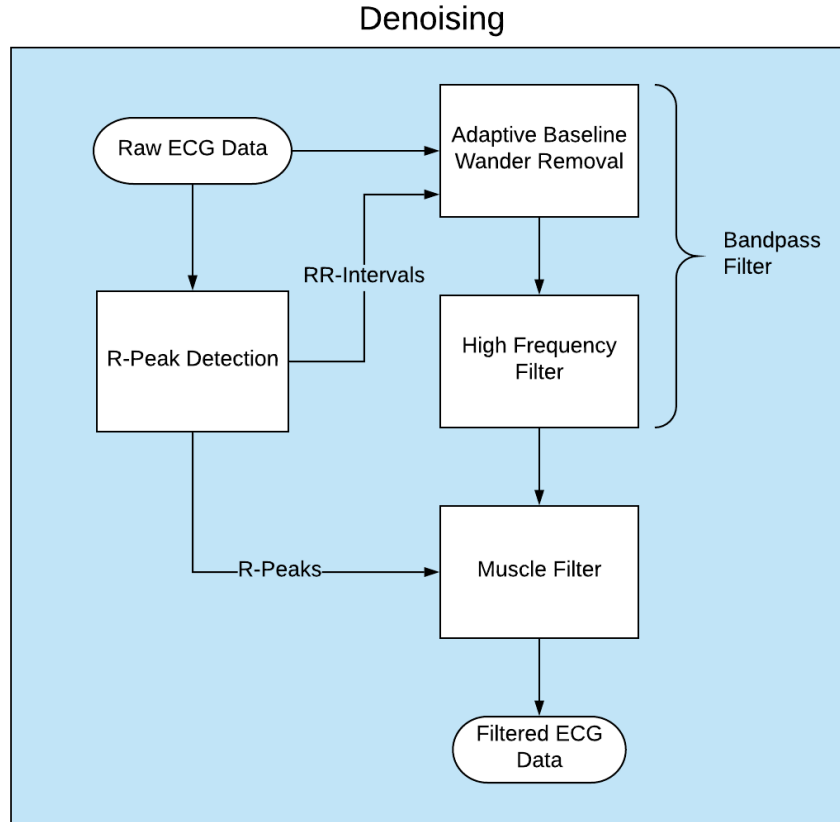
Figure 3.6: Pre-processing flow chart

Adaptive Baseline Wander Removal

Low-frequency noise due to wire movement and respiration affects the ECG signal by making a non-zero and dynamic baseline. High pass filters have been used to remove the low-frequency baseline drift. The baseline drift mostly comes from respiration and wire movement. In most cases, the use of a high pass filter with cut-off frequencies at approximately 0.5 Hz removes the baseline, as the baseline components usually are in a range below 0.5 Hz [18]. Though, in this project, different datasets with excessive baseline wander were present. Using a higher cut-off frequency would overlap with the frequencies of the P- and T-wave as they have low-frequency components. This problem was solved by the implementation of a FIR time-variant baseline filter like the one proposed by Sörmo [39]. This filter uses a cut-off frequency that is proportional on the heart rate, thus making it inversely proportional to the RR interval. This was implemented to preserve as much as possible from the ECG signal; however, at the same time, remove the baseline drift. Limitations on 50 and 180 bpm were set to limit the cut-off frequencies produced by this method. The filter was implemented

as a FIR filter with a filter length of 1145, thus providing 20 dB stopband attenuation [18]. Estimation of heart rate was done according to Equations 3.3.

$$HR_i = \frac{60 \cdot f_s}{R_{i+1} - R_i} \tag{3.3}$$

The heart rate in Equation 3.3 is 60 times the sampling frequency divided by the number of samples between two detected R-peaks. the cut-off frequency range for the filter was set between 0.5 HZ and 3 Hz, and were used on heart rates of 50 bpm and 180 bmp respectively. This filter was used to estimate the baseline, and was thereafter subtracted from the signal.

High Frequency Filter

From Table 3.2, the highest frequency of interest is up to 50 Hz. Thus, higher frequencies than 50 Hz does not contribute to more information but provides noise. To remove these noise components, a FIR lowpass filter with a cut-off at 45 Hz was implemented. In addition to the baseline filter described previously, they formed a cascade time-variant bandpass filter. It could also be used as a bandstop filter, to estimate the baseline with additional noise, and could thereafter be subtracted from the input signal.

Muscle Filter and Smoothing

Noise because of muscle tension, in contrast to other types of noise in the ECG signal, cannot be removed by narrowband [18]. This is because the frequency components of muscle noise are in the range of 5-50 Hz. A standard method for filtering this kind of noise is by signal averaging; in which, multiple heartbeats are necessary. Another technique for muscle noise filtering is by using dynamic Gaussian functions [40]. This was implemented by using QRS detection to locate the QRS-complex, and after that, using Gaussian smoothing on the parts of the signal not containing QRS-complex. The flow chart is shown in Figure 3.7. In addition to filtering the muscle noise this filter compensates for the noise due to low quantification.

The smoothing windows were created from a FIR window filter design by providing a window of 51 samples generated with different sigma values. Also, smoothing requires the area under the window to sum up to one. Thus the Gaussian functions have been scaled accordingly. The two Gaussian windows created are as shown in Figure 3.8
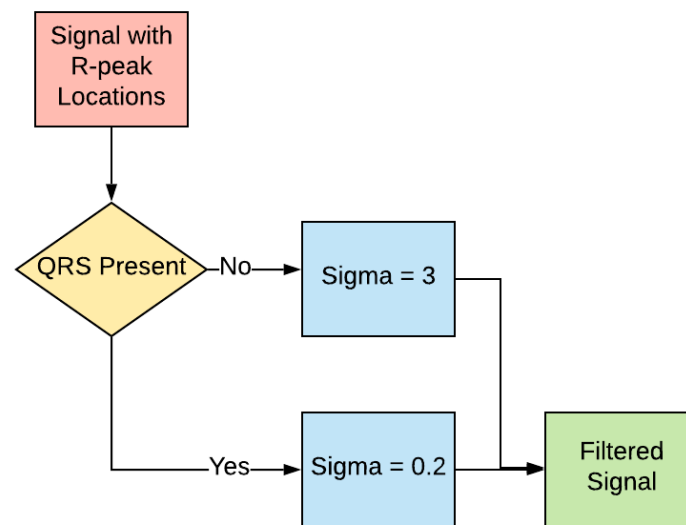
Figure 3.7: Flow chart of the dynamic Gaussian smoothing method.
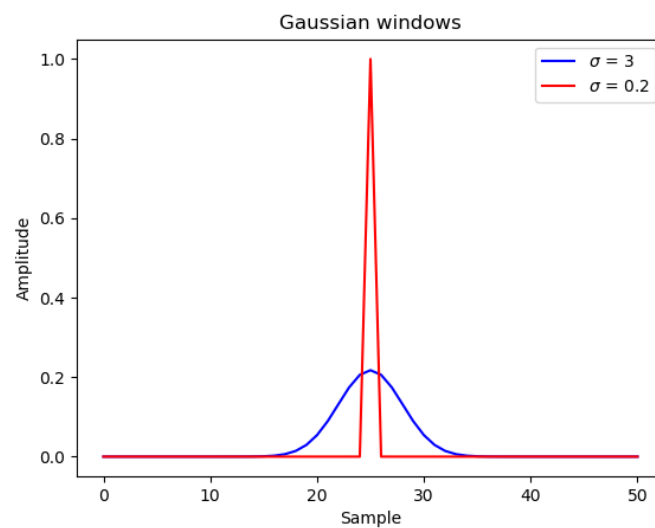
Figure 3.8: Scaled Gaussian windows for $\sigma = 3$ and $\sigma = 0.2$.

3.2.2   Feature Extraction

Feature extraction is a crucial part of supervised learning[3] where points that provide unique characteristics for the different classes of input signals are being used. From Section 2.1, some significant points of the ECG signal based on the various aspects of the heartbeat has been explained. These amplitudes and intervals are some of the key features used in ECG analysis, as they can provide valuable information about the health of the heart.

Feature
R Amplitude
Q Amplitude
S Amplitude
P Amplitude
T Amplitude
ST Elevation
QR Slope
RS Slope
P Interval
P+ Interval
P- Interval
T Interval
PR Segment
QTc Interval
QRS Onset
QRS Interval

Table 3.3: Full list of features used for this project. The list is an adaptation from the feature list by Biel et al. [10].

QRS Detection

A QRS detection algorithm is conceivably an essential tool for ECG signal processing, as most analysis starts with detecting the heartbeats. From Section 2.4.1, it was suggested that the implementation of QRS detection in this thesis would be the method due to Pan and Tompkins.

The modified Pan-Tompkins method proposed by Sathyapriya et al. uses a moving average filter[4] as a replacement for the integration step [7].
The different parameters used for the algorithm is as follows:

- Band Pass filtering with 5-15 Hz passband.

---

[3]Some classification methods such as neural networks perform feature selection, therefore manual feature extraction is not necessary.
[4]Moving average filters averages the n nearest samples, thus provides smoothing the signal.
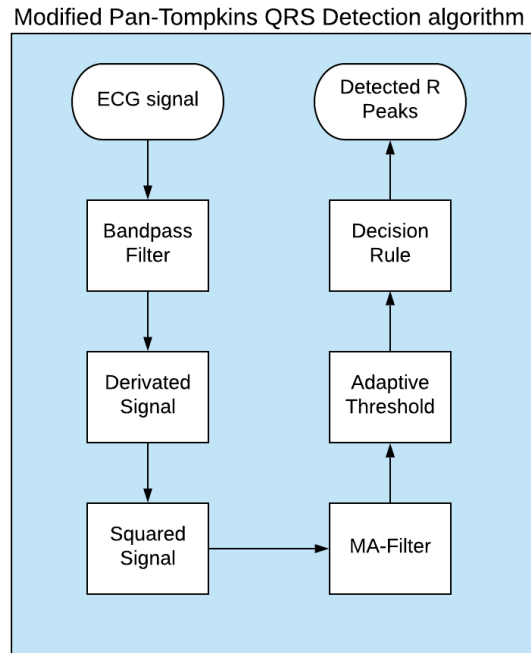
Figure 3.9: Modified Pan-Tompkins QRS detection algorithm. Adaption from [7].

- Forward-Backward MA-Filter with a window size of 15 samples.

- Adaptive threshold to identify peaks.

- Decision rule based on prior knowledge of the QRS complex and time intervals.

The decision rule for the QRS detection algorithm applies two thresholds to decide if the current peak is from noise or is, in fact, the R-peak. The thresholds are continuously changing with the signal to get the best estimate for the level of noise in the signal. The decision method used for this thesis is an adaption of the Matlab implementation by Sedghamiz [41]. The method provides the modified Pan-Tompkins method, where an additional search-back function for detecting missed peaks has been implemented. The decision function can mainly be divided into a four-step process:

1. Detect peaks.

2. Search for missed peaks.

3. Elimination of multiple detections.

4. T-wave discrimination.

The algorithm starts with a one and a half seconds of training time to find the maximum peak to provide a signal threshold and takes one-third of the meaned signal to provide a

threshold for the noise. As the algorithm proceeds, the thresholds for signal and noise are continuously being updated as a new peak for either signal or noise was located. Time limits on 200ms and 360ms were implemented to avoid classifying a T wave or another peak as an R-peak. Since the refactorization time of a heartbeat has a minimum of 200ms, no beats closer to 200ms were added to the list of peaks, and if the time was lower than 360ms the possibility for the signal to be a T-peak was high. This was handled by not allowing the decreased threshold unless 360ms had passed.

Heartbeat Segmentation

Before feature extraction was performed, each of the recordings was separated into single heartbeats. The motivation for heartbeat segmentation was to only provide the P-QRS-T complex in each of the signals; thus, more general feature extraction methods could be performed. The changing heart rate provides a different length for each of these complexes. Therefore, the length of each segmented heartbeat was decided from the RR intervals found from the QRS detection algorithm. The QT-Interval for each heartbeat corresponds to approximately 40% of the RR intervals [9]. However, additional margins were desirable to locate the T-wave ending point. Therefore, the final heartbeat segmentation for the RR intervals became 40% of the RR interval for the signal before the R-peak, and 60% of the RR interval for the signal after the R-peak. The rest of the signal was zero-padded to introduce constant signal length for each of the heartbeats, and providing that the R-peak had the same index for all the segmented heartbeats.

Points of Interest

Each of the heartbeats does contain three main points of interest. These points are the P-wave, QRS complex, and the T-wave. The location of these points is used to calculate the full feature list in Table 3.3. The location of peaks was done by prior knowledge of the ECG signal, and could, therefore, be accomplished as illustrated in Figure 3.11. At first, the fiducial feature extraction presented by Choi at el. [13] was tested. However, it was desirable to locate the onset and offset of the P, S and T waves, thus wave delineation as presented in Chapter 2.4.2 was added to the algorithm.

Key Points and Feature Computation

The feature extraction algorithm placed ten points for each of the heartbeats. These points were used to produce the feature vectors for each of the subjects. Figure 3.10 represents all the significant points found from QRS detection and wave delineation.
Combinations of indexes and amplitudes of each of the ten fiducial points were after that used to find and calculate the 16 different features found in table 3.3.
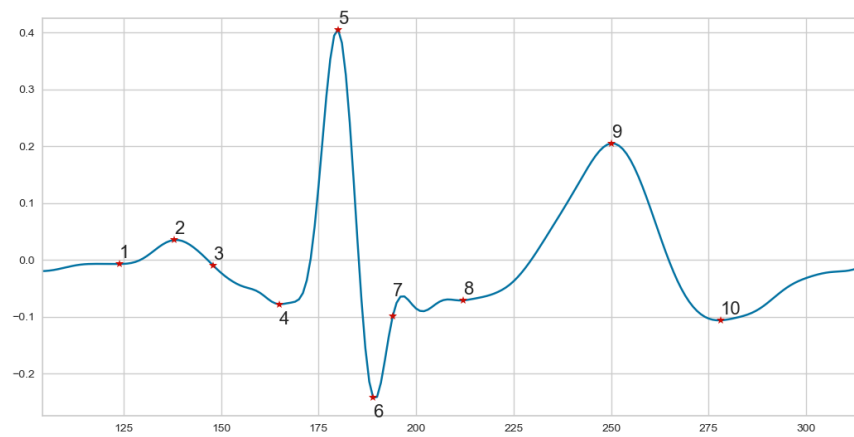
Figure 3.10: Illustration of the keypoins for a given P-QRS-T complex. (1) P_{on} (2), P_{Peak} (3), P_{off} (4), Q_{Pit} (5), R_{Peak} (6), S_{Pit} (7), S_{off} (8), T_{on} (9), T_{Peak} (10), T_{off}
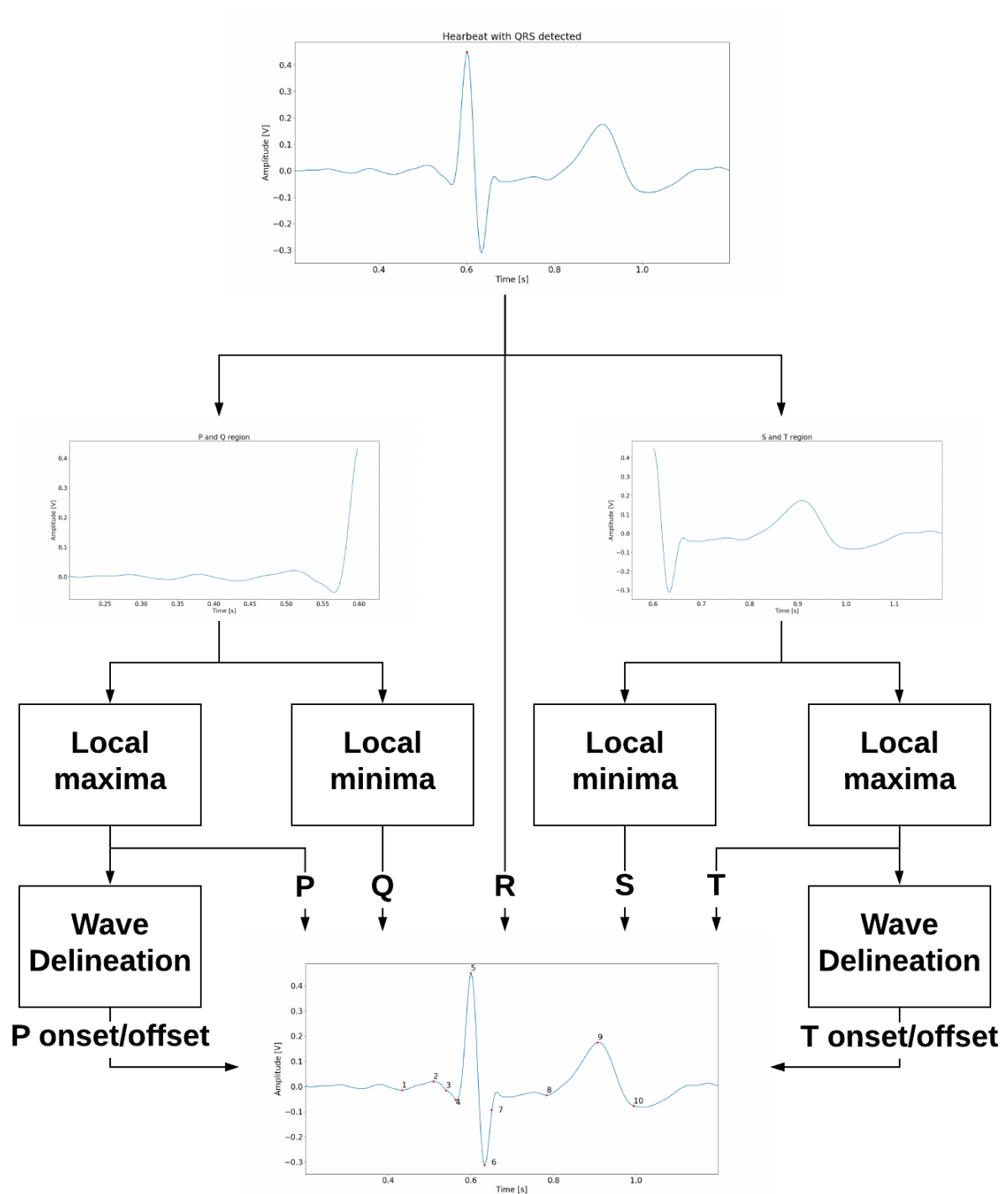
Figure 3.11: Simplified fiducial feature extraction algorithm.

3.2.3 Heart Rate Normalization

Extracting fiducial features from ECG signals makes the features notably conditioned on the heart rate. To make a generalized classifier that could tell each of the subjects apart, it was beneficial to normalize the extracted features for the estimated heart rate. Heart rate normalized features could enhance the robustness of ECG signals for the cases of increased heart rates.
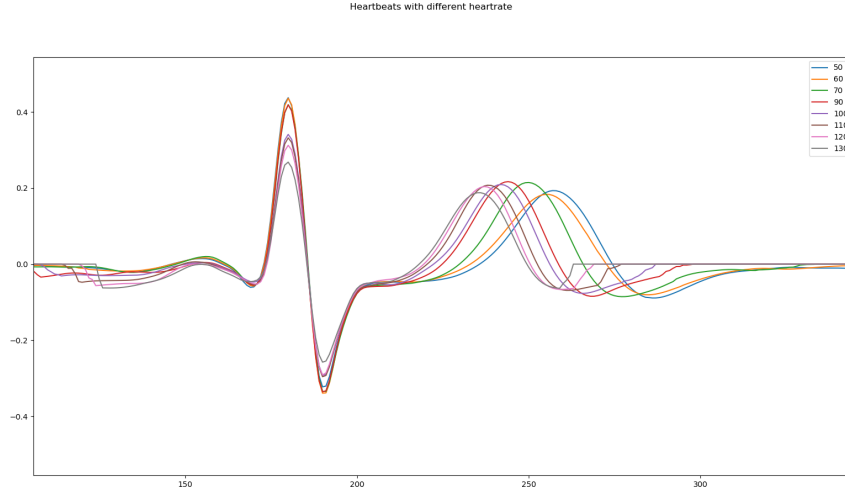
Figure 3.12: The changes on the different aspects of the heartbeat with elevated heart rate.

Some changes due to various heart rates have been illustrated in Figure 3.12. The amplitudes of the R- and S-peak decreases as the heart rate increases, and the QT interval got shifted towards the QRS complex as the heart rate increased.

Corrected QT Interval

The QT interval is the interval of ventricular de- and repolarization. As the heart beats faster, this period gets shorter. Therefore, the QT-interval changes with increased heart rate. Identification based on intervals that change with increased heart rate causes additional intraindividual variance for each of the subjects, which can contribute to a reduction in classification accuracy. However, a variety of methods for QT-interval correction has been proposed. Calculation of the Corrected QT interval (**QTc**) is a challenging task, as it has been found from regression analysis that the QR-RR ratio differs with age and sex. In practice, this means that all the methods for QTc corrections provide some error to the QTc interval. A 1992 study proposed "Framingham's method" to be more reliable than other methods [42]. The result of their study is given in Equation 3.4.

$$QT_C = QT + 0.154(1 - RR) \tag{3.4}$$

3.3 Machine Learning

ML provides powerful tools to find patterns in data, as explained in Section 2.3. In this project, the data consists of 20 different classes and 60 datasets. The desired result was to find a classifier that could provide sufficient accuracies for the 20 classes on all the datasets; however, only providing data from the R datasets to train the classification model. Machine learning in Python is accomplished by using the Scikit Learn library [43]. This library provides a powerful toolbox with various ML models, feature scaling and much more.

3.3.1 Classification Model

For this project, a handful of different classification models have been tested. The models have been listed in Table 3.4. The model consists of a mixture of different models, where most are based on models used by Choi et al. in their study with noisy ECG signals [13].

Most of the classification models are made for binary classification. However, using One-vs-Rest (**OVR**) classification methods provide one classifier for each of the 20 classes. This is accomplished by training one class at the time against each of the other classes. In Scikit Learn this was accomplished using the OnevsRest classifier [44].

In this project, it was decided that only the R dataset would be used to find and train the classification model that would be used for authentication. From the Scikit Learn library [43] the different classification models were trained and validated by performing 10-fold cross validation. Given the size of the dataset this was a reasonable way to estimate the accuracy for the different classification models on a relatively[5] small dataset. In addition, the hyperparameters for the classification models were found by performing a grid search, as explained in Section 2.3.5

Model	Name
LR	Logistic Regression
LDA	Linear Discriminate Analysis
KNN	K Nearest Neighbour
CART	Classification And Regression Trees
NB	Naive Bayes
SVM	Support Vector Machine
MLP	Multilayer perception
RF	Random Forest

Table 3.4: List of different Machine learning models tested for this project.

---

[5]ML requires a large scale of data, hence in a small dataset 10-fold cross validation is a simple, yet effective way to "expand" the dataset

3.3.2 Classifier Evaluation Method

Evaluation of the different classifiers to decide the best fit for an identification algorithm, some factors needed to be considered. The number of TP, TN, FP, and FN can be used to calculate additional evaluation metrics than ACC and was used for evaluation of the different classifiers.

$$TPR = \frac{TP}{TP + FN} = 1 - FNR \tag{3.5}$$

$$TNR = \frac{TN}{TN + FP} = 1 - FPR \tag{3.6}$$

$$FPR = \frac{FP}{TN + FP} = 1 - TNR \tag{3.7}$$

$$FNR = \frac{FN}{TP + FN} = 1 - TPR \tag{3.8}$$

In addition to the ACC, other factors of interest were the true positive rate (sensitivity) from Equation 3.5, and the true negative rate (specificity) from Equation 3.6 and their opposites, from Equation 3.7 and 3.8 describing the fall-out rate and miss rate respectively [32].
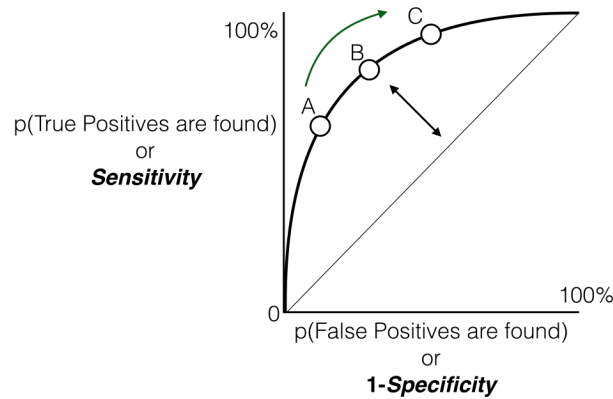
Figure 3.13: ROC-curve providing the relationship between sensitivity and 1-specificity. Point B illustrates the point where sensitivity is equal to 1-specificity[6][8].

The Receiver Operating Characteristics (**ROC**) curve provides insight in the decision boundary. and can distinguish operationally between discriminability and decision bias [30]. By calculation of Area Under the Curve (**AUC**) and Equal Error Rate (**EER**), where the latter is the sensitivity or 1-specificity in point B from figure 3.13. For this project, the ROC curves for each of the classification models must be averaged, as each class in the OVR classifier has its own classifier.

---

[6]CC BY-SA 3.0 https://creativecommons.org/licenses/by-sa/3.0/legalcode

For these evaluation metrics to be used for a multiclass classifier, the metrics for each of the individual classes had to be found and averaged. However, only calculating the mean of the already calculated values produces the macro-average, and it provides little information about class imbalance. Micro-average was used to take the class imbalance into account. [45].

$$Precision_i = \frac{M_{ii}}{\sum_j M_{ji}} \tag{3.9}$$

$$Recall_i = \frac{M_{ii}}{\sum_j M_{ij}} \tag{3.10}$$

$$F1_i = 2 \cdot \frac{Precision_i \cdot Recall_i}{Precision_i + Recall_i} \tag{3.11}$$

Using the combined measurements for the classifiers gave a full view of how the classifiers compared to each other. The most crucial metrics for identification is perhaps the F1-score, which is a combination of precision and recall. Calculation for precision, recall and F1-score for an arbitrary class i have been done according to Equations 3.9, 3.10 and 3.11. In addition to F1-score and ACC, identifying the prediction probability for the correct and wrong classifications provided insight into the certainty of the classifiers.

In addition to estimating the model for single heartbeats, it was of interest to find how additional heartbeats could provide better accuracies for the classifiers. This was accomplished by taking the sum of the predicted probabilities[7] for multiple heartbeats. After that, the OVR classifier is using the maximum of these probabilities to present the predicted class. This provided additional information that was lost in the ROC curves if either the sensitivity or specificity reached zero or one. Providing box-plots on the correct and wrong classifications provides estimates of mean and variance for the decision probabilities for the validation data.

3.4 Identification

The identification experiment for this project was performed by looking into each of the datasets individually. By analyzing the datasets individually, the accuracy for the different recording conditions could be calculated. The identification procedure was first tested on the single heartbeat case, where each of the datasets was tested before and after additional training data was supplemented. Identification for the 20 individuals in the datasets has been presented as multiclass confusion matrices. The confusion matrices are presented as

---

[7]The predicted probability is the probability each of the classification models in the OVR classifier predicts. The classifier with the highest prediction probability gets decided. Some classification models such as; RF and MLP also provides predicted probabilities.

illustrated in Figure 3.14a, where the perfect results are provided with all classifications placed on the diagonal.

Figure 3.14: (a) Illustration of the confusion matrix used to display how the different individuals got classified in relation to each other. (b) Confusion matrix illustrating the classification of known individuals and unknown individuals.

3.4.1 Biometric Authentication System

The final experiment was performed by the implementation of a simple authentication system. The authentication system was first tested with only data from the R dataset. By analyzing the predicted probabilities received from the classifier evaluation procedure, a threshold was set. In these experiments, 19 of the 20 individuals had been used to generate the classifier, and one had been excluded. This was done to evaluate how the unknown class would be handled in the authentication process. To illustrate this, a new confusion matrix was produced. This confusion matrix handled the 19 known classes as the positive state, and the unknown class as the negative state. The confusion matrix has been illustrated in Figure 3.14b. For this experiment, the unknown getting identified as a known subject is denoted as a type II error, and a known subject getting referred to as an unknown subject is referred to as type I error. This confusion matrix does not discriminate between the known subjects but sees the known subjects as the positive class, and the unknown subject as the negative class.

For the second part of this experiment, the authentication of all the collected datasets was used with the same conditions as earlier. However, a new decision threshold was found from the HRV validation test. Also, an increasing amount of the HRV dataset was used for training. The HRV training data was sorted by heart rate as illustrated in Figure 3.15, this was desired to keep the R dataset as the basis of the classifier and use the additional data to extend the "limits". This part was performed with 0%, 25% and 50% of the HRV dataset as training.
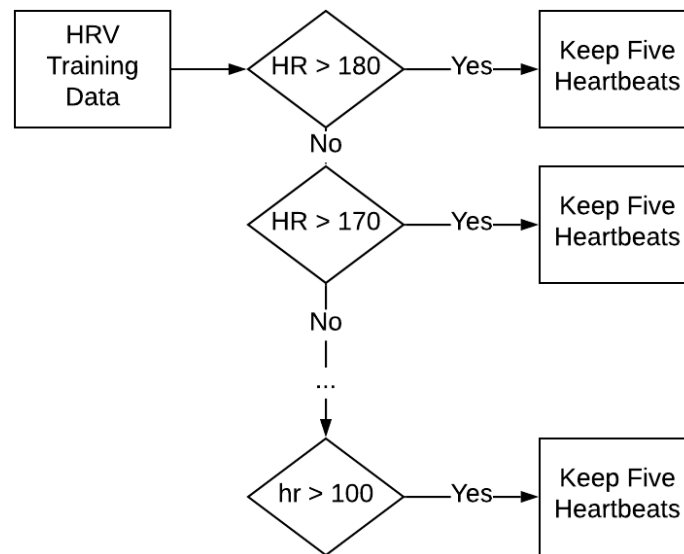
Figure 3.15: Extraction of five heartbeats for each tenth heart rate interval.
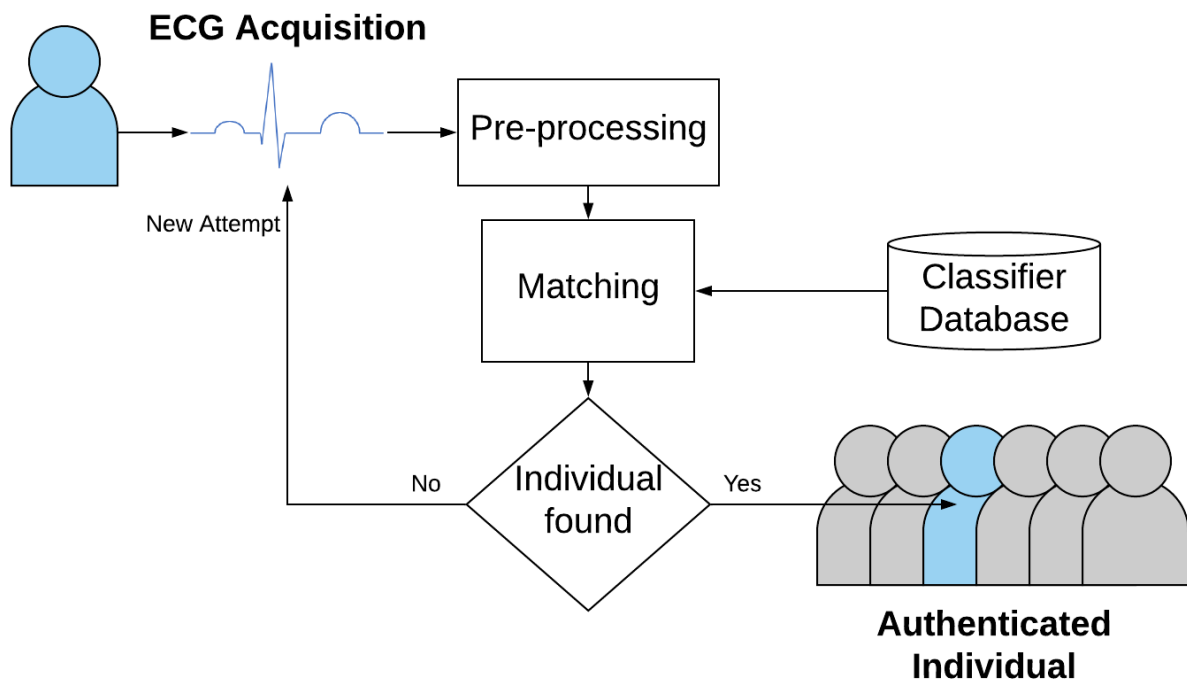
Figure 3.16: Flow chart of the implemented biometric authentication system.

The authentication system allowed each of the individuals to provide up to five heartbeats. In the case where the predicted probability did not exceed the set threshold had not been satisfied within this time, the individual was classified as unknown and had to possibility to try again. The authentication system has been illustrated in Figure 3.16.

Chapter 4

Results

This chapter manifests the results obtained from the method presented in Chapter 3. This chapter has been decomposed into four primary results. First, some results on the preprocessing method, have been presented; after that, some results from the model verification. The two last sections provide results obtained from single beat identification and finally, the results from the biometric authentication system.

4.1   Preprocessing

The preprocessing algorithm was developed to handle the different types of recorded ECG signals. The feature extraction method was strict and needed some requirements to add the given heartbeat to the list of features, and required that the denoising, QRS detection, and wave delineation provided the expected results. As the data has been collected during this project, the number of valid heartbeats was unknown. The ratio calculated is, therefore, the number of heartbeats located vs. the number of heartbeats kept after the strict feature extraction algorithm had proceeded. These results have been listed in table 4.1 and provide the number of heartbeats kept for each of the 20 individuals for all the three different datasets. The table is denoted such that the "true" value is the amount of QRS complexes found by the algorithm, and the kept rate is the percentage of heartbeats that was kept after the fiducial feature extraction algorithm.

From the table a relatively high rate of heartbeats was kept for most of the 20 subjects in the dataset. With a minimum at 90% for subject 20 in the R dataset, 90% for subject 18 in the HRV dataset, and 86% for subject 10 in the M dataset.

- 99.4% for the R dataset.

- 98,4% for the HRV dataset.

- 96,95% for the M dataset.

Subject	R true	R kept[%]	HRV true	HRV kept [%]	M true	M kept [%]
01	69	100	246	100	25	100
02	47	100	120	98	13	100
03	67	100	227	100	29	93
04	81	100	261	96	29	100
05	57	100	189	99	29	100
06	56	100	195	100	27	96
07	53	100	160	100	21	100
08	72	100	175	100	23	100
09	55	100	170	100	11	90
10	55	100	235	96	22	**86**
11	71	100	226	94	27	100
12	57	100	180	100	16	100
13	61	98	156	100	17	94
14	70	100	225	96	38	100
15	69	100	149	100	16	93
16	65	100	198	100	30	100
17	76	100	138	99	30	100
18	96	100	251	**90**	33	87
19	58	100	185	100	19	100
20	60	**90**	208	100	25	100

Table 4.1: Relationship between found and retained heartbeats for the three datasets for each of the 20 subjects in this project.

From table 4.1 it has been shown that the different subjects provided different amounts of heartbeats for each of the tests. This is a result of different heart rates and possibly poor performance of the QRS detection algorithm.

4.2 Model verification

Verification of which classifiers had the potential to provide accurate biometric identification the classifiers were tested and validated with a 10-fold cross validation method. The parameters for the classifiers were found by the experiment in Appendix A.1. The classifiers were provided with parameter grids and were used to find which of the parameters provided the highest accuracy. The results from cross validation have been provided in Figure 4.1 and Table 4.2.

In addition to accuracy, evaluation metrics such as; EER, AUC, and F1-score was calculated, and the micro-average ROC curves for the classifiers plotted. For this case, the training provided further splitting to generate 70% training and 10% validation data.

Figure 4.2 shows the micro-average ROC curves for the different classifiers for the R dataset. The performance metrics are listed Table 4.3. The highest performance for single beat

Model	Mean Accuracy	Standard Deviation
LR	0.897	0.044
LDA	0.859	0.061
KNN	0.893	0.031
CART	0.711	0.056
NB	0.935	0.029
SVM	0.888	0.043
MLP	**0.950**	**0.019**
RF	0.933	0.032

Table 4.2: Mean accuracy and standard deviation for the 10-fold cross validation of the different classification models listed in Table 3.4.
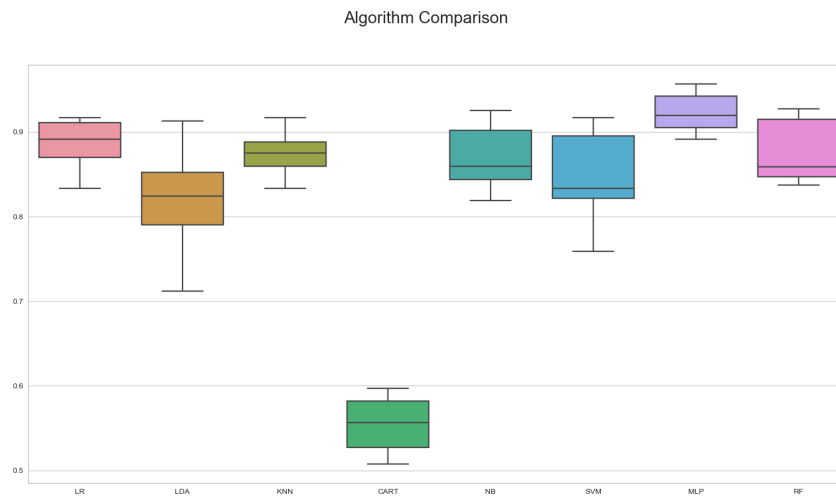

Figure 4.1: Box plot illustrating how the algorithms compare to each other after hyper parameter tuning with 10-fold cross validation

	LR	LDA	KNN	CART	NB	SVM	MLP	RF
ACC	0.555	0.449	0.823	0.381	0.660	0.600	0.872	0.551
AUC	0.978	0.969	0.972	0.785	0.971	0.979	0.992	0.988
EER	0.074	0.083	0.033	0.039	0.090	0.083	0.039	0.060
F1 score	0.718	0.623	0.853	0.538	0.795	0.740	0.910	0.707

Table 4.3: Evaluation metrics for the different classifiers where only R data was used for validation. The ROC curve from this test has been displayed in Figure 4.2

classification was with the MLP classifier.

Figure 4.3 shows the micro-average ROC curves for the different classifiers for the R and HRV datasets. The performance metrics are listed Table 4.4. The highest performance for the single beat classification was with the MLP classifier. Also, the micro-average ROC curve for five heartbeats was found. This has been shown in Figure 4.4, and the evaluation metrics for the classifiers have been listed in Table 4.5.

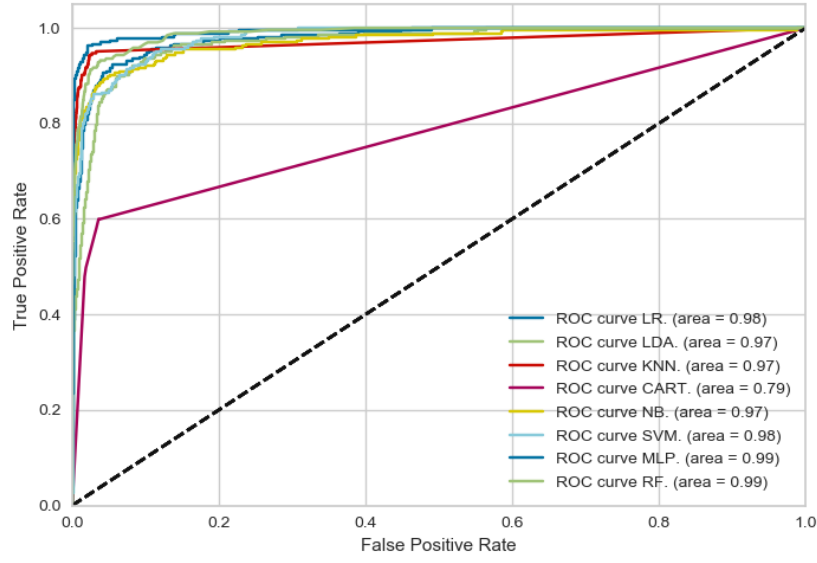Micro-Average ROC curve for the different classifiers

Figure 4.2: Micro-average ROC curves for the different classifiers. Training and validation data from the R dataset on single heartbeats.

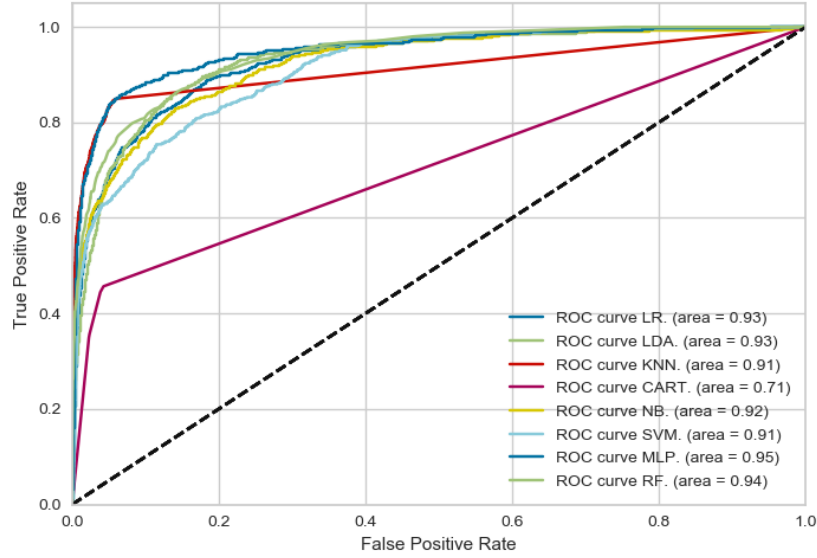Micro-Average ROC curve for the different classifiers

Figure 4.3: Micro-average ROC curves for the different classifiers. Training from the R dataset, and validation from the HRV dataset for single heartbeats

.

The presented evaluation metrics did not provide enough certainty for which of the classifiers that would provide the best results for both the R and HRV datasets. To decide, all the classification models except for the CART classifier was tested for one and five heartbeats. By storing the predicted probabilities for both correct classifications and false classifications it was possible to see how big the difference in predicted probabilities was for both correct and wrong classifications. The full experiment has been placed in Appendix A.2. The results

	LR	LDA	KNN	CART	NB	SVM	MLP	RF
ACC	0.393	0.341	0.622	0.253	0.402	0.414	0.607	0.335
AUC	0.929	0.930	0.913	0.709	0.923	0.912	0.951	0.941
EER	0.152	0.141	0.060	0.043	0.159	0.186	0.111	0.142
F1 score	0.546	0.504	0.684	0.395	0.575	0.564	0.685	0.488

Table 4.4: Evaluation metrics for the different classifiers where both R and HRV data was used for validation. The ROC curve from this test has been displayed in Figure 4.3.
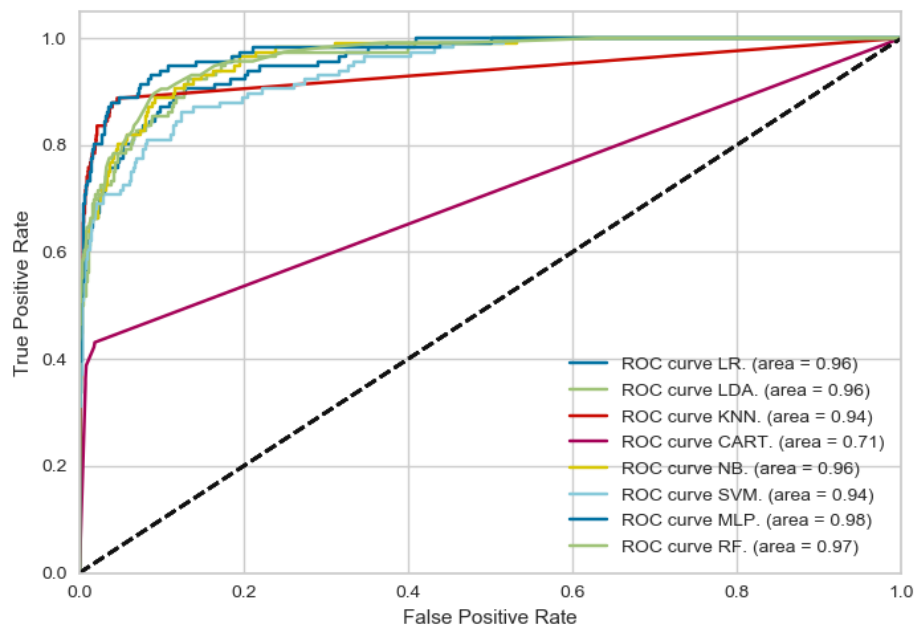

Figure 4.4: Micro-average ROC curves for the different classifiers. Both the training and validation data consists of the HRV datasets with taking the majority voting after five heartbeats.

	LR	LDA	KNN	CART	NB	SVM	MLP	RF
ACC	0.431	0.431	0.724	0.353	0.517	0.474	0.672	0.371
AUC	0.957	0.960	0.936	0.708	0.964	0.942	0.975	0.966
EER	0.114	0.117	0.050	0.019	0.114	0.138	0.084	0.099
F1 score	0.590	0.602	0.771	0.503	0.697	0.633	0.759	0.541

Table 4.5: Model evaluation with HRV and R evaluation and training data. The ROC curve for this test has been displayed in Figure 4.4.

from this test showed that the Random Forest classifier produces the best limits between predicted probabilities for correct and wrong classifications. Therefore, the RF classifier was used for the next experiments.

4.3   Single Beat Identification

This section provides the result from the identification experiment proposed in Section 3.4. The test was performed on single heartbeats, and the experiment included two different training sets. The first test was performed where only the test data from the R dataset was used for training. The second test consisted of both the R dataset, and 50% of the HRV dataset for training. The goal of this experiment was to establish how the use of additional training data based on ECG signals with increased heart rate and increased noise could potentially improve accuracy for the single heartbeat identification.
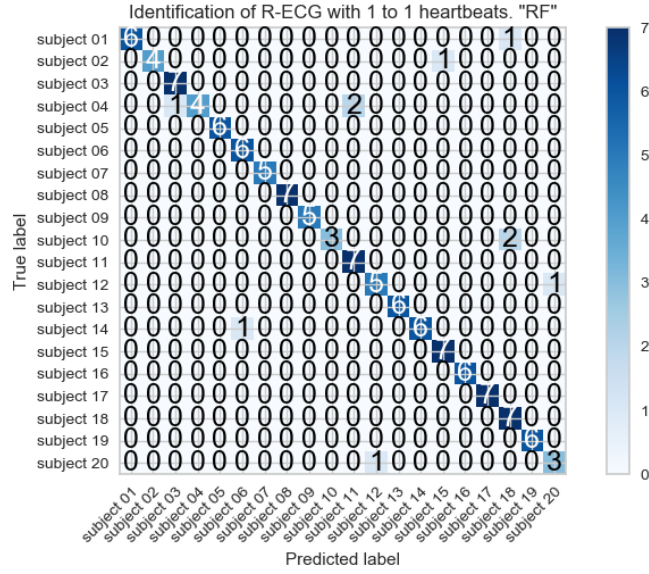
Figure 4.5: Confusion matrix for single beat identification for the R dataset.
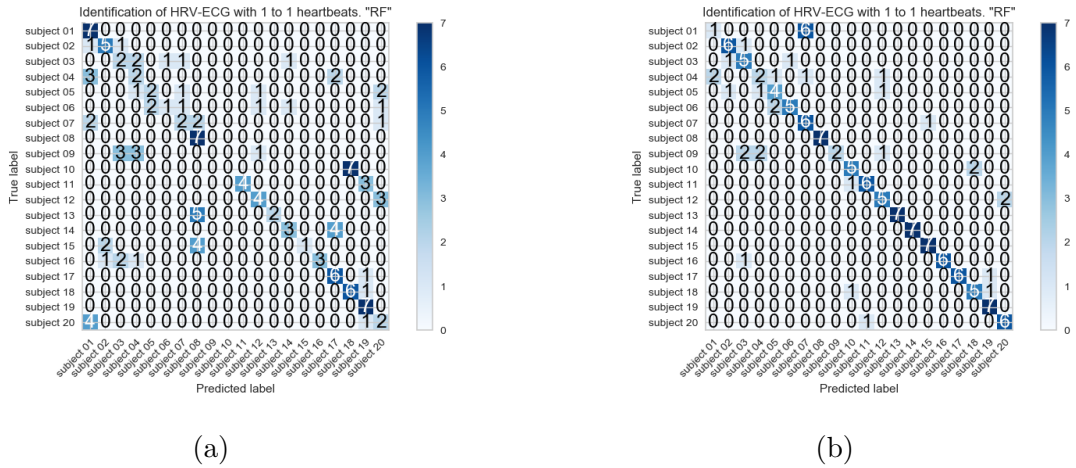
Figure 4.6: (a) Confusion matrix for HRV single beat identification without HRV training data. (b) Confusion matrix for HRV single beat identification with 50% HRV training data.
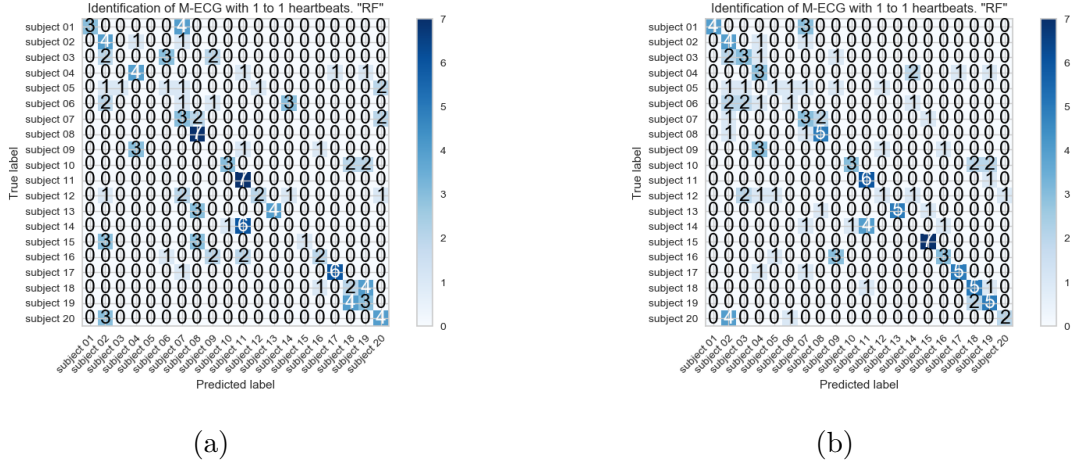
(a) (b)

Figure 4.7: (a) Confusion matrix for M single beat identification without HRV training data. (b) Confusion matrix for M single beat identification with 50% HRV training data.

Figures 4.5, 4.6 and 4.7 show how the accuracy for single beat identification increases by including additional HRV data for training. Percentages of accuracy before and after additional training data has been provided in Table 4.6

	R	HRV test 1	HRV test 2	M test 1	M test 2
Accuracy [%]	91.8	47.1	82.9	40.1	45.9
Identified	20	18	20	15	18

Table 4.6: Accuracy and number of subjects identified from single beat identification. Test 1 consisted of training on only the R dataset and test 2 included additional HRV training data.

By additional training data from the HRV data set the accuracy and number of correctly identified subjects increased drastically for the HRV dataset. The M dataset increased as well. However, did not have the same effect as for HRV.

4.4 Biometric Authentication

The biometric authentication has, as explained in Section 3.4.1, been divided into two separate authentication experiments. The first experiment only tests on the R dataset, and the second experiment uses the R, HRV and M datasets. The thresholdsh have been found from Appendix A.2.

4.4.1 Authentication Experiment 1

This experiment evaluated how the recorded R ECG signals recorded from the set of individuals was handled by the authentication system. From Appendix A.2 in Figure A.15 it was shown that a threshold on 0.7 should, based on the HRV validation data, not give misclassifications.

| (a) | (b) |
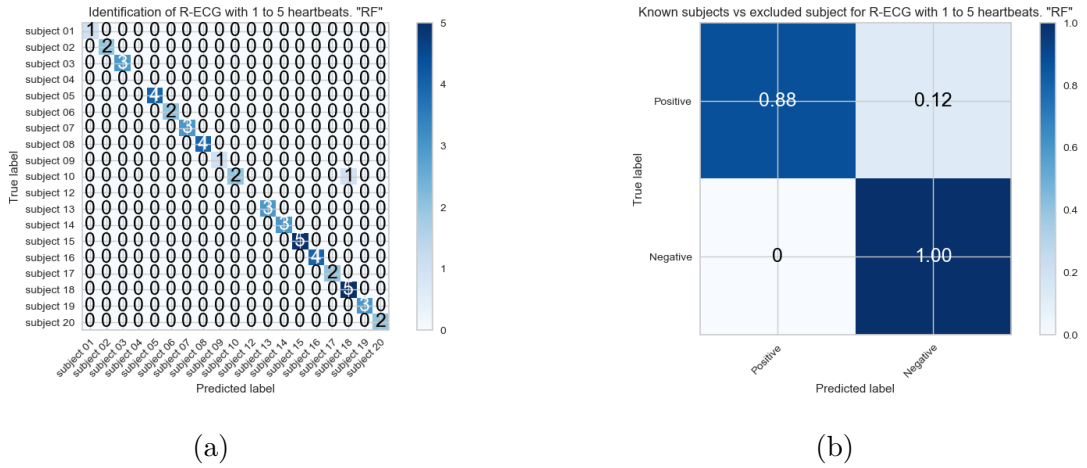
Figure 4.8: (a) Confusion matrix for the identification test with the R-ECG dataset (b) Confusion matrix for the identification process illustrating the amount of test being classified as unknown subject, and the unknown subject classified as a known subject.

Figure 4.8a displays the ratio of subjects classified, from the figure it is shown that 17/19 individuals got classified as themselves, one known individual got classified as another known individual. However, the unknown individual did not get classified as a known subject, and 12% of the tests on known subject got classified as an unknown subject. These numbers are shown in Figure 4.8b.

4.4.2 Authentication Experiment 2

This experiment evaluated how the recorded R, HRV and M ECG signals perform with a varying amount of HRV training data. In this experiment, the threshold value was found from Appendix A.2. Figure A.16 gave reason to believe that a threshold at 0.75 should avoid misclassifications. The results have been presented in confusion matrices.

0% HRV-ECG Training Data
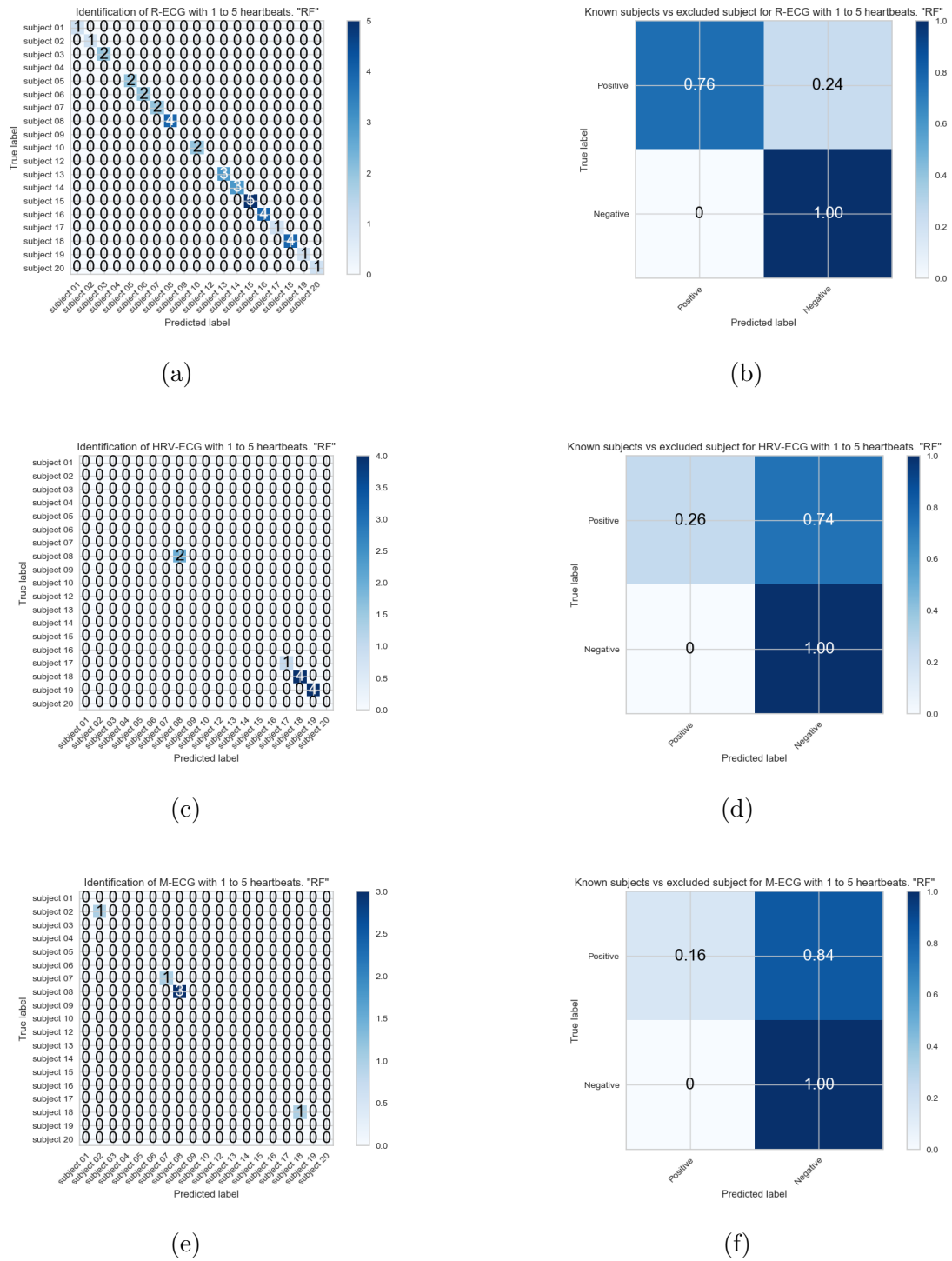

(a)

(b)

(c)

(d)

(e)

(f)

Figure 4.9: Authentication with 0% HRV. (a)-(b) consists of R, (c)-(d) consists of HRV, and (e)-(f) consists of the M dataset.

Figure 4.9 shows how the authentication process performed without additional training data.
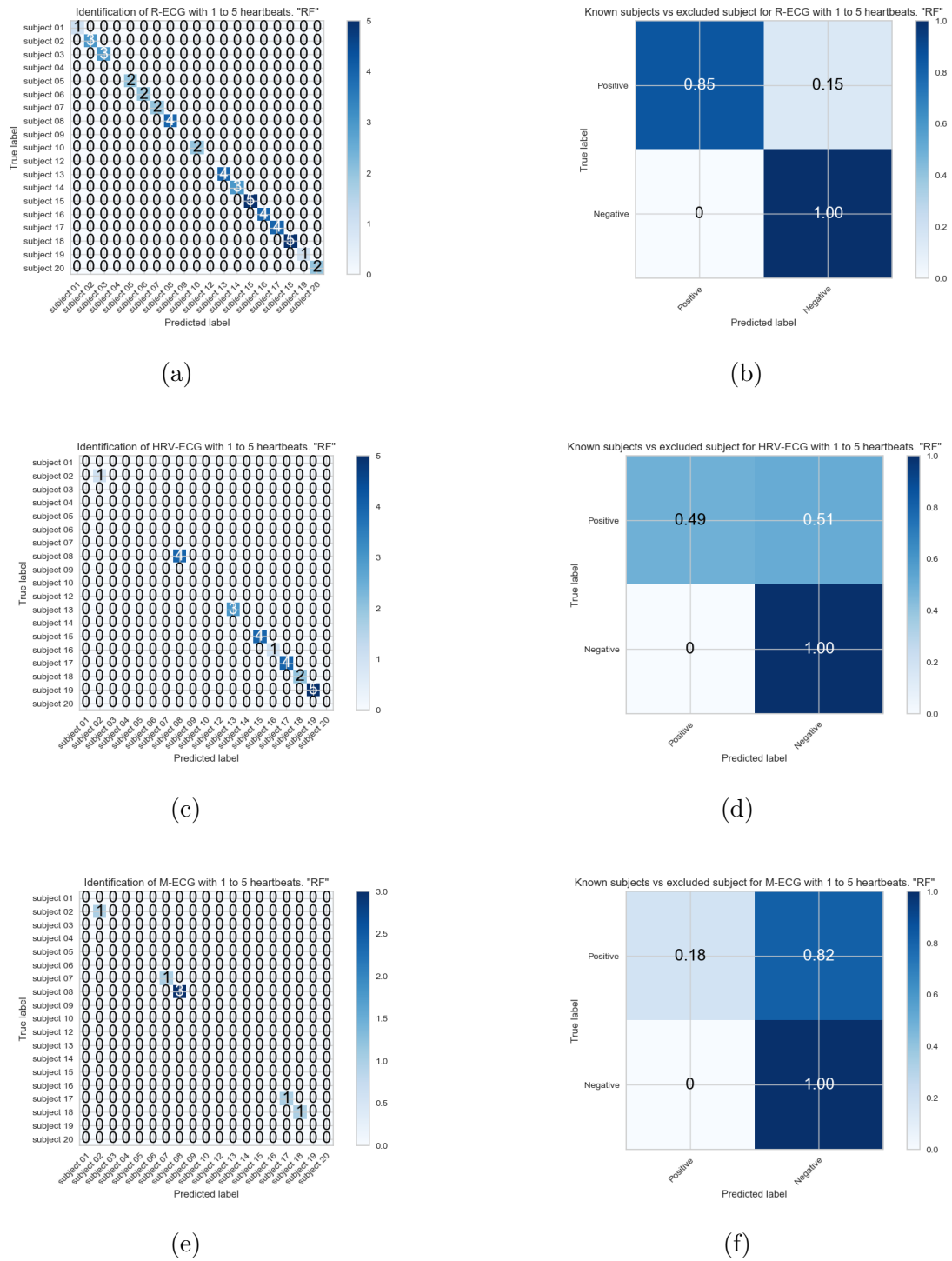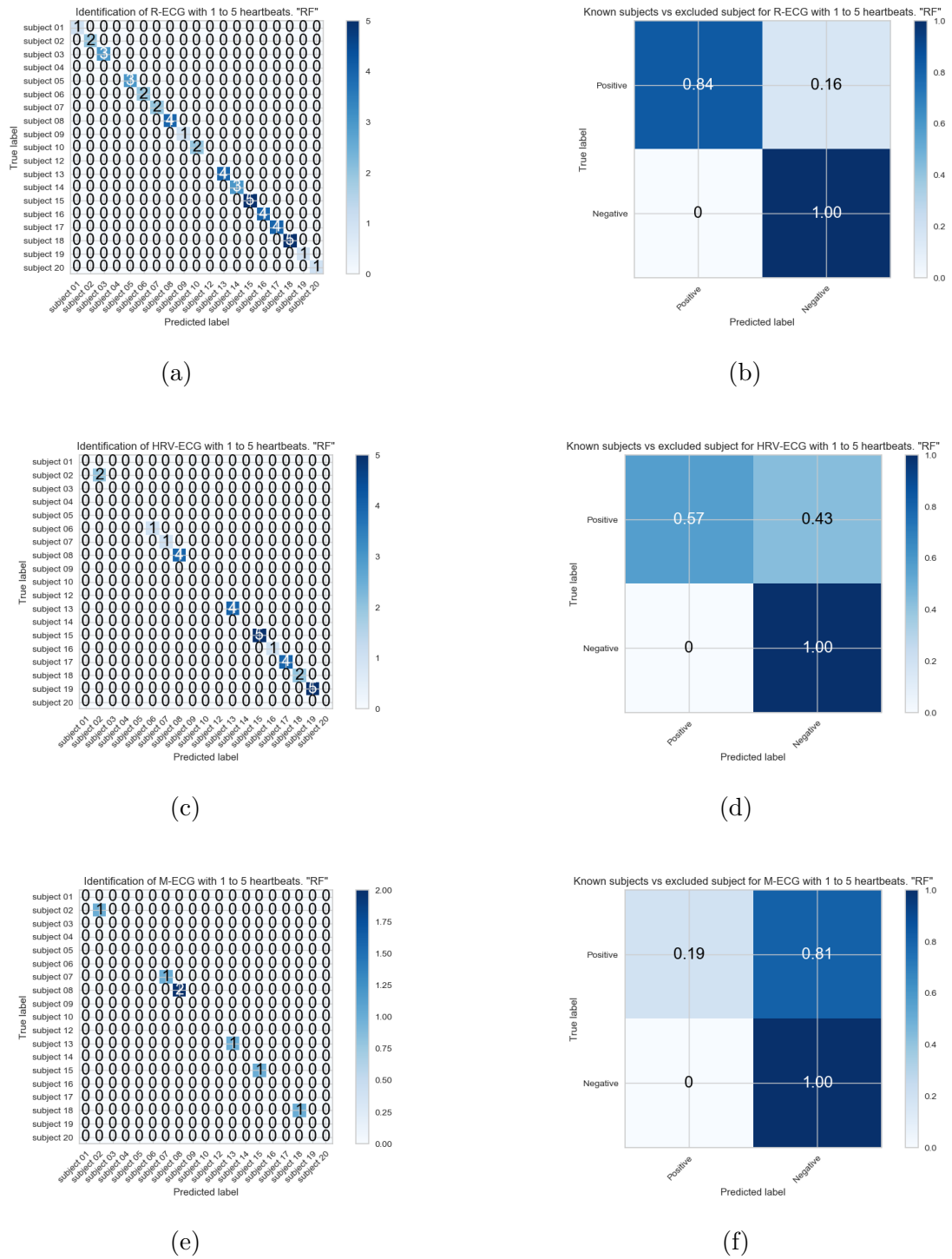
25% HRV-ECG Training Data

(a)

(b)

(c)

(d)

(e)

(f)

Figure 4.10: Authentication with 25% HRV. (a)-(b) consists of R, (c)-(d) consists of HRV, and (e)-(f) consists of the M dataset.

Figure 4.10 shows how the authentication process performed with 25% additional training data from the HRV dataset.

50% HRV-ECG Training Data

(a)

(b)

(c)

(d)

(e)

(f)

Figure 4.11: Authentication with 50% HRV. (a)-(b) consists of R, (c)-(d) consists of HRV, and (e)-(f) consists of the M dataset.

Figure 4.11 shows how the authentication process performed with 50% additional training data from the HRV dataset.

Summary of Authentication Experiment 2

Dataset	Correct	Wrong	Type I Error[%]	Type II Error[%]
R	16/19	0	24	0
HRV	4/19	0	74	0
M	4/19	0	84	0

Table 4.7: Results from the 0% HRV test.

Dataset	Correct	Wrong	Type I Error[%]	Type II Error[%]
R	16/19	0	15	0
HRV	8/19	0	51	0
M	5/19	0	82	0

Table 4.8: Results from the 25% HRV test.

Dataset	Correct	Wrong	Type I Error[%]	Type II Error[%]
R	17/19	0	16	0
HRV	10/19	0	43	0
M	6/19	0	81	0

Table 4.9: Results from the 50% HRV test.

From Tables 4.7, 4.8 and 4.9 it has been shown that an increasing amount of training data for the HRV provides a higher acceptance rate for the HRV dataset. It also increased the accuracy for the authentication of the R dataset. The increase of HRV data did slightly provide better authentication results for the M dataset as well. Not including HRV data provided a low number of authenticated subjects for the HRV and M datasets. However, no type II error was found in any of the experiments.

Chapter 5

Discussion

This chapter discuss the results and the method. Further some threats to validity, future work and some concluding remarks.

5.1 Analysis of the Results

The results obtained for this thesis has been provided over a handful of experiments. The results obtained by the classification model evaluation indicated that the Multilayer Perceptron classifier would give the desired results "hinted at" in the introduction. However, by validation over one and five heartbeats, the decision probabilities proved to be uncertain for validation of the HRV dataset. Thus, the Multilayer Perceptron classifier did not provide satisfactory results for a biometric authentication system. Further analysis of the predicted probability for the different classifiers, it was found that the Random Forest increased the certainty distinguished between the different individuals for the R and HRV dataset.

Identification from single heartbeats provided compelling results, revealing how the Random Forest classifier provided reasonable identification accuracies for both the HRV and M datasets. Without using any of these datasets for training, accuracies of 47.1% and 40.1% respectively were obtained. After additional training data had been appended to the training set for the classifier, accuracies of 82,9% and 45.9% were achieved for the HRV and M datasets respectively. Further analysis of the prediction probabilities achieved for correct and wrong classification, unfortunately, revealed that, even though the individuals got correctly classified, the predicted probabilities that was obtained from the classification model was rather low.

The biometric authentication system was tested under three different conditions where different amounts of the HRV datasets were provided as additional training data, where 0%, 25%, and 50% were tested. First, by only testing on the R dataset with a threshold of 0.7 provided a correct classification of 16 of the subjects, and avoided classifying a subject that

had been excluded from the dataset as another subject, thus resulting in 0% type II error. Further testing on the HRV dataset found that the authentication, with strict thresholds, had 0% type II error. However, without additional training data, the system performed poorly, only correctly classifying four subjects. Additional training provided eight and ten correctly classified subjects for the 25% and 50% cases respectively. The M dataset provided four, five, and six correct classifications for the cases of 0%, 25% and, 50% respectively. However, like for the R and HRV dataset no type II error was found.

5.2 Reflections on the Method

5.2.1 Pre-processing

The pre-processing methods used for this thesis have been adaptions of different adaptive methods used for the pre-processing of ECG signals under noisy conditions. The different filters and feature extraction methods have not been implemented for real-time systems, however, it was used for offline analysis. The adaptive filtering methods for baseline wander and muscle noise provided reasonable results. The QRS detection algorithm implemented did not detect all the peaks in some of the ECG signals acquired. The QRS algorithm based its decision on the slopes of the ECG signal. Although, for some of the recorded signals in the database, the slopes did not produce high enough amplitudes in the moving average signal. This resulted in some peaks not being identified and made the estimated heart rate for the signals less accurate. This could potentially have provided the wrong scaling of the QTc intervals.

5.2.2 Classification and Evaluation Methods

The classification methods tested for this thesis was used as OVR classifiers. The OVR classification provides one trained classifier for each of the classes provided. The problem with using OVR is that each of the classifiers is trained against each other; this means that this method requires at least two classes to produce one classifier. This type of classification model would perhaps not be desirable for authentication applications on wearable devices where there is most likely only one user. A more suitable solution would preferably be independent of other classes.

For evaluation, the provided metrics did not provide the anticipated differences; therefore, an additional test to find the difference in prediction probabilities were implemented. This could potentially have been solved by introducing additional evaluation metrics. Evaluation metrics that include threshold in their evaluation could have provided the results from Appendix A.2 on a more professional standard.

5.3   Threats to Validity

This thesis was originally aimed at Apple Watch series 4. However, Apple did not want to provide API[1] for ECG signal extraction. Different equipment with a similar sample rate and quantification was found, however, due to price and lack of accessible APIs for data extraction, his project had to proceed on equipment that provided a lower sampling rate and quantification. The recorder provided lower precision and included more noise. This provided smaller differences in both amplitudes and intervals, and may, therefore, have resulted in smaller interindividual variance between the subjects.

For data acquisition, a set of 20 individuals have been used. The volunteers were mostly male and were around the same age. The dataset did, therefore not provide the diversity desired. A generalized approach can therefore not confidently propose, as the data does not represent a broad enough population. This could be solved by additional data collection in a broader spread of individual's age, sex, etc.

5.4   Conclusion

This thesis examined the possibility of implementation of a robust ECG biometric authentication system for wearable/portable devices. The motivation was to find a model that could provide reasonable results for ECG signals recorded with resting heart rate, increased heart rate after physical activity, and noisy ECG signals recorded during motion. By performing adaptive denoising and fiducial feature extraction, various classification models were trained and evaluated to find a classification model that could be used for a biometric authentication system. Finally the Random Forest classification model was found to provide the best results, and was therefore used for the experiments executed in this project.

Results in this thesis indicated that an ECG based biometric authentication procedure provided that 17/19 subjects got authenticated, corresponding to **89.5%** accepted individuals. This was reduced to **52.6%** (10/19) and **31.6%** (6/19) for ECG signals acquired after physical activity and during motion, respectively. By only providing ECG signals recorded with a resting heart rate to train the biometric authentication system, the acceptance rate was reduced to 84.2% for resting heart rate, and 21.1% for ECG signals after physical activity and while in motion. The last subject who had been excluded from the system did not get accepted for any of the three recordings provided. Additionally, no misclassification between the known subjects occurred. The test on single heartbeat identification provided accuracies of 91.8%, 82.9% and 45.9% for the different recordings. This indicates that a higher acceptance rate could be obtainable.

---

[1]Application Programming Interface.

Overall the results obtained from this project indicate that biometric authentication on wearable devices potentially could provide the desired accuracy for a biometric authentication system. However, it would require new and more robust methods in order to be generalized for ECG signals obtained under different conditions.

5.5   Future Work

Many different methods for classification and feature extraction have been left for future work. The lack of API access for ECG signals collected on the Apple Watch resulted in alternative equipment. The use of Apple Watch for data collection would have been of significant interest as it provides high-quality data. For the feature extraction method presented in Chapter 3 it should be considered using features from the frequency domain rather than those of the time domain, as used for this thesis. Also, alternative methods for QRS detection and filtering should be tested, as it may give more accurate QRS detection. The amount of noise found in the different recording of the ECG signals provided more noise than what could be filtered, which made accurate fiducial feature extraction difficult.

For the classification method, it could be considered using outlier detection as a classification method that would be independent of the other classes. The generalization of the outlier detection algorithm for the HRV and M datasets could potentially be problematic, thus reversing the heart rate normalization algorithm to simulate heartbeats with increased heart rate could be a potential solution. Additionally, using multiple methods for increased decision certainty may boost the acceptance rate.

Bibliography

[1] Wapcaplet, "Diagram of the human heart, created by Wapcaplet in Sodipodi. Cropped by Yaddah to remove white space (this cropping is not the same as Wapcaplet's original crop)." Jun. 2006, accessed: 29.4.2019. [Online]. Available: https://commons.wikimedia.org/wiki/File:Diagram_of_the_human_heart_(cropped).svg

[2] E.-P. s. E.-P. s. S. s. C. b. Agateller, "Schematic diagram of normal sinus rhythm for a human heart as seen on ECG, two periods forming a RR-interval." Sep. 2009, accessed: 7.5.2019. [Online]. Available: https://commons.wikimedia.org/wiki/File:ECG-RRinterval.svg

[3] Npatchett, "English: Spatial orientation of EKG leads," Mar. 2015, accessed: 4.6.2019. [Online]. Available: https://commons.wikimedia.org/wiki/File:EKG_leads.png

[4] aconcagua, "Deutsch: Apple Watch, 40mm AluminiumEnglish: Apple Watch, 40mm AluminumFrançais: Apple Watch, 40mm Aluminium," Sep. 2018, accessed: 28.3.2019. [Online]. Available: https://commons.wikimedia.org/wiki/File:Apple_Watch_Series_4_40mm_space_gray_Aluminum.jpg

[5] A. Lumini and L. Nanni, "Overview of the combination of biometric matchers," *Information Fusion*, vol. 33, pp. 71–85, Jan. 2017, accessed: 17.1.2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1566253516300446

[6] S. s. C. b. Agateller, "Schematic diagram of normal sinus rhythm for a human heart as seen on ECG (with Czech labels)." Sep. 2009, accessed 10.3.2019. [Online]. Available: https://commons.wikimedia.org/wiki/File:ECG-PQRST%2Bpopis.svg

[7] L. Sathyapriya, L. Murali, and T. Manigandan, "Analysis and detection R-peak detection using Modified Pan-Tompkins algorithm," in *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*, May 2014, pp. 483–487.

[8] Masato8686819, "English: ROC curve image," Nov. 2013, accessed: 4.5.2019. [Online]. Available: https://commons.wikimedia.org/wiki/File:ROC_curve.svg

[9] M. S. Thaler, *The only EKG book you'll ever need*, 7th ed. Wolters Kluwer/Lippincott Williams & Wilkins, 2012.

[10] L. Biel, O. Pettersson, L. Philipson, and P. Wide, "ECG analysis: a new approach in human identification," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 3, pp. 808–812, Jun. 2001.

[11] G. B. Moody and R. G. Mark, "MIT-BIH Arrhythmia Database," 1992, type: dataset. [Online]. Available: https://physionet.org/physiobank/database/mitdb/

[12] T. Lugovaya, "The ECG-ID Database," 2011, type: dataset. [Online]. Available: https://physionet.org/physiobank/database/ecgiddb/

[13] H. Choi, B. Lee, and S. Yoon, "Biometric Authentication Using Noisy Electrocardiograms Acquired by Mobile Sensors," *IEEE Access*, vol. 4, pp. 1266–1273, 2016.

[14] R. Salloum and C.-J. Kuo, "ECG-based biometrics using recurrent neural networks," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2017, pp. 2062–2066.

[15] T. S. Lugovaya, "Biometric Human Identification based on ECG," Jun. 2005, accessed: 17.1.2019. [Online]. Available: https://www.physionet.org/pn3/ecgiddb/biometric.shtml

[16] S. Pathoumvanh, S. Airphaiboon, and K. Hamamoto, "Robustness study of ECG biometric identification in heart rate variability conditions," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 9, no. 3, pp. 294–301, 2014. [Online]. Available: http://onlinelibrary.wiley.com/doi/abs/10.1002/tee.21970

[17] J. R. Pinto, J. S. Cardoso, and A. Lourenço, "Evolution, Current Challenges, and Future Possibilities in ECG Biometrics," *IEEE Access*, vol. 6, pp. 34 746–34 776, 2018.

[18] L. Sörnmo and P. Laguna, *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. Elsevier Academic Press, 2005.

[19] M. AlGhatrif and J. Lindsay, "A brief review: history to understand fundamentals of electrocardiography," *Journal of Community Hospital Internal Medicine Perspectives*, vol. 2, no. 1, Apr. 2012, accessed: 5.6.2019. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3714093/

[20] L. Potter, "Understanding an ECG," Mar. 2011, accessed: 14.6.2019. [Online]. Available: https://geekymedics.com/understanding-an-ecg/

[21] Apple, "Using Apple Watch for Arrhythmia Detection December 2018," 2018.

[22] C. Watford, "Understanding ECG Filtering | EMS 12 Lead," Mar. 2014, accessed: 9.4.2019. [Online]. Available: http://ems12lead.com/2014/03/10/understanding-ecg-filtering/

[23] A. Goode, "Biometric Identification or Biometric Authentication?" Jul. 2018. [Online]. Available: https://www.veridiumid.com/blog/biometric-identification-and-biometric-authentication/

[24] "Biometric authentication (What is biometrics?) | 2018 Review." [Online]. Available: https://www.gemalto.com/govt/inspired/biometrics

[25] Z. Rui and Z. Yan, "A Survey on Biometric Authentication: Toward Secure and Privacy-Preserving Identification," *IEEE Access*, vol. 7, pp. 5994–6009, 2019.

[26] M. Korolov, "What is biometrics? And why collecting biometric data is risky," Feb. 2019, accessed: 21.5.2019. [Online]. Available: https://www.csoonline.com/article/3339565/what-is-biometrics-and-why-collecting-biometric-data-is-risky.html

[27] M. G. Adams-Hamoda, M. A. Caldwell, and N. A. Stotts, "Factors to consider when analyzing 12-lead electrocardiograms for evidence of acute myocardial ischemia," *AMERICAN JOURNAL OF CRITICAL CARE*, vol. 12, no. 1, p. 9, 2003, accessed: 28.2.2019. [Online]. Available: http://ajcc.aacnjournals.org/content/12/1/9.full.pdf

[28] B. J. A. Schijvenaars, G. van Herpen, and J. A. Kors, "Intraindividual variability in electrocardiograms," *Journal of Electrocardiology*, vol. 41, no. 3, pp. 190–196, May 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0022073608000447

[29] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms.* Cambridge: Cambridge University Press, 2014. [Online]. Available: http://ebooks.cambridge.org/ref/id/CBO9781107298019

[30] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, second edition ed. Canada: Wiley-Interscience, 2001.

[31] Wikipedia, "Feature scaling," Jun. 2019, accessed: 4.6.2019. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Feature_scaling&oldid=899790585

[32] ——, "Sensitivity and specificity," May 2019, accessed: 13.6.2019. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Sensitivity_and_specificity&oldid=895891646

[33] J. Bergstra and Y. Bengio, "Random Search for Hyper-Parameter Optimization," p. 25, Feb. 2012.

[34] C. Carreiras, A. P. Alves, A. Lourenco, F. Caneto, and H. Silva, "Biosignal Processing in Python." Apr. 2019, accessed: 28.4.2019. [Online]. Available: https://github.com/PIA-Group/BioSPPy

[35] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985. [Online]. Available: http://ieeexplore.ieee.org/document/4122029/

[36] "Alive Bluetooth Heart & Activity Monitor," accessed: 28.3.2019. [Online]. Available: https://www.alivetec.com/pages/alive-bluetooth-heart-activity-monitor

[37] "Key Changes with the General Data Protection Regulation – EUGDPR," accessed: 24.4.2019. [Online]. Available: https://eugdpr.org/the-regulation/

[38] D. Ross, "Processing biometric data? Be careful, under the GDPR." [Online]. Available: https://iapp.org/news/a/processing-biometric-data-be-careful-under-the-gdpr/

[39] L. Sörnmo, "Time-varying digital filtering of ECG baseline wander," *Medical and Biological Engineering and Computing*, vol. 31, no. 5, p. 503, Sep. 1993. [Online]. Available: https://doi.org/10.1007/BF02441986

[40] A. Hashemi, M. Rahimpour, and M. R. Merati, "Dynamic Gaussian filter for muscle noise reduction in ECG signal," in *2015 23rd Iranian Conference on Electrical Engineering*, May 2015, pp. 120–124.

[41] H. Sedghamiz, "Matlab Implementation of Pan Tompkins ECG QRS Detector." p. 3, Mar. 2014, acessed: 28.05.2019. [Online]. Available: https://www.researchgate.net/profile/Hooman_Sedghamiz/publication/313673153_Matlab_Implementation_of_Pan_Tompkins_ECG_QRS_detector/data/5ac701130f7e9bcd51932f19/Sedghamiz-2014-Pan-Tompkins-Matlab.pdf

[42] A. Sagie, M. G. Larson, R. J. Goldberg, J. R. Bengtson, and D. Levy, "An improved method for adjusting the QT interval for heart rate (the Framingham Heart Study)," *The American Journal of Cardiology*, vol. 70, no. 7, pp. 797–801, Sep. 1992.

[43] "scikit-learn: machine learning in Python." Apr. 2019, accessed: 28.04.2019. [Online]. Available: https://github.com/scikit-learn/scikit-learn

[44] "sklearn.multiclass.OneVsRestClassifier — scikit-learn 0.20.3 documentation," accessed: 4.05.2019. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.multiclass.OneVsRestClassifier.html#examples-using-sklearn-multiclass-onevsrestclassifier

[45] V. V. Asch, "Macro-and micro-averaged evaluation measures [[BASIC DRAFT]]," Sep. 2013.

[46] "sklearn.linear_model.LogisticRegression — scikit-learn 0.20.3 documentation," accessed: 3.05.2019. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

[47] "sklearn.discriminant_analysis.LinearDiscriminantAnalysis — scikit-learn 0.20.3 documentation," accessed: 3.05.2019. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.discriminant_analysis.LinearDiscriminantAnalysis.html

[48] "sklearn.neighbors.KNeighborsClassifier — scikit-learn 0.20.3 documentation," accessed: 3.05.2019. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html#sklearn.neighbors.KNeighborsClassifier

[49] "sklearn.tree.DecisionTreeClassifier — scikit-learn 0.20.3 documentation," accessed: 3.5.2019. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html

[50] "sklearn.naive_bayes.GaussianNB — scikit-learn 0.20.3 documentation," accessed: 3.5.2019. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html

[51] "sklearn.svm.SVC — scikit-learn 0.20.3 documentation," accessed: 3.05.2019. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html

[52] "sklearn.neural_network.MLPClassifier — scikit-learn 0.20.3 documentation," accessed: 3.05.2019. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn. neural_network.MLPClassifier.html

[53] "sklearn.ensemble.RandomForestClassifier — scikit-learn 0.20.3 documentation," accessed: 3.05.2019. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn. ensemble.RandomForestClassifier.html

Appendix A

Experiments

A.1 Grid Search

From the table 3.4 all the models that was tested to find the best classifier for this project has been listed. By feeding all the R training data into these models with 10-fold cross validation the accuracy for each classifier is as illustrated in figure A.1 and listen in table A.1.
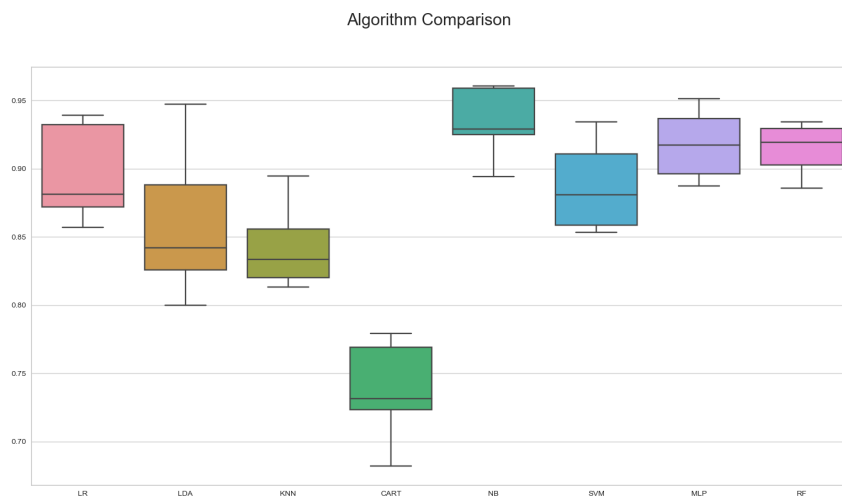
Figure A.1: Algorithms comparison with 10-fold cross validation

```python
lr_paramgrid = dict(solver=['newton-cg', 'lbfgs', 'liblinear', 'sag', 'saga'],
                    multi_class=['ovr', 'auto'])

lda_paramgrid = dict()

knn_paramgrid = dict(n_neighbors=[3, 5, 11, 15, 19],
                     weights=['uniform', 'distance'],
                     metric=['euclidean', 'manhattan'])

cart_paramgrid = dict(max_depth=range(1, 11),
                      min_samples_leaf=range(1, 6),
                      max_features=['auto', 'sqrt', 'log2'])

nb_paramgrid = dict()

lsvc_paramgrid = dict(C=np.arange(0.01, 100, 10))

mlp_paramgrid = dict(solver=['lbfgs'],
                     max_iter=[1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000],
                     alpha=10.0 ** -np.arange(1, 10),
                     hidden_layer_sizes=np.arange(10, 15),
                     random_state=[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
                     )

svm_paramgrid = dict(kernel=['linear', 'rbf'],
                     C=[1, 0.25, 0.5, 0.75],
                     gamma=[1, 2, 3, 'auto'],
                     decision_function_shape=['ovo', 'ovr'],
                     shrinking=(True, False))

rf_paramgrid = dict(bootstrap=[True, False],
                    n_estimators=[100, 200, 300, 400, 500, 600, 700, 800, 900, 1000],
                    random_starte=range(50, 150, 5)
                    )
```

Figure A.2: Parameter grid used for the grid search

Model	Mean Accuracy	Standard Deviation
LR	0.888	0.049
LDA	0.859	0.061
KNN	0.841	0.045
CART	0.733	0.048
NB	0.935	0.029
SVM	0.886	0.039
MLP	0.919	0.027
RF	0.911	0.027

Table A.1: Results before the grid search, values from figure A.1.

To tune the classifiers to give the best possible classifier, the grid search method as explained in chapter A.1 is used. By generating grids based on input values from the Scikit learn documentation page for the classifiers [46, 47, 48, 49, 50, 51, 52, 53]

Model	Mean Accuracy	Standard Deviation
LR	0.897	0.044
LDA	0.859	0.061
KNN	0.893	0.031
CART	0.711	0.056
NB	0.935	0.029
SVM	0.888	0.043
MLP	**0.950**	**0.019**
RF	0.933	0.032

Table A.2: Results after the grid search, as illustrated in figure 4.1.

Comparing table A.1 and A.2 shows that the RF and MLP classifiers have gotten higher accuracy's after hyper parameter tuning using grid search.

A.2 Predicted Probability Experiment

In this experiment the validation data for both the R and HRV datasets were tested for single heartbeat identification and for five heartbeats identification. The goal of this experiment was to visually inspect the identification, and in addition find the mean and standard deviation for the correct predictions and wrong predictions to find which classifiers provides the highest certainty in the predictions. The box plot shows the true predictions in label 2, and false predictions in label 1.

Logistic Regression
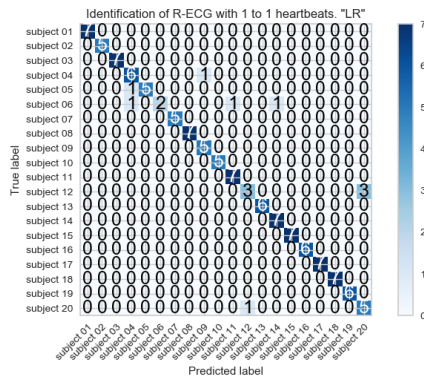
(a)
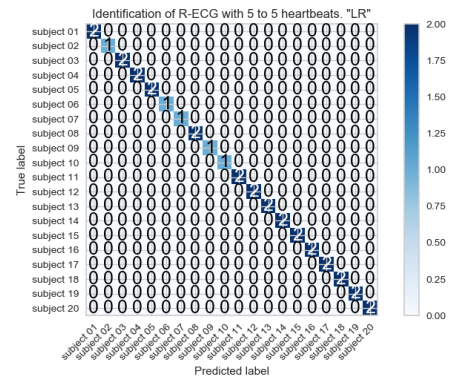
(b)
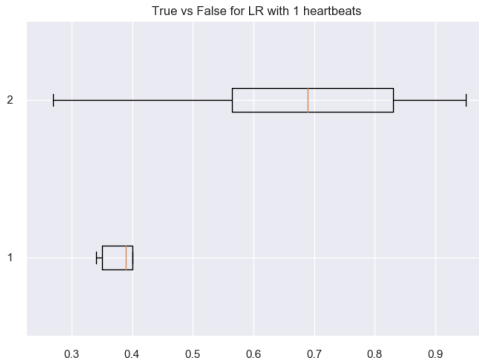
(c)
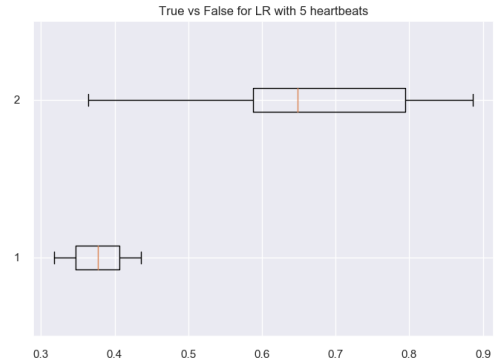
(d)

Figure A.3: Predicted probability experiment for the Logistic Regression classifier with the R validation set. (a) and (c) are from one heartbeat. (b) and (d) are from five heartbeats.
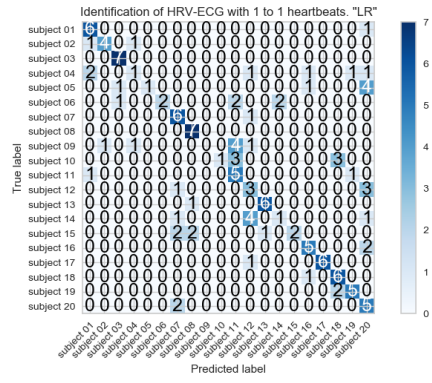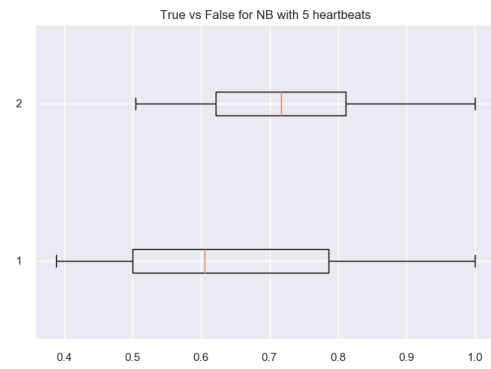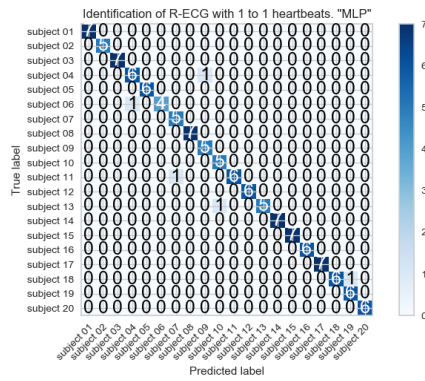
(a)
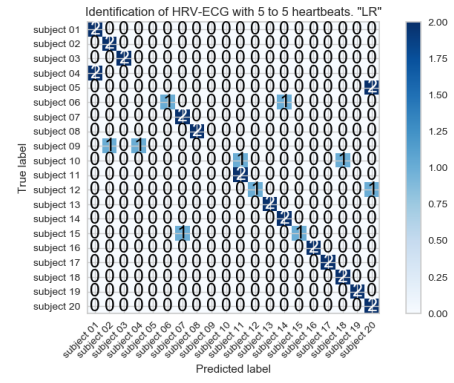
(b)
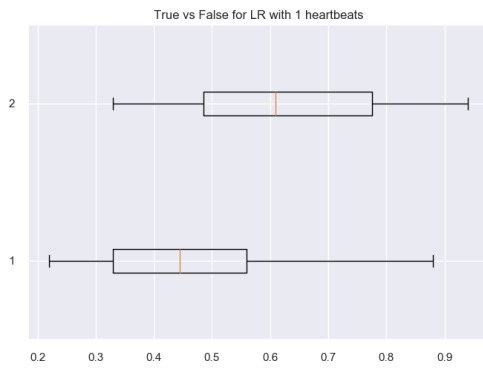
(c)
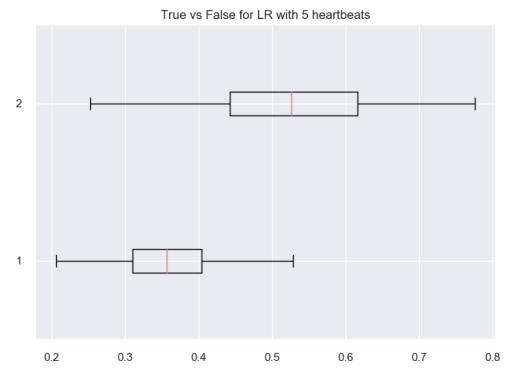
(d)

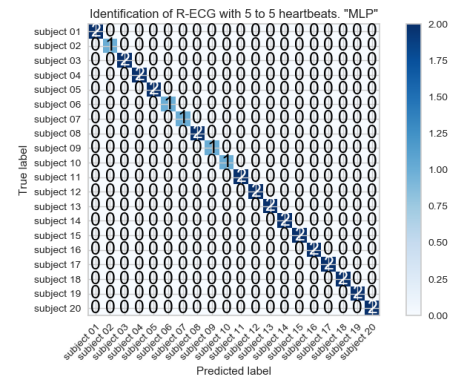Figure A.4: Predicted probability experiment for the Logistic Regression classifier with the HRV validation set. (a) and (c) are from one heartbeat. (b) and (d) are from five heartbeats.
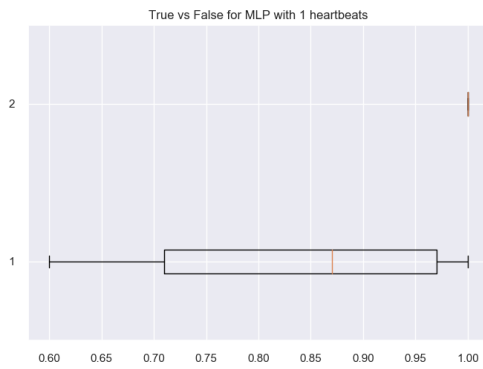
Linear Discriminant Analysis
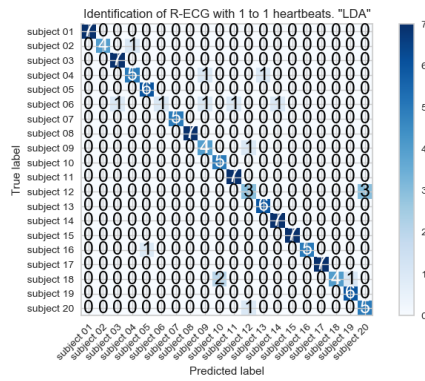
(a)
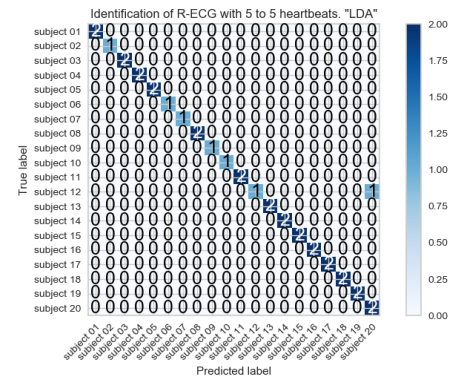
(b)
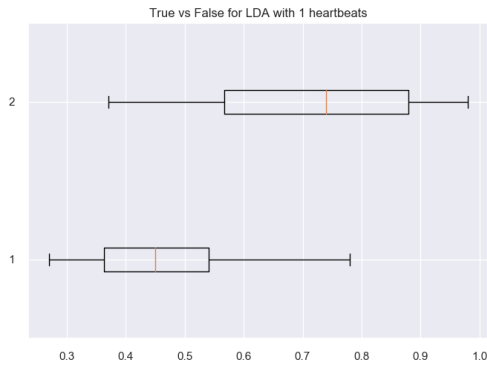
(c)
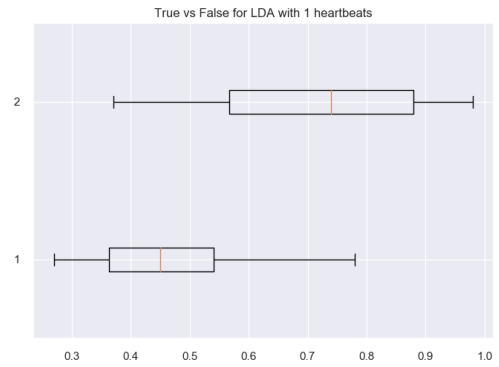
(d)

Figure A.5: Predicted probability experiment for the Linear Discriminant Analysis classifier with the R validation set. (a) and (c) are from one heartbeat. (b) and (d) are from five heartbeats.
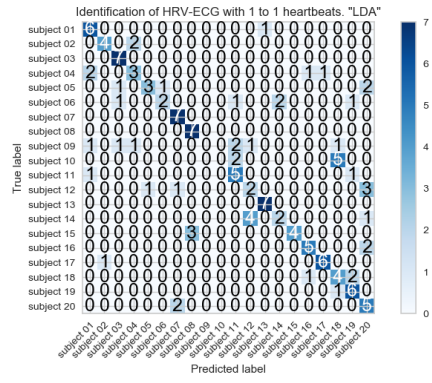
(a)
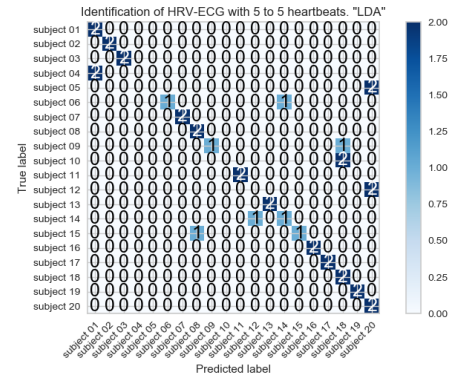
(b)
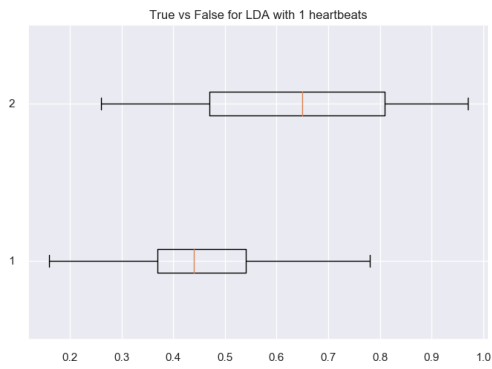
(c)
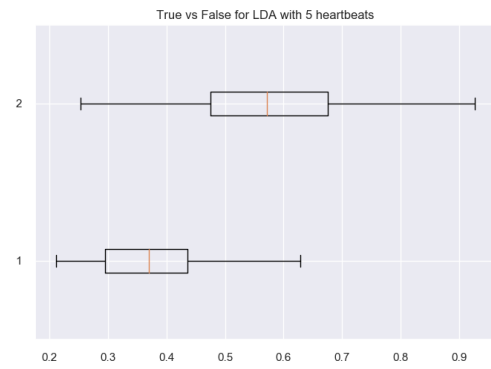
(d)

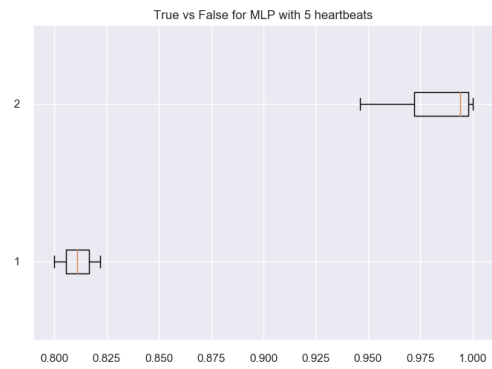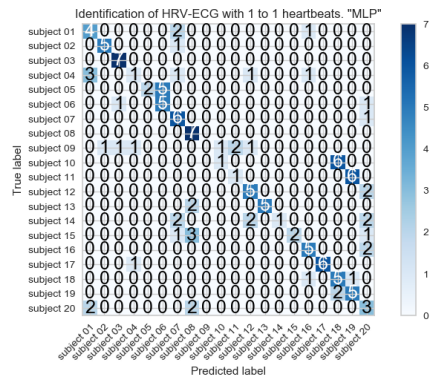Figure A.6: Predicted probability experiment for the Linear Discriminant Analysis classifier with the HRV validation set. (a) and (c) are from one heartbeat. (b) and (d) are from five heartbeats.

K-Nearest Neighbors
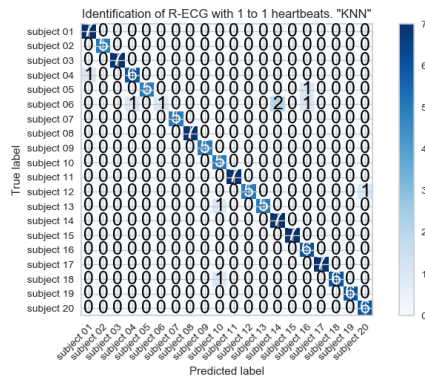
(a)
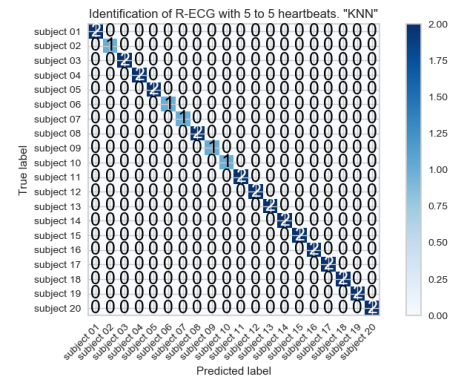
(b)
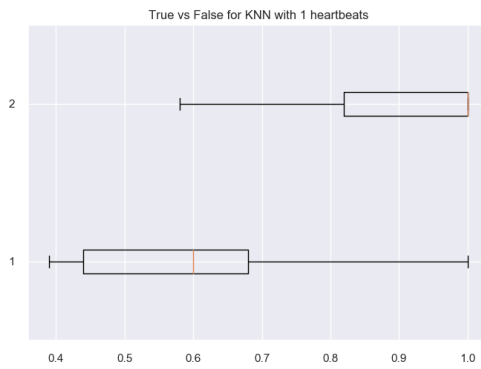
(c)
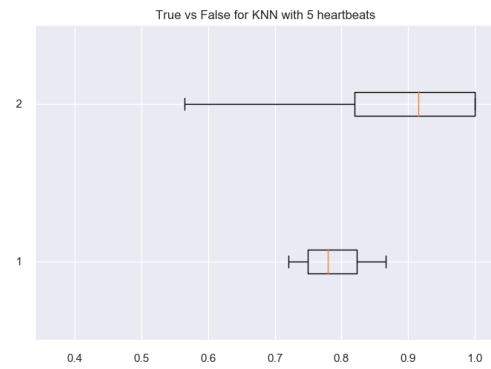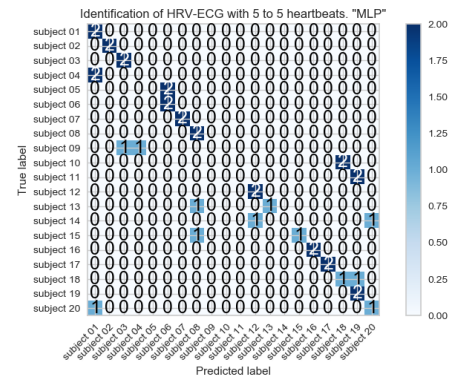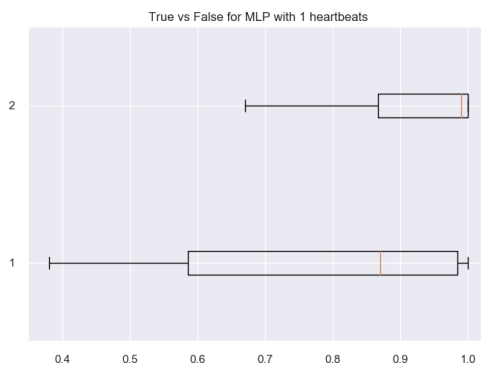

(d)

Figure A.7: Predicted probability experiment for the K-Nearest Neighbors classifier with the R validation set. (a) and (c) are from one heartbeat. (b) and (d) are from five heartbeats.
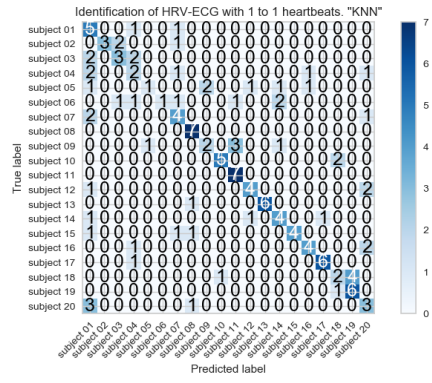
(a)
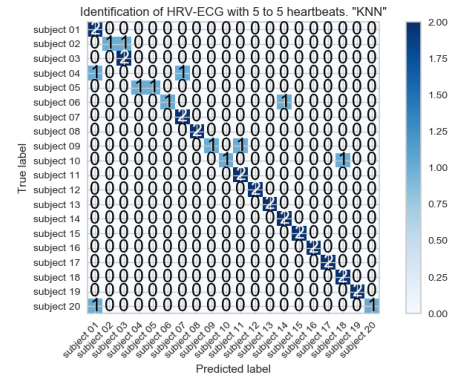
(b)
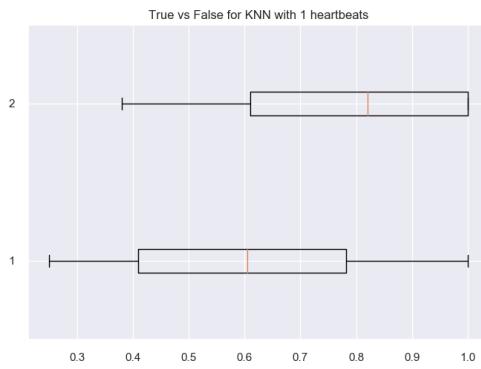
(c)
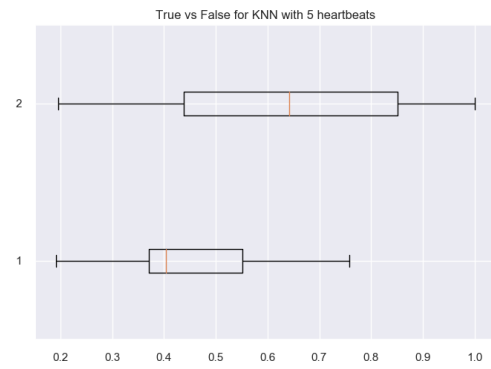
(d)

Figure A.8: Predicted probability experiment for the K-Nearest Neighbors classifier with the HRV validation set. (a) and (c) are from one heartbeat. (b) and (d) are from five heartbeats.
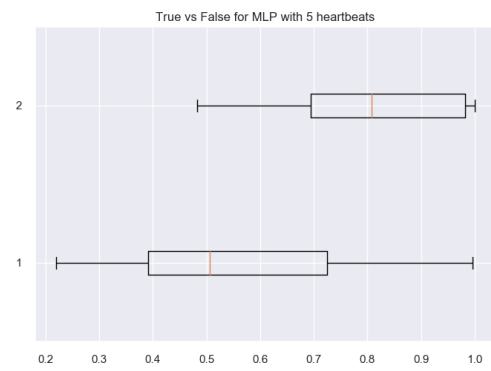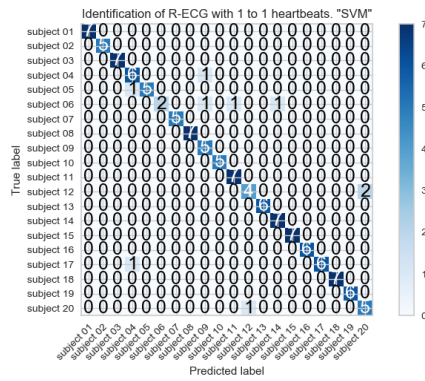
Naive Bayes
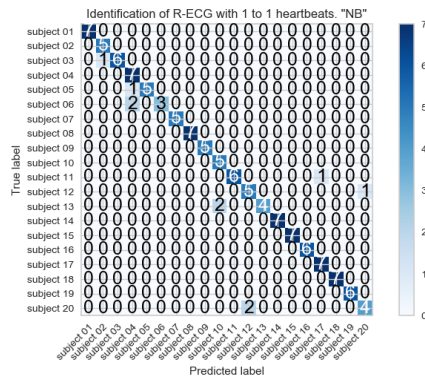
(a)
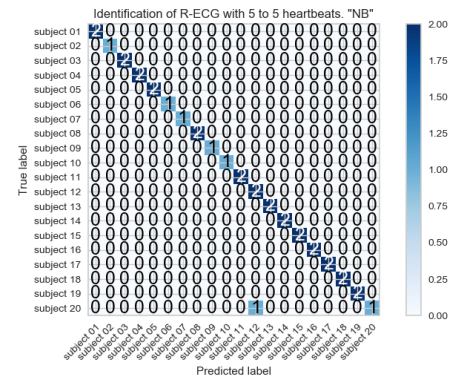
(b)
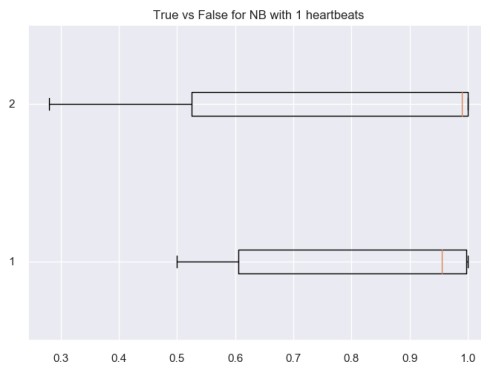
(c)
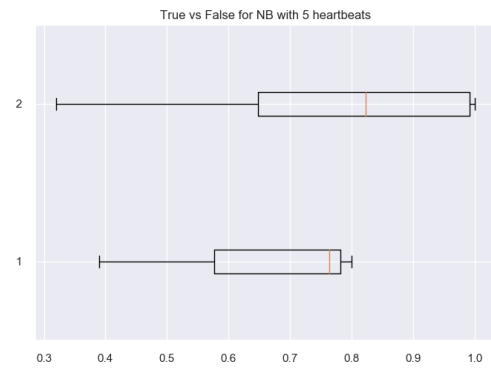
(d)

Figure A.9: Predicted probability experiment for the Naive Bayes classifier with the R validation set. (a) and (c) are from one heartbeat. (b) and (d) are from five heartbeats.
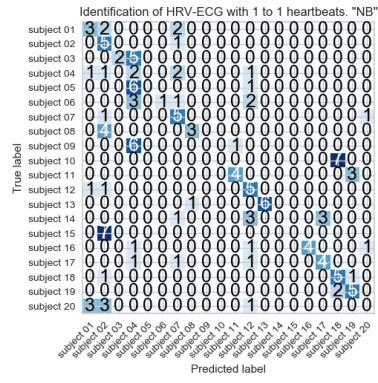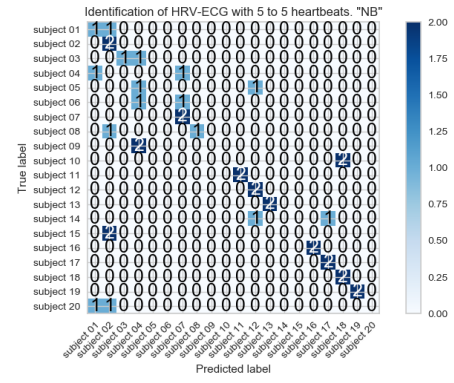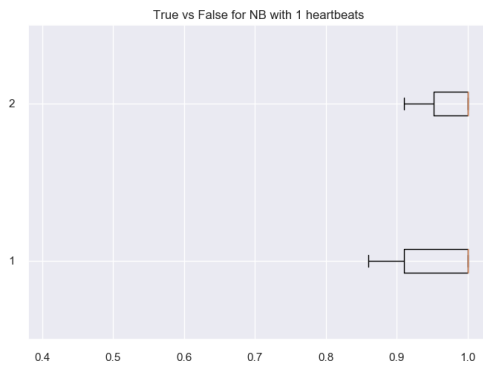
Figure A.10: Predicted probability experiment for the Naive Bayes classifier with the HRV validation set. (a) and (c) are from one heartbeat. (b) and (d) are from five heartbeats.

Multilayer Perceptron

(a) (b)

(c) (d)

Figure A.11: Predicted probability experiment for the Multilayer Perceptron classifier with the R validation set. (a) and (c) are from one heartbeat. (b) and (d) are from five heartbeats.
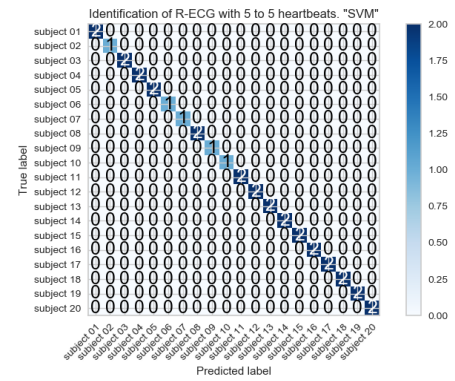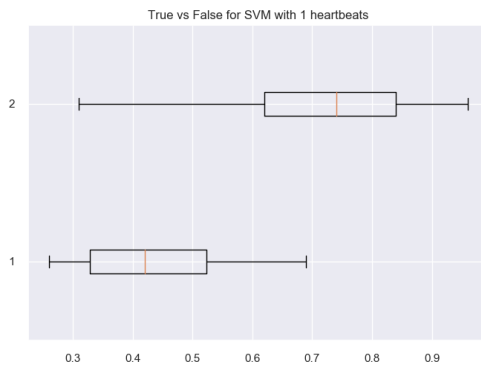
(a)

(b)

(c)

(d)

Figure A.12: Predicted probability experiment for the Multilayer Perceptron classifier with the HRV validation set. (a) and (c) are from one heartbeat. (b) and (d) are from five heartbeats.
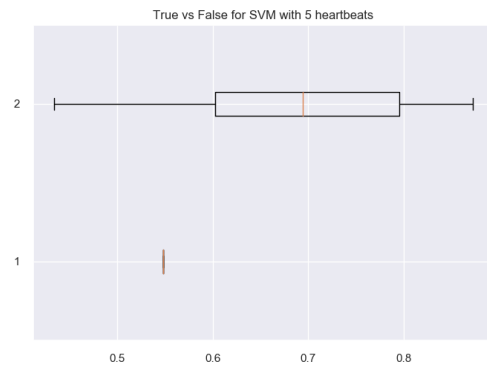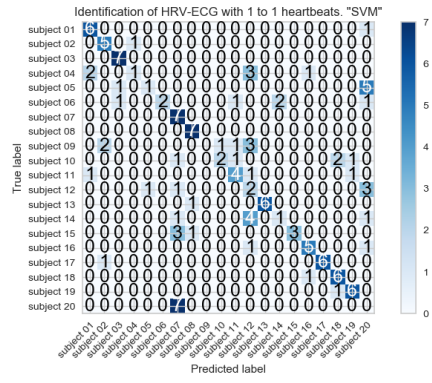
Support Vector Machine
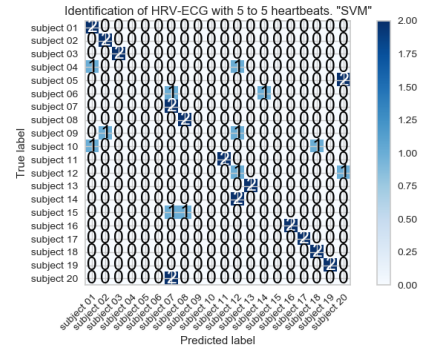

(a)

(b)

(c)

(d)

Figure A.13: Predicted probability experiment for the Support Vector Machine classifier with the R validation set. (a) and (c) are from one heartbeat. (b) and (d) are from five heartbeats.
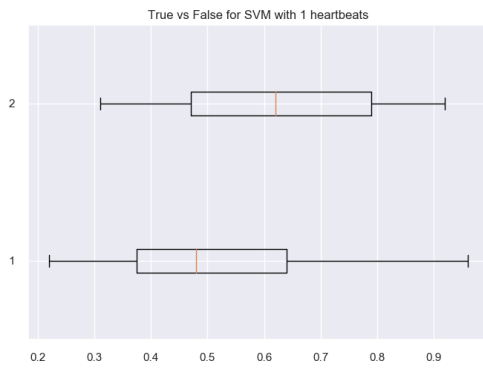
Figure A.14: Predicted probability experiment for the Support Vector Machine classifier with the HRV validation set. (a) and (c) are from one heartbeat. (b) and (d) are from five heartbeats.
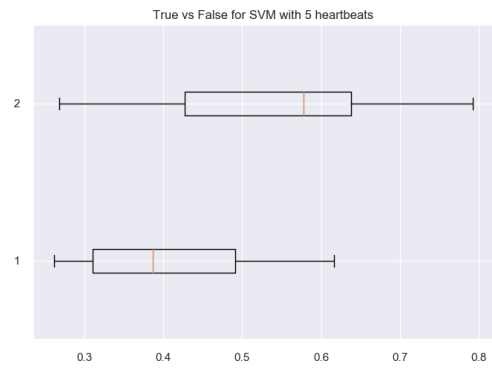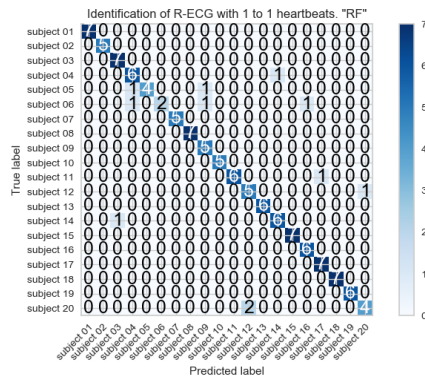
Random Forest
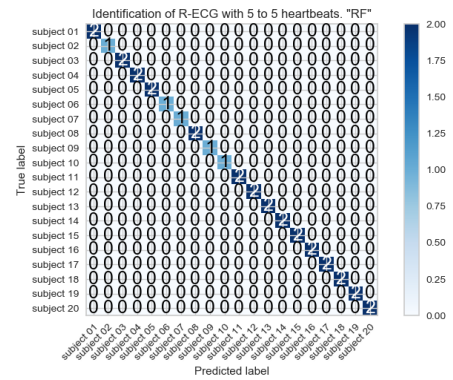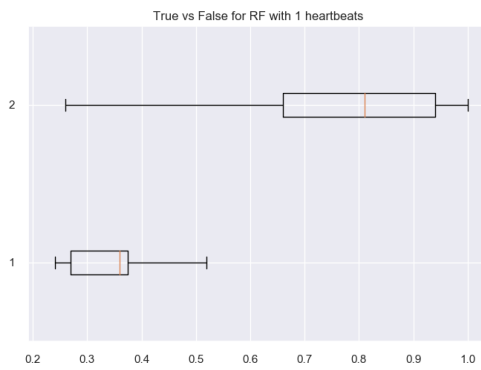

(a)

(b)

(c)

(d)

Figure A.15: Predicted probability experiment for the Random Forest classifier with the R validation set. (a) and (c) are from one heartbeat. (b) and (d) are from five heartbeats.
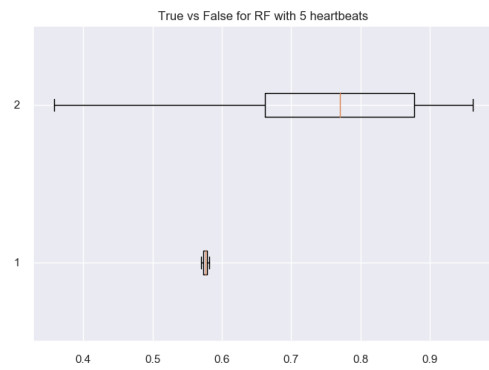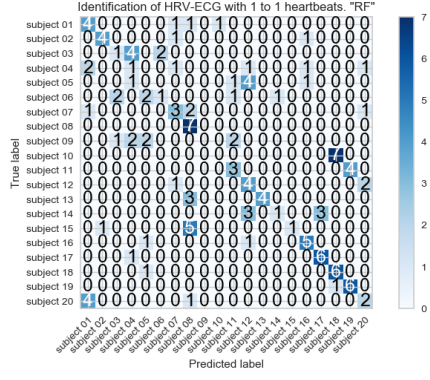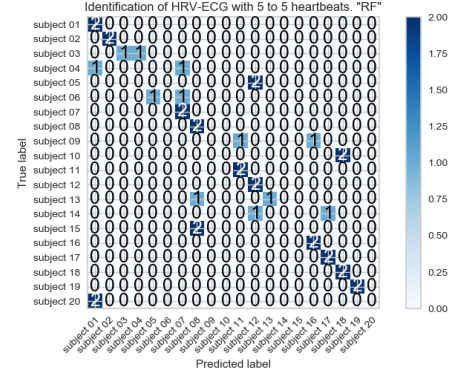
(a)

(b)
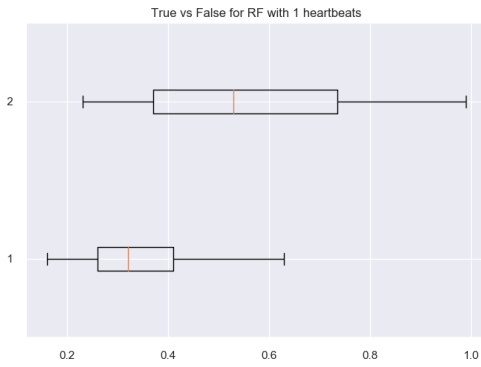
(c)
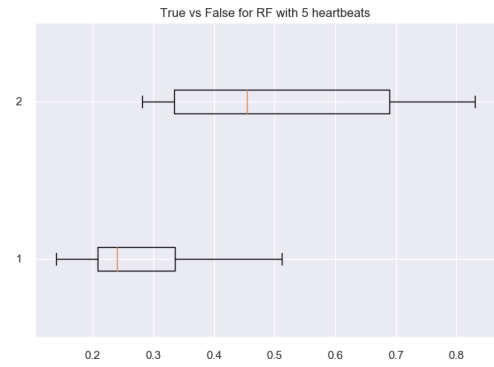
(d)

Figure A.16: Predicted probability experiment for the Random Forest classifier with the HRV validation set. (a) and (c) are from one heartbeat. (b) and (d) are from five heartbeats.

Results

From visual inspection on the predicted probabilities for correct and false classification for both the R and HRV validation sets, it showed that though most of the models had a clear separation in the predicted probabilities, only the Linear Discriminant Analysis and Random Forest classifiers had the desired results for this project. However, as the Random forest had better separation in both the R and HRV validation sets, the Random Forest classifier was evaluated to be the best fit for this project. The predicted probabilities have been calculated from the validation sets. Therefore, some additional certainty should be considered before using the highest prediction probability for false classifications as a threshold for the test sets.

Appendix B

Software

The code used for this thesis and the dataset collected has been embedded into this thesis in the "attachment.zip" file.

B.1   The Dataset

The dataset for this thesis consists of 20 subjects, where each of the subject has a recording for the R, HRV and M data.

B.2   Preprocessing.py

Preprocessing.py is a class which contains static methods for preprocessing the ECG signals.

B.3   Features.py

Features.py is a class which contains static method for heartbeat extraction, feature extraction and more.

B.4   MachineLearning.py

MachineLearning.py is the code which is used for training the classifiers, performing grid search, ROC-curves and the identification/authentication based on input parameters.

B.5   Randomizer.bat

Randomizer.bat is a file that allows the subject to choose a random number between 0 and 99, such that his or her data is saved as this folder number.

Appendix C

Data Collection Protocol

This Chapter contains the data collection protocol.

Protocol

Biometric authentication from ECG signals

Responsible: Vebjørn Kaldahl Bottenvik

Date: 12.03.2019

Place: University of Stavanger

Purpose:

The purpose of the project is to determine if a person's ECG signal can be used for biometric authentication. The data collected for this study will be used to generate and test an algorithm for such biometric authentication. The collected data will be anonymized and cannot be used to identify any individual taking part in the study.

Anonymization Procedure:

The anonymization procedure will contain a method to ensure the anonymity of each participant.

1. The data of each participant will be collected as explained in the collection Procedure.
2. The data will be saved to a temporary local folder.
3. The participant will run a script that gives the possibility save the data to a new folder that will be named based on the user input. If the folder already exists, user will be asked to try again.
4. When data from all the participants has been collected all the folders will be renamed in a chronological order from "1" and up.
5. The data in the temporary folder will manually be deleted.
6. The data is now fully randomized.

The reasoning for this method is the fact that the data only will contain the ECG signal itself. There is therefore no personal information saved. This method has the purpose of randomizing the order of the data collected.

Collection Procedure:

1. The participant will measure the ECG sitting still, having approximately a resting heart rate. This measurement will go on for 1 minute.
2. The participant will disconnect from the ECG electrodes, walk over to the stairs where the participant will run up and down 3 times to get an elevated heart rate.
3. The participant will again measure ECG sitting down. This time the signal will be measured as it goes from a high heart rate. This will go on for 2 minutes.

Flow Chart

Information regarding the data collection method

Gives the volunteer a few minutes to get calm, getting the best estimate for a resting heart rate

Resting Heart ate ECG collection

R-ECG data for training and testing

Collect one minute of R-ECG data

Physical exercise (A brief running exercise to increase heart rate)

Heart Rate Variation ECG collection

HRV-ECG data for training and testing

Collect two minute of HRV-ECG data, this collecting is longer due to getting different ranges of heart rate variance

A minute of resting

Walking with rigid arms ECG collection

M-ECG for testing

Collecting data from a person moving with his/her hands on the electrodes. Purely for testing purposes.