WORKSHEETS and ARTICLE (Patient identification based on unfiltered ECG using Hierarchical Temporal Memory)

Aalborg University
10th semester
School of Medicine and Health

Project group: 17gr10416
Rune Kongsgaard Rasmussen

07/06/2017

In the present study the following tools and software were used to produce results, statistics and figures: MATLAB® NuPIC and Draw.IO.

School of Medicine and Health
Biomedical engineering
Frederik Bajers Vej 7
Telefon 96 35 97 31
Fax 98 13 63 93
http://smh.aau.dk

Title: Patient identification based on unfiltered ECG using Hierarchical Temporal Memory

Keywords: ECG, HTM, Machine learning, AI, electrocardiogram, memory, intelligence

Project period: ST10/BME4, Spring semester 2017

Project group: 17gr10416

Participants:

Rune Kongsgaard Rasmussen

Supervisor: Johannes Struijk

External supervisor: Rasmus Wiberg Nedergaard

Page count (Article): 10

Page count (worksheets): 43

Abstract:

Introduction Over 17.5 million people die every year due to cardiovascular disease. The most widespread used to detect heart disease is the 12-Lead ECG. In ECG research there has been a lot of research in understanding its features and the underlying grammar of the ECG not only for diagnostics but also for identification of a specific person or even mood. Hierarchical Temporal Memory (HTM) seemed to be able to identify the underlying grammar automatically in a robust matter. It was therefore evaluated to which extent HTM could identify and use biometric grammar of the ECG in the context of identifying subjects based on their heartbeat from unfiltered ECG.
Methods Data subsets with different number of subjects were created from an ECG database with 25,000 subjects. Two sessions were included from each subject of 500 Hz 10 seconds ECG. Lead II was used from these ECGs and the dataset contained healthy and unhealthy subjects. A HTM model was configured and built using Numenta NuPic software and exposed to the different subsets. Subsets were applied with different number of iterations spanning from 1 to 1,000.
Results The maximum accuracy was achieved from using only 10 subjects where the accuracy was found to be 31.3 % and down to 0.07 % for 1,250 subjects.
Conclusion The number of subjects decreased this accuracy and the number of iterations had no effect. It was not possible to determine if the accuracy found was due to limitation of inter-subject variability of the ECG or to the configuration of HTM. The authors of the software used in this current study now provide tools that might be a logical next step in improving the results.

Resume (Dansk)

I dette projekt er der udarbejdet en artikel og en rapport om Hierarchical Temporal Memory. I denne sektion er det engelske resume oversat til dansk.

Introduktion

Over 17.5 milioner af mennesker dør hvert år pga kardiovaskulære problemer hvoraf størstedelen af tilfældene er mulige at forhindre. Forskellige metoder kan bruges til at detektere hjerteproblemer og hjertesygedomme men den mest brugte metode er det 12-ledet EKG. Egenskaber fra et EKG kan udtrækkes og bruges i Machine Learning. Det viser sig dog at det ikke nødvendigvis er en god ide at anvende såkaldte eksperter til at udtrække disse egenskaber fra EKGet. I EKG forskninger der lavet en masse forskning i at forsøge at forstå disse egenskaber og den underlæggende grammatik af EKGet, ikke kun til diagnostisk brug men også til at identificere en person eller deres humør. I en litteratursøgning om Machine learning blev det fundet at Hierarchical Temporal Memory virker til at kunne identificere de underliggende egenskaber og grammatik i EKG'et på en automatisk og robust måde. Det var derfor evalueret hvor godt at en implementing af Hierarchical Temporal Memory kunne identificere og bruge den underliggende grammatik af EKG'et til at klassificere hvilken person et hjerteslag tilhørte baseret på et hjerteslag afledt fra ufiltreret EKG.

Metoder

Flere del-sæt hvert indeholdende et forskelligt antal personer var generet fra en EKG database med over 25,000 personer. To sessioner blev inkluderet fra hver person hvor hver enkelt session indeholdet 10 sekunders data og var optaget med 500 Hz. Denne data indeholdte flere led men kun II-ledet var brugt. Det samlede datasæt indeholdte både sunde og raske personer men også personer der led af forskellige sygdomme der havde betydning for deres hjertefunktion. Hvert enkelt del-sæt blev splittet i træning og test data hvor der var fire gange så meget trænings data som test data. En implementering af Hierarchical Temporal Memory blev fortaget ved at bruge software fra Numenta kaldet NuPic. Hvert eneste del-sæt der indeholdte træningsdata blev anvendt i denne implementering. Disse datasæt blev vist med et forskelligt antal iterationer hvor kun en enkelt iteration blev fortaget for noget data og op til 1,000 iterationer for andet data.

Resultater

Den største nøjagtighed blev opnået ved kun at bruge 10 personer hvor at denne nøjagtighed blev målt til at være op til 31.3 % og mindst for 1,250 patienter hvor den var under et procent. I del-sæt hvor flere iterationer var brugt var der ikke fundet bedre resultater. Ved at kombienere forskellige del-sæt sås der heller ikke en forbedring af resultater.

Konklusion

Det var muligt at identificere hjerteslag med op til en nøjagtighed på 31.3 % men ved at forøge antal af personer inkluderet i klassifikationen faldt denne nøjagtighed. Slev ved at forøge antal af iterationer hvormed data blev brugt i systemet var der ingen forbedring af resultaterne. Resultaterne af dette studie er stadigvæk dog forholdsvis nye og interessante ide at Hierarchical Temporal Memory ikke har været brugt i en biometrisk applikation før. Derudover er antal af personer høje i forhold til mængden af data per person når man sammenligner med andre studier fundet under litteratursøgningen i dette projekt. Det var ikke muligt at finde ud af om den fundne nøjagtighed

var begrænset pga variabiliteten mellem personer i datasættet eller fordi at konfigurationen af Hierarchical Temporal Memory ikke var god nok. Fabrikanten af det software der blev brugt i dette studie er udkommet med et nyt værktøj kaldet "swarming" til at hjælpe med at forbedre konfigurationer af Hierarchical Temporal Memory. Det virker derfor som et logisk valg at bruge dette værktøj til vider undersøgelser i forbedring af nøjagtigheden med Hierarchical Temporal Memory.

Patient identification based on unfiltered ECG using Hierarchical Temporal Memory

R. Kongsgaard Rasmussen
Aalborg University
17gr10416@hst.aau.dk

Abstract

Introduction Over 17.5 million people die every year due to cardiovascular disease where the majority of these cases are preventable. Different methods can be used to detect heart disease but the most widespread used is the 12-Lead ECG. Features can be extracted from the ECG and used in machine learning. Using an expert to extract these features does not seem to be the correct way. In ECG research there has been a lot of research in understanding these features and the underlying grammar of the ECG not only for diagnostics but also for identification of a specific person or even mood. Hierarchical Temporal Memory (HTM) seemed to be able to identify the underlying grammar automatically in a robust matter. It was therefore evaluated to which extent HTM could identify and use biometric grammar of the ECG in the context of identifying subjects based on their heartbeat from unfiltered ECG.
Methods Data subsets with different number of subjects were created from an ECG database with 25,000 subjects. Two sessions were included from each subject of 500 Hz 10 seconds ECG. Lead II was used from these ECGs and the dataset contained healthy and unhealthy subjects. Each data subset was split into a 4:1 training to test data ratio. A HTM model was configured and built using Numenta NuPic software and exposed to the different subsets. Subsets were applied with different number of iterations spanning from 1 to 1,000.
Results The maximum accuracy was achieved from using only 10 subjects where the accuracy was found to be 31.3 % and down to 0.07 % for 1,250 subjects. Increasing the number of subjects and combining subsets decreased the accuracy of the system. The different number of iterations had no effect on the accuracy of the system.
Conclusion It was possible to identify with up to an accuracy of 31.3 % but the number of subjects decreased this accuracy and the number of iterations had no effect. The results of this study are however somewhat novel compared to other studies because of the high number of subjects compared to the amount of data for the individual subject. It was not possible to determine if the accuracy found was due to limitation of inter-subject variability of the ECG or to the configuration of HTM. The authors of the software used in this current study now provide tools to improve configurations of HTM. Using these tools might therefore be a logical next step.

1. Introduction

Cardiovascular disease is the leading cause of mortality and illness in the western world [1] with over 17.5 million deaths worldwide and 80 % of them being due to heart attacks and strokes with the majority of the deaths being preventable [2]. Heart attacks (ischemic heart disease) alone resulted in more than 7.4 million deaths [2].

The 12-lead ECG is the gold standard for diagnosing and monitoring coronary heart disease patients [3]. It is also the most widely used in the diagnosis of suspected heart disease [1].

There is a physiological reason for the usage of ECG in diagnostics. Since ECG is the readout of the electrical activity of the heart, any electrical malfunction will change the recorded signal [4, 1]. The heart and its electrical activity is modulated by both intrinsic and extrinsic factors [5, 1], which means that not only heart problems are detectable but also problems occurring extrinsically from the autonomic nervous system (ANS) would be detectable. The ANS also regulates internal organs other than the heart [6] thereby the 12-lead ECG could potentially be used for a more general health status. One limitation of the ECG is that it doesn't indicate the actual mechanical contraction and pump function of the heart [7].

Since ECG features indicate the heart's electrical activity [7] it can be used by a doctor to perform further investigations or diagnosis. [1] Some of the useful features that can be derived from the ECG are for example Heart Rate Variability (HRV), QT interval (to diagnose Long QT and Short QT syndromes) and QRS width (to diagnose e.g., bundle branch block and ectopic beats) [1].

There is a lack of a gold standard feature set [8, 9] as new feature sets are derived each time from disease annotated original ECG databases [10, 11] instead of from a pool of already defined features. Therefore, behind much of ECG research there is a lot of work being done to establish an underlying grammar of the ECG as the minimal ideal reduced dataset. From this dataset, one should be able to not only diagnose suspected heart disease but also identify a specific person or even their mood. The idea of using an expert for feature extraction does not seem to be the correct way but rather, features should be extracted synaptically or statistically [8, 12, 13, 14].

The papers by Casarella M. [8], Oster et al. [12], Agrafioti [13] and Lu et al. [14], investigating ECG with features not extracted by an expert, show promising results thereby bypassing the feature extraction problem. Some of these results which were found in Casarella M. [8], Masetic and Subasi [15] and Seera et al. [9] showed the accuracy of close to 100 % where some algorithms used in these showed to be more robust to noise than others. These results could be due to: overfitting, the grammar in the dataset being ideal for the application of the respective algorithm or a general descriptive grammar of the ECG was found.

This project therefore investigated if this general descriptive grammar could be found by using methods applied to another application than their original one. The Hierarchical Temporal Memory (HTM) methodology used in Casarella M. [8] was chosen since the research field fits within the scope of this article and provides a novel feature extraction and classification that does not require an ECG expert. The application chosen follows Agrafioti [13] which performed identification of subjects based upon ECG heartbeats. The application was chosen since it had higher error rates than other application areas found, but was still found to be within an area that showed promise [13].

The following was investigated: *To what extent can the underlying biometric ECG grammar be found with HTM on subjects heartbeats over multiple sessions?*

2. Methods

In the current study parts of the methodologies applied in Casarella M. [8] are used, that is a dataset was prepared and a Hierarchical Temporal Memory model (HTM) was built and configured. The HTM model is illustrated in figure 1. Multiple experiments were run in different configurations in order to investigate the influence of number of subjects and number of iterations on the results. The non-shared configuration for each of these experiments is described under the subsection experiments.

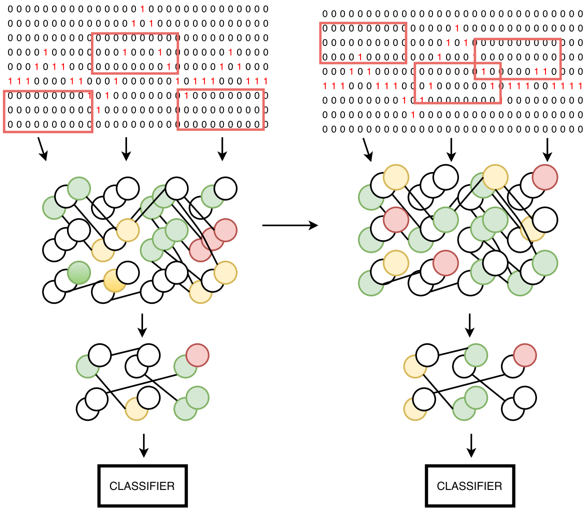

Figure 1: *The goal of the HTM model is to mimic the human neocortex columnar structure and algorithmic property as these properties are a big part of human intelligence [8]. This illustrates the pyramidal nature of HTM where initially the ECG is encoded to a binary 2D image where features get extracted and reduced layer by layer until they are fed to a classifier. Different parts of this image marked with red squares act as input for a set of 2D columns each with cells marked as circles. On the left some cells are active due to an input to their respective column. These active cells are marked green. The yellow cells are connected to two other cells that are active and are therefore in a predictive sate. Not every connection is illustrated here for the sake of simplicity. The red cells are only connected to a few cells so they are not in a predictive state. On the right a new input has arrived. Every cell in a column is active since none of the cells were in a predictive state before. In the other columns the cells that were predictive before are now active. The model learns by adjusting the weight of connections between the input and columns and between cells through iterations. A classifier is added in the end and performs classifications based upon the features derived in the layers of HTM*

2.1. Application

The real life application used in this study is based on the literature Agrafioti [13]. Described in Agrafioti [13] are three biometric applications. The best results in terms of having a lower error rate, was in the application where heartbeats derived from a finite known population's ECG were to be identified over time. In that application a continuous classification on a non overlapping dataset of five seconds ECG was used. Temporal information can be learned without the overlap in HTM. [8] That application is therefore mimicked in the current study so that only single consecutive heartbeats are used to identify a subject over time.

2.2. Software

Some preprocessing had to be done to the data since images need to be generated from the original data set. This was done in MatLab® 9.1.0.441655 (R2016b). The actual HTM model is however not supported in MatLab®, therefore software by Numenta NuPIC 0.7.0-dev [16] was used. Experiments were run on 2015 MacBook Pro 15 Inch 2.8 GHz i7 with 16 GB of ram. MacOS version 10.12.4. Python 2.7.10. Furthermore a MongoDB 3.04 database was created with the data to avoid memory leaks.

2.3. Dataset

The origin of the dataset is a database by the general practitioners of Copenhagen, Denmark. It is 500 Hz sampled unfiltered ECG for 25,000 patients each with two sessions. There were 14,225 female and 10,775 male with a total mean age of 64 years and a standard deviation of 16.64. From each session the full length of 10 seconds was chosen. In the application being mimicked from Agrafioti [13] they used multiple leads but in this study the lead II was chosen to match the methodology by Casarella M. [8]. The dataset contained both healthy and unhealthy patients which is a broader spectrum than both Agrafioti [13] and Casarella M. [8] who only used either one of the two patient groups.

2.3.1. **Subsampling of the dataset**

Two subsampled dataset called optimisation and control were substracted from the entire dataset. Samples were assigned to the two subsets in a random manner automatically by the file system of the computer running the software. 1,250 subjects (5 % of entire dataset) were included in each of the subsets. This number was chosen in order to reduce the runtime of the software, since initial investigations indicated that it would not be able to finish within time scope otherwise. In order to evaluate the influence of number of subjects, subsets of 10 and 100 subjects were also included.

2.3.2. **Segmentation of heartbeats**

In order to segment the heartbeats from the ECG the Pan Tompkins algorithm was used as suggesed by Elgendi [17], Agrafioti [13] and Dubois et al., 2007 [10]. The Pan Tompkins algorithm, developed by Pan and Tompkins [18], provided an accuracy of 99.3 % on 12-lead ECG arrhythmic data and can be used on the unfiltered ECG signal.

The width of the heartbeat was defined to be 740 ms in this current study. A fixed length was chosen to match with the methodologies by Casarella M. [8]. The exact length of a heartbeat is affected by both disease and heart rate, but was calculated based upon it normally being within; 80-100 ms for the P-wave, 120-200 ms for PR-interval and 300-440 ms for QT-interval resulting in a maximum total length of 740 ms. [8].

2.3.3. **Encoding images**

Images were encoded so that each sample corresponded to a pixel. Since the length of a heartbeat was calculated to be 740 ms, and each sample corresponds to 2 ms, a width of 370 samples (370 pixels) were selected for the image. This corresponds to the length of the heartbeat normalized to the sampling frequency.

In Casarella M. [8] the height of images were 96 pixels so the same height for images were selected for the current study.

In order to determine the vertical pixel a sample belonged to, a grid was created with 96 values each representing a pixel. The distance between the value of a sample and the 96 pixels in the grid were calculated and the pixel with the minimum distance was selected to be the corresponding pixel of the sample.

In case of a tie between distance of pixels, the upper pixel was selected to be the corresponding one.

Grids were created for each individual recording so that the maximum value of the grid corresponded to the maximum value of the session and the minimum to the minimum value. Each step in the grid was calculated as described in expression 2.

$$value(pixel) = min + pixel * (max - min)/96 \quad (1)$$

2.3.4. **Training and test groups**

The very essence of machine learning is the ability to predict future values based upon previous experience. Subjects data from each of their respective subsets were therefore split into training and test groups.

In Casarella M. [8] different ratios between the two groups are described. They describe that the implementers of Numenta NuPIC suggest a 8:1 training to test data ratio but that they usually used a 1:1 ratio. Casarella M. [8] uses roughly a 1:2 ratio for their own study but the ratio is not the same for each class they investigated.

In the current study the 4:1 ratio was chosen. The data amount is limited for each class in the current study compared to what was used in Casarella M. [8] and increasing the ratio gives more training data and thereby better accuracy might be achieved.

All data on each subject was used as a combined dataset where a 4:1 ratio of data was extracted for each of the subjects. This gave roughly 25 beats for training and 6 beats for test for each subject.

2.4. Configuration of algorithm

There are not yet any gold standards for selecting the values for parameters in an implementation of HTM as described in Casarella M. [8] and Hawkins et al. [16]. In this section the high level parameters, such as number of layers, are selected before the low level parameters, such as number of columns in each layer.

As described by Casarella M. [8] whose methodology is mimicked in the current study, parameters had to be tweaked to fit within the problem domain. Since the application or problem domain in the current study is biometrics, values in this current study would not necessarily be the same as of [8].

2.4.1. **Number of layers**

The number of layers to use is not directly specified by Numenta, but experiments have been made with both three and four layers. More layers increased the robustness of the system but decreased the accuracy. [8] For image classification problems, three layers seemed to produce the best results [8], therefore a three layer structure was used in the current study.

2.4.2. **Configuration of layers**

The configuration of the two bottom layers were made with the following considerations from Casarella M. [8]: The image size, patterns within the image, the correlation between number of layers and number of columns in each layer. Following the methodology described in that literature, two prime factors were derived based upon the image size and used to calculate the configuration of columns. These calculations are made in expression 2 and 3.

$$96 \Rightarrow 3 \times 2 \times 2 \times 2 \times 2 \times 2 \quad (2)$$

$$370 \Rightarrow 2 \times 5 \times 37 \quad (3)$$

A pyramidal regression was found in Casarella M. [8] to be required for the proportion of the images. It is specified that the first layer should be an eighth of the original data set and the second layer should be a half of the first layer. The last layer is the classifier. It is however not possible to fulfil these requirements with a width of 370 pixels. Therefore the images were adjusted to a width of 384 pixels (by chaining the width of the window) and new primes were calculated in expression 4.

$$384 \Rightarrow 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 5 \quad (4)$$

By using the primaries it was possible to calculate the following properties of the layers in calculation 5 and 6. This means 48 × 12 columns (width × height) in the first layer and 12 × 3 (width × height) in the second layer.

$$96/8 = 12 || 12/4 = 3 \quad (5)$$

$$384/8 = 48 || 48/4 = 12 \quad (6)$$

2.4.3. **Capacity of learning**

The capacity of a layer defines how many underlying patterns can be stored. Essentially this is the configuration of the SDR. The expression for the capacity in a layer is given in expression 7. n is number of columns where k is number of active columns. By selecting a rounded 2% suggested in Hawkins et al. [16]

the number of columns for layer one is $48 \times 12 = 576$, the number of active columns is 12 and the capacity is $2.4810e + 24$. In the second layer the number of columns for layer one is $12 \times 3 = 36$, active columns is only 1 and the capacity is $1.2517e + 09$.

$$\frac{n!}{k!(n-k)!} = \binom{n}{k} \tag{7}$$

2.4.4. Selecting number of cells

The number of temporary contexts that can be represented is given in 8 and limited by c number of active columns and n cells in each of the columns. Examples are given for a layer of 4 cells in each columns in Hawkins et al. [16]. This number allows for a big temporal memory and it is argued in Hawkins et al. [16] that a layer of above the 4 is usually not needed. The temporal patterns (context) that can be learned in this implementation is therefore 16,777,216 and 4 in the first and second layer respectively. This guideline is also followed in Casarella M. [8].

$$n^c \tag{8}$$

2.4.5. Classifier and performance measures

The top layer (classification algorithm) was chosen to be Naive Bayes Classifier (NBA) since it was also used in Casarella M. [8] and a class was defined for each subject. This classifier outputs a sample's probability to belong to each of the classes as a test is performed. The three most likely classes and corresponding probability for each of those classes were used in further analysis from every test. The further away from the most likely class a prediction is, the less likely it is for the system to be able to classify such a sample and an arbitrary cut-off of three classes was therefore chosen.

2.5. Experiments

Experiments were conducted with different numbers of iterations in this present study as of [8]. The span of iterations were set to be between 9,000 to 60,000 times in the study by Casarella M. [8] and was only done on the spatial pooler. No improvement in classification was seen for over 21,000 iterations. Essentially it was theorised that the more times an image is exposed to the spatial pooler the better the underlying patterns can be detected. For this application the ideal number of iterations needed were not known and multiple experiments were therefore conducted with an arbitrary number of 1, 100 and 1,000 iterations respectively for the control subset. The trial with 1 and 100 iteration(s) were then run again but with the optimisation subset instead. An experiment with only 1 iteration but with a combined dataset of the control and optimisation dataset was also run. This last experiment was conducted to investigate the influence of the size of the dataset on accuracy of classification. Investigation of the size of the dataset was also further supported by using the subsets that included 10 and 100 subjects respectively. The number of iterations used for these were 1, 10 and 100.

3. RESULTS

The results are represented in table 1 for the different number of iterations used and the optimisation and control subsets.

Table 1: *Results of using HTM with the different subsets are given in this table. The accuracy of correctly classifying a sample as the most likely or within the top three most likely is given*

Iteration	Subset	Acc (1)	Acc (3)
1	Optimise	0.07 %	0.24 %
100	Optimise	0.07 %	0.24 %
1,000	Optimise	0.07 %	0.24 %
1	Control	0.09 %	0.27 %
100	Control	0.09 %	0.27 %
1	Optimise & Control	0.04 %	0.01 %

The results did not improve with the number of iterations. By creating a dataset consisting of both of the subsets, the accuracy dropped to half of the mean of the accuracies of each of the subsets. The results from evaluating for number of subjects are presented in table 2.

Table 2: *Results of using HTM with the different subsets are given in this table. The accuracy of correctly classifying a sample as the most likely or within the top three most likely is given*

Iterations	Num subjects	Acc (1)	Acc (3)
1	10	8.89 %	31.3 %
10	10	8.89 %	31.3 %
100	10	8.89 %	31.3 %
100	100	4.47 %	18.5 %

The results of these showed an improvement compared to the former results of this current study with the number of subjects influencing the accuracy of the system. The number of iterations did not change the accuracy. These findings align with what was already found earlier in this study. Two classifications made by the classifier has been included in figure 2 and 3 respectively. Figure 2 is a correct match where figure 3 is not. These figures illustrate how similar both the correctly and wrongly classified images look with their matches.

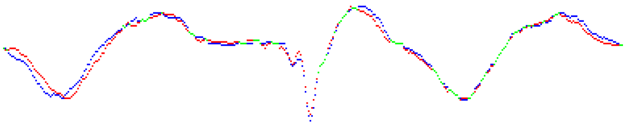

Figure 2: *Correct match from random subject's encoded heartbeat. The shared pixels are marked with green, the non-shared are marked with red and blue where each of these colours represents two different beats*

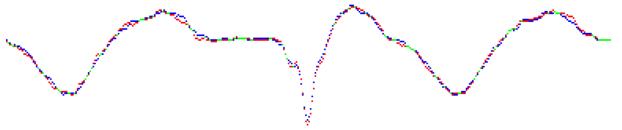

Figure 3: *Wrong match from random subject's encoded heartbeat. The shared pixels are marked with green, the non-shared are marked with red and blue where each of these colours represents two different beats*

4. Discussion

The results found in this study were less accurate than what was found by Agrafioti [13], Gregg et al. [4] and Casarella M. [8]. In Casarella M [8] the application was not biometric but the HTM was also used and an almost 100 % accuracy was found. That study did however only differentiate between four classes and in this study it was found that increasing the number of classes lowered the accuracy of the system. Furthermore these results were achieved using 60,000 iterations. When only 3,000 iterations were used the results were an accuracy of 70 %. These number of iterations are higher than what was used in the current study.

In the biometric application in Agrafioti [13], mimicked in the current study, the lowest equal error rate was 10 %. This was however only on a database with up to 52 subjects whereas the current study used up to 2,500 subjects. More data was used for classification in Agrafioti [13] as they used five seconds of ECG recordings providing more data for the classifier to make its decision than in the current study where single heartbeats were used. Using more data should improve the accuracy of the system [19, 20, 21].

Filtered ECG was used in Agrafioti [13] as recommended in literature [22, 23, 24] but this was not done in the current study. Noises might therefore be so strong that they suppress the information about the identity of the subject in this current study.

The study Gregg et al. [4] reported, just as the findings in the current study, that sensitivity decreased with an increasing number of subjects. A reason for this could be that with a large number of subjects the intra-subject variability might be too high for the inter-subject variability creating a critical overlap in the distributions of features for each subject. This can decrease not only sensitivity but also specificity.

Training was made in this study on 294 subjects but included more data in the form of 15 leads instead of one lead used in the current study. Gregg et al. [4] reported a sensitivity of 37 % but also included 8369 subjects for test that were not trained on. Since Gregg et al. [4] tested the system by matching a trained subject's ECG with an untrained subject's ECG no match should be found, thereby increasing the sensitivity of the system without increasing the actual accuracy. In Gregg et al. [4] the entire ECG was used to determine if a match was found of a subject where only single heartbeats were used in this current study. By using the entire ECG instead of a single heartbeat, the subject who has been suggested most frequently for any of the heartbeats could be selected as the subject the ECG belonged to and therefore improve the results.

4.1. Application

The current study mimicked the application described in Agrafioti [13] but did have some differences. In Agrafioti [13] the application was biometric identification of subjects on ECG signals. In the current study only single beats were used instead of five seconds of ECG. Chaining the application in the current study to also contain a consecutive five seconds of ECG would allow HTM to learn more of the underlying patterns as the number of columns to handle the input data would increase. By introducing an overlap in the images it would also be possible to strengthen the temporal information in the images in HTM and thereby potentially increase the accuracy of the system. In this current study only 10 seconds ECG were available and a five seconds dataset would therefore result in a very limited size of a training and test set for each subject. This issue could have been overcome by having multiple either increased length of sessions or more sessions for each subject.

The application in the current study was adapted to also fit Casarella M. [8] who used unfiltered ECG. In a real world application there might still be some kind of filtering because of the issues with ECG quality [3, 25]. Good results using unfiltered ECG does however show a more robust solution.

4.2. Software

It was not possible to use the same software version as used in Casarella M. [8] even though attempts were made to contact the author. Changes in software can influence both function and stability of the software [19] and therefore it might have affected the outcome of this study. This did limit the possibilities of using and replicating the experience gained by Casarella M. [8].

4.3. Dataset

One lead was used in this study with two sessions where each session was 10 seconds long. In the cases of Casarella M. [8] and Agrafioti [13] they both used ECG recordings with a length more than 100 times greater than that of the recordings used in this current study.

Since Casarella M. [8] showed that an increased amount of data showed higher accuracy it might be worth investigating the effect of an increased data size for outcome.

Some of the ways to increase the data size could be to introduce more sessions since the data size for each session in this study is limited to 10 seconds. This will also give the possibility of gaining more complex temporal information from the ECG since only two sessions are available. Whether or not this would be feasible with more sessions can be discussed since it would require resources from healthcare professionals and patients to collect the extra amount of data.

Different leads might provide information different from each other about the heart [1, 7]. In HTM no additional knowledge about what the input represents (such as its diagnostic value) is required according to Casarella M. [8], meaning that even thought the exact value of these leads is not fully understood in a biometric application it would indeed be possible to use multiple leads to increase the dataset.

The dataset contained a broader spektrum of patients than both Agrafioti [13] and Casarella M. [8] since both unhealthy and healthy subjects were included in this current study. In Agrafioti [13] it was shown that the ability to perform the biometric application was affected by emotions. Whether or not accuracy of the system could also be affected in the same way but for health status of the subject is not known, but could be controlled for by using either only healthy subjects or for subjects of a specific disease.

4.3.1. **Subsampling of the dataset**

The subsets from the entire dataset were subtracted in a random manner. This does however also mean that a balance between male/female, age and different diagnosis were not known for these subsets. Introducing a pseudo randomisation could ensure an equal distribution of these. It is however not known how these demographics actually influence the accuracy of classification. The size of the subset did however have an influence of the accuracy of the system as an increased number of subjects decreased the accuracy of the system. Databases can however require even larger group sizes than used in this current study if comparisons were to be made for the entire database.

4.3.2. **Segmentation of heartbeats**

Pan Tompkins was used to segment heartbeats. This method has proven to be accurate and robust [18]. A specific width was selected even though some variations in the length of a heartbeat would be present due to factors such as the heart rate [8] thereby creating an overlap between heartbeats. The purpose of segmenting heartbeats might therefore be lost and a fixed length of time could be used instead.

4.3.3. **Encoding images**

The height of the image was chosen to follow Casarella M. [8]. Even though it matched with that literature, increasing the number of pixels would allow for more columns in the HTM layers thus increasing the capacity of learning for each layer.

4.3.4. **Training and test groups**

Recommendations from Numenta [19] and Casarella M. [8] were not directly followed for training to test data ratio as they had more data available, therefore the amount of training data was increased in this current study. In order to avoid over-fitting of the entire dataset and investigating the effect of different subsets, an optimisation and control dataset was included. This was done according to what is described by Holzinger [20] and Daume [21]. They argue that this is a robust method to overcome over-fitting issues when optimisation of algorithm has to be done.

The number of beats for testing were however very limited for each subject as only 2-3 beats were available. Since temporal information is stored as consecutive beats and only a few are available, the change from one subject to another might disrupt the ability to classify. Using one session as training and another one as test would increase the amount of data available for temporal information.

4.4. Configuration of algorithm

Only Casarella M. [8] was found to use HTM on ECG from the literature review. There were no further investigations of literature to find out how to optimise this HTM model other than reading through the software documentation of HTM. Important knowledge could therefore have been lost that could help optimise the HTM configuration.

Numenta [19] has created a new tool since the publication of Casarella M. [8] to help optimise the model, called swarming. This tool was however not used since it would move the methodology of this study further away from Casarella M. [8] but could potentially improve the results according to Numenta [19]. Casarella M. [8] did also notice that a correct configuration of the algorithm to the problem domain would significantly improve the results and a correct configuration was not necessarily found in this study even though attempts were made. It is also worth noticing that the HTM algorithm is not completely implemented in the software tool [19] used in this project.

4.4.1. **Number of layers**

The number of layers did follow the guidelines by Numenta [19] and Casarella M [8]. It is therefore not likely that another number would change the outcome of this current study.

4.4.2. **Configuration of layers**

Each of the layers were configured to follow what was found in Casarella M. [8]. These recommendations were made where more temporal information was available. Increasing the number of columns while still maintaining the pyramidal form, would increase the amount of spatial information that could be deducted from the images available.

4.4.3. **Capacity of learning**

It was selected that 2 % of columns should be active. This did follow all recommendations set forward by Casarella M. [8] and Numenta [19]. Due to the limited number of columns used in this current study however, this value did decrease the amount of temporal information that could be derived from the already limited amount of temporal information that was caused by only having a few test beats. Increasing this number might therefore in this case prove to be beneficial for the accuracy of the system. This could be tested with the new swarming methods by Numenta [19].

4.4.4. **Selecting number of cells**

The number of cells did not have a big effect on the amount of temporal information that could be stored due to the limited number of active columns selected. This number should therefore potentially be much higher as examples given by Numenta [19] normally include thousands of temporal patterns.

4.4.5. **Classifier and performance measure**

Other classifier types used than the one in the software for this project have shown better results on other datasets [26] and this might therefore limit results. Furthermore the number of classes (subjects) to differentiate between in this study was more than 100 times that in other studies compared under results. The amount of information about the underlying ECG grammar from the layers before the classifier might not be sufficient to differentiate between that number of subjects. It was found that increasing the number of subjects decrease the accuracy of the classifier.

The actual value of the probabilities of a sample belonging to a subject were not further investigated in the current study. With a high accuracy of the system these might carry more information beneficial for the understanding of underlying grammar of the ECG but were not used here due to the low accuracy when the higher number of classes were used.

4.4.6. **Experiments**

The number of iterations chosen were much lower and below the number of iterations tested in Casarella M. [8]. They show a somewhat sigmoid form of the relationship between number of iterations and classification accuracy thereby using just 1,000 iterations in the current study might have resulted in the same accuracy for Casarella M. [8]. In other words the number of iterations used in this current study might have been insufficient, but were however limited due to the time frame of the current study. Furthermore the number of experiments conducted on different iterations and subsets were not great enough to actually perform any statistics and thereby prove the effect of the subset size and number of iterations.

4.5. Technical issues

In this project a MongoDB server had to be created to distribute the memory load as the software had problems crashing midway throughout the calculations. This was due to the very large dataset above 30 GB. Furthermore it took roughly a full week for the computer to calculate results of just 100 iterations and above for the respective subsets. This massive requirement of hardware resources might limit the real life feasibility for usage of HTM especially in the initial training phase. After the training phase less than a minute was required to perform the actual classifications.

One solution could therefore be to apply online learning or train on just a specific subject with abnornamlity detection enabled [19]. Only training on a specific person and then detecting similiar ECGs from that person was done by Gregg et al. [4]. Here, time of training could be reduced since the amount of training data would be limited and only two classes would be needed: the specific subject and others (abnormal).

Another solution could be to create multiple HTM instances. Calculations could then be distributed and each HTM instance and in parallel detect and classify different underlying parts of the ECG that then can be used in a HTM instance that combine these results to classify the subject.

4.6. Future perspective

Finding the underlying ECG grammar, especially in an automated manner with machine learning, can be proven to be an essential tool in both biometrics but also in diagnosis. Casarella M. [8] points out that one of the benefits of HTM is that no prior knowledge is required about the underlying parameters of ECG and their origin in the leads. This means more leads could be included as dataset for HTM even though their representation of the ECG grammar is not yet known and that representation could be automatically investigated. Indeed this current study shows that even though no prior knowledge is known about the grammar representing biometric information from ECG lead II, some matches could still be found. Finding the underlying grammar would also help illuminate the limitation of ECG due to for example inter- and intra-subject variability within the grammar. This current study's findings show that increasing the number of subjects and thereby introducing more variability challenges the classifier as the accuracy drops. The study did however also find that new configurations and methods has been introduced to counter some of these issues as Numenta [19] introduced a tool called swarming for HTM. Finding the right configuration for HTM rather than completely discarding it as a tool due to the initial low accuracy might therefore be more feasible for the research of underlying grammar of ECG.

For ECG applications and especially the biometric application, rethinking the way that classifiers are trained when using large datasets might be beneficial for the accuracies of these systems due to inter- and intra-subject variablity. A greater amount of heartbeats can be used from the ECG and profiling can be made for values of the underlying patterns for each individual subject such as done in Chandrakar and Monisha [27]. In that study they managed to utilize the intra-subject variability as a tool to detect patterns for heart diseases specifically for each subject thereby improving the accuracy of their system.

5. CONCLUSION

In order to test if the underlying grammar of ECG could be found with HTM, another application than that originally tested in Casarella M. [8] was chosen. This application involved identifying a patient based on the heartbeats derived from the ECG. Limitations of HTM was evaluated with respect to number of subjects and number of iterations.

It was possible to use the underlying biometric grammar of the ECG to classify heartbeats with an accuracy up to 31.3 &. When the number of subjects increased the accuracy of the system decreased and eventually dropped below 1 %. These results persisted with different number of iterations.

The results of this study are however somewhat novel because of the limited dataset available for each of the subjects compared to other studies of ECG in a biometric application, and the high number of subjects compared to amount of data for the individual subject. Whether or not the results are due to the limitations of HTM or the inter-subject variability of ECG, would require a more similar dataset with other biometric applications.

Furthermore, the application of ECG in HTM for a biometric application is also novel so the right configuration for a biometric application of HTM is therefore not necessarily used in this project. It is worth investigating if a different configuration HTM could improve the results. Numenta [19] have developed a new tool called swarming that might assist in finding the correct configuration.

References

[1] A. Bayés Bayes de Luna, *Clinical Electrocardiography*. John Wiley & Sons, Incorporated, 2012.

[2] WHO, "Global status report on noncommunicable diseases 2014," *World Health*, p. 176, 2014.

[3] K. Khunti, "Accurate interpretation of the 12-lead ECG electrode placement: A systematic review," *Health Education Journal*, vol. 73, no. 5, pp. 610–623, 2013.

[4] R. E. Gregg, S. H. Zhou, and S. Babaeizadeh, "Finding Similar ECGs in a Large 12-lead ECG Database Introduction," vol. 43, 2016.

[5] F. H. Martini, J. L. Nath, and E. F. Bartholomew, *Fundamentals of Anatomy & Physiology*. Pearson, 2012.

[6] A. Zygmunt and J. Stanczyk, "Methods of evaluation of autonomic nervous system function," *Archives of Medical Science*, vol. 6, no. 1, pp. 11–18, 2010.

[7] S. Silbernagl and A. Despopoulos, *Color Atlas of Physiology*. 2009.

[8] J. Casarella M, *The Classification of Un-preprocessed ECG Waveforms through the Application of the Hierarchical Temporal Memory Model*. Doctor thesis, Pace University, 2012.

[9] M. Seera, C. P. Lim, W. S. Liew, E. Lim, and C. K. Loo, "Classification of electrocardiogram and auscultatory blood pressure signals using machine learning models," *Expert Systems with Applications*, vol. 42, no. 7, pp. 3643–3652, 2015.

[10] R. Dubois, P. Maison-Blanche, B. Quenet, and G. Dreyfus, "Automatic ECG wave extraction in long-term recordings using Gaussian mesa function models and nonlinear probability estimators," *Computer Methods and Programs in Biomedicine*, vol. 88, no. 3, pp. 217–233, 2007.

[11] Y.-C. Yeh, W.-J. Wang, and C. W. Chiou, "Feature selection algorithm for ECG signals using Range-Overlaps Method," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3499–3512, 2010.

[12] J. Oster, J. Behar, O. Sayadi, S. Nemati, A. E. W. Johnson, and G. D. Clifford, "Semisupervised ECG Ventricular Beat Classification with Novelty Detection Based on Switching Kalman Filters," 2015.

[13] F. Agrafioti, *ECG in Biometric Recognition : Time Dependency and Application Challenges by Foteini Agrafioti A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy Graduate Department of Electrical and Computer Engineering Univers*. Doctorate, University of Toronto, 2011.

[14] L. Lu, J. Yan, and C. W. de Silva, "Feature Selection for ECG Signal Processing Using Improved Genetic Algorithm and Empirical Mode Decomposition," *Measurement*, vol. 94, pp. 372–381, 2016.

[15] Z. Masetic and A. Subasi, "Congestive heart failure detection using random forest classifier," *Computer Methods and Programs in Biomedicine*, vol. 130, pp. 54–64, 2016.

[16] J. Hawkins, S. Ahmad, and D. Dubinsky, "Cortical Learning Algorithm and Hierarchical Temporal Memory," *Numenta Whitepaper*, pp. 1–68, 2011.

[17] M. Elgendi, "Fast QRS Detection with an Optimized Knowledge-Based Method: Evaluation on 11 Standard ECG Databases," *PLoS ONE*, vol. 8, no. 9, 2013.

[18] J. Pan and W. J. Tompkins, "A Real-Yime QRS Detection Algorithm," *IEEE transactions on bio-medical engineeringBiomedical Engineering*, vol. BME-32, no. 3, pp. 230–236, 1985.

[19] Numenta, "NuPIC 0.7.0.dev0 API Documentation," 2017.

[20] A. Holzinger, *Machine Learning for Health InformaticsState-of-the-Art and Future Challenges*. Springer, 2016.

[21] H. Daume, *A course in machine learning Ch3*. Self Published, 2012.

[22] F. Sufi, A. Mahmood, and I. Khalil, "Diagnosis of cardiovascular abnormalities from compressed ECG: A data mining based approach," *2009 9th International Conference on Information Technology and Applications in Biomedicine*, vol. 15, no. 1, pp. 1–4, 2009.

[23] M. M. A. Rahhal, Y. Bazi, H. Alhichri, N. Alajlan, F. Melgani, and R. R. Yager, "Deep learning approach for active classification of electrocardiogram signals," *Information Sciences*, vol. 345, pp. 340–354, 2016.

[24] C. Chandrakar and M. Sharma, "A real time approach for classification of ECG beats using repetition-based pattern detection and cardiac profiling scheme," *International Conference on Computing, Communication and Security, ICCCS 2015*, 2015.

[25] N. Kalkstein, Y. Kinar, M. Na'aman, N. Neumark, and P. Akiva, "Using machine learning to detect problems in ECG data collection," *2011 Computing in Cardiology (CinC)*, pp. 437–440, 2011.

[26] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," *Proceedings of the 23th International Conference on Machine Learning*, pp. 161–168, 2006.

[27] C. Chandrakar and S. Monisha, "Qualitative features selection techniques by profiling statistical features of ECG," *EEE India Conference (INDICON)*, 2015.

WORKSHEETS

Preface

The present report has been produced on the forth semester master of Biomedical Engineering and Informatics on Aalborg University during a semester abroad at the Centre for Chiropractic Research at New Zealand College of Chiropractic. The supervisors associated with the present project were Johannes Jan Struijk from the Department of Health Science and Technology of Aalborg University and Rasmus Nedergaard from Centre for Chiropractic Research at the New Zealand College of Chiropractic. The purpose of the project was to perform machine learning on health record data. This projected is made in collaboration with the New Zealand College of Chiropractic. They want to take a novel approach to patient management and record systems by developing their own system with research usage orientation. Further more they seek to combine their electronic system with raw data samples from the patients such as electrocardiogram (ECG) to create a big data source for researching. Their goal is for this system to be able to collect and contain a big interconnected data pool that can be used in data mining and thereby improved the quality of care and data-driven research in the field of chiropractic. Since the type of data source used in this project is similar as of the system, this project member has therefore assisted in the development of the system alongside with working with machine learning.

Reading instruction

The Harvard method was used for references in this report. If the reference is placed before the dot it is referring to the sentence, if it is placed after the dot it is referring to the section. If the source has multiple authors the surname of the first author is written followed by *et al.* - an example: [Kelly et al., 2012]. Figures such as Figure 1.2 and tables without any references in the caption are self-made. Any abbreviations used in the report will be written to its full extend followed by the abbreviation the first time mentioned and included in headlines - an example: Machine learning (MS).

Table of contents

Chapter 1

Background

In this section, the problem, its domain, and origin in the surface electrocardiogram (ECG) is outlined. A literature review is conducted and a problem statement is established from that review. This section also covers a brief introduction to the data source of this project.

1.1 Introduction

Cardiovascular disease is the leading cause of mortality and illness in the western world [Bayés Bayes de Luna, 2012] with over 17.5 million deaths worldwide. Approximately 80 % of these cases are due to ischaemic heart disease and stroke with the majority of cases being preventable. [WHO, 2014]. Heart attacks (ischemic heart disease) alone resolved in more than 7.4 million deaths [WHO, 2014].

In order to detect these cardiovascular diseases different methodologies have been developed such as coronarography, contrast-enhanced cardiac magnetic resonance, ECG [Bayés Bayes de Luna, 2012], Doppler echocardiography [Strasburger et al., 1986] and seismocardiography [Wick et al., 2015].

The 12-lead ECG is the golden standard for diagnosis of coronary heart disease patients [Khunti, 2013] and it is also the most widespread used to diagnose suspected heart disease [Bayés Bayes de Luna, 2012].

There is a physiological reason for the usage of ECG in diagnostics. Since ECG is the readout of the electrical activity initiating a contraction of the heart, an electrical malfunction will change the recorded signal [Gregg et al., 2016, Bayés Bayes de Luna, 2012]. The heart and its electrical activity is modulated by both intrinsic and extrinsic factors [Martini et al., 2012, Bayés Bayes de Luna, 2012] which means that not only heart problems are detectable but also problems occurring extrinsically from the autonomic nervous system (ANS) would be detectable. The ANS also regulates internal organs other than the heart [Zygmunt and Stanczyk, 2010] thereby the 12-lead ECG could potentially be used for the assessment of a more general health status. One limitation of the technique is that it doesn't indicate the actual mechanical contraction and pump function of the heart [Silbernagl and Despopoulos, 2009].

Since ECG features indicate the heart's electrical activity [Silbernagl and Despopoulos, 2009] it can be used by a doctor to perform further investigations or diagnosis. [Bayés Bayes de Luna, 2012] Some of the useful features that can be derived from the ECG are for example Heart Rate Variability (HRV), QT interval (to diagnose Long QT and Short QT syndromes) and QRS width (to diagnose e.g., bundle branch block and ectopic beats) [Bayés Bayes de Luna, 2012].

Detecting and using features derived from the 12-lead ECG in a form of automation of diagnosis with machine learning on a large data set has been suggested to show potential by Gregg et al. [2016], Eftestol et al. [1998], Bayés Bayes de Luna [2012]. In fact one of the biggest limitations with ECG was suggested by Bayés Bayes de Luna [2012] to be the lack of expertise by the physician doing a manual inspection and diagnosis. The idea of using medical data from a subject is further supported by Jacobson and Dalianis [2016], Holzinger [2016], Daume [2012], Eftestol et al. [1998] who outline some of the machine learning that can be applied to different kinds of data.

Since the field of using ECG in machine learning shows great potential, the state of the art of that procedure along with its limitations was investigated in a literature review with the initial problem being *how to investigate the ECG with machine learning?*

1.2 Data source

In order to better understand the data source and its limitation, the following two subsections have been established to understand the heart's conduction system and theory about ECG and features in it.

1.2.1 The heart's conduction system

In order for the heart to pump blood efficiently in the body, a coordinated contraction of the tissue is important [Martini et al., 2012, Balakrishnan et al., 2015] Failure results in cardiovascular problems [Balakrishnan et al., 2015]. This coordination is autorhythmic since it is controlled by the Sinuatrial node and Atrioventricular node. They do not require any input to initiate a contraction of the heart muscle. [Berdajs and I. Tuna, 2011, Bayés Bayes de Luna, 2012, Christoffels and Moorman, 2009] The frequency of these spontaneous activities in the nodes are 80-100 bpm and 40-60 bpm for the Sinuatrial node and Atrioventricular node respectively [Martini et al., 2012] . The last backup, if any of the nodes is failing, is the Bundle of His [Martini et al., 2012] that makes sure that at the end of the Purkinje fibers the firing rate would be approximately 20 bpm [Balakrishnan et al., 2015].

The conducting pathway is illustrated in figure 1.1 and the innervation of the different parts of the ANS on the system is noted in the figure text. One important pathway of conduction that is not mentioned in the figure is the Intercalated discs. These Intercalated discs are gap junctions in the sarcolemma of adjacent cells allowing the electrical signal to propagate between muscle cells without direct neural innervation. This is one of the main differences between cardiac muscle tissue and other muscle tissue. [Martini et al., 2012]

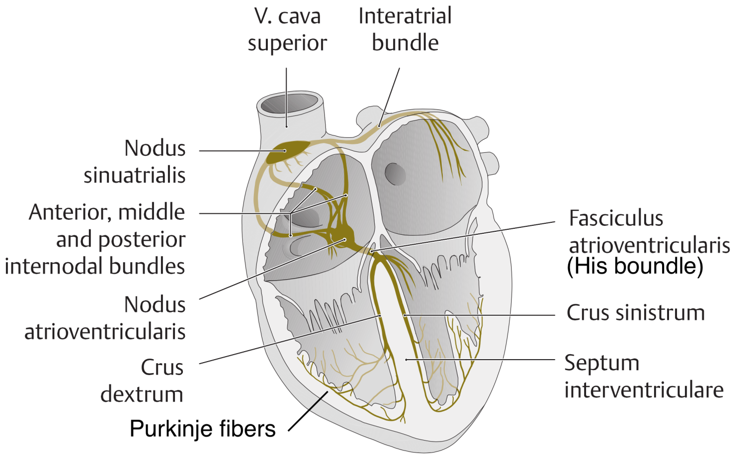

Figure 1.1: The conducting system of the heart with some labels of the different elements in the conducting system. Used and modified from [Gilroy et al., 2008]. Parasympathetic and sympathetic ANS innervates the Sinuatrial node, Atrioventricular node along with atrial and ventricular muscle cells [Gilroy et al., 2008, Martini et al., 2012]

The conducting system of the heart is under constant regulation from the ANS of either sympathetic or parasympathetic activity which leads to an increase or decrease in cardiac output. The sympathetic system is for fight-or-flight responses where the parasympathetic activity is for rest-and-digest responses. The base regulation is called the autonomic tone and is the resulting cardiac output is a sum of not just ANS activity but also hormones epinephrine and norepinephrine, drugs and abnormalities in the homeostasis such as increased or lowered temperature. [Martini et al., 2012, Silbernagl and Despopoulos, 2009]

The cardiac output is a function of the heart rate times the stroke volume [Martini et al., 2012]. Only heart rate is directly affected by the ANS where as the stroke volume is the difference between how much blood is in the ventricle prior to and after a contraction [Martini et al., 2012]. The ANS does however regulate how much blood is left in the ventricles after their systole, this is called inotropism and works with a direct regulation of the muscle cell metabolism controlling the force of the contraction of the muscle cells [Martini et al., 2012, Silbernagl and Despopoulos, 2009]

Furthermore, stroke volume is also indirectly affected by ANS as an increase in heart rate decreases the time available for filling, thereby decreasing the duration of the diastole. [Martini et al., 2012].

Other factors affecting the stroke volume are venous return and afterload where an increased return will cause a stretch reflex to produce a more forceful contraction. An increased afterload is an increased force that the heart has to work against to pump out blood and that will ultimately lower the stroke volume. [Martini et al., 2012, Silbernagl and Despopoulos, 2009]

The alteration of heart rate is called chronotropism and it is managed via a change of the rate of spontaneous depolarization and repolarisation in the Sinuatrial node and Atrioventricular node along with a change of velocity in signal conduction (dromotropism). [Silbernagl and Despopoulos, 2009, Martini et al., 2012] The change in conduction velocity is mainly for the internodal pathway to the Atrioventricular node and through it. [Silbernagl and Despopoulos, 2009]

1.2.2 Electrocardiography

The ECG is the recording of the electrical activity of the heart [Bayés Bayes de Luna, 2012, Silbernagl and Despopoulos, 2009] in the frequency spectrum from 0.05 Hz to at least 250 Hz for children and 150 Hz for adults [Bayés Bayes de Luna, 2012]. The recorded signal is however not without noise as surrounding muscle activity and breathing is also recorded. Muscle activity is in the frequency range of roughly 5 Hz to 500 Hz [Van Boxtel, 2001, de Luca, 1997] suggesting that the signal should be sampled in higher frequencies to avoid aliasing. This baseline wandering or artifact due to breathing has been thought to be below 1 Hz [Shin et al., 2015], but just removing that frequency component has been proven to be insufficient [He et al., 2001, Shin et al., 2015].

When the electrical activity is recorded 10 electrodes are placed in well defined locations on the chest and arms of the subject [Bayés Bayes de Luna, 2012]. However as pointed out by Khunti [2013] the placement is not necessarily the same each time it is placed as 50 % of nurses and more than 20 % of cardiologists misplace some of the electrodes.

The placement of these electrodes are divided into chest and limb leads making it possible to track the summation vector of the electrical activity of the heart in different plans and angles over time. Electrical activity perpendicular to a lead cannot be detected in that lead. [Bayés Bayes de Luna, 2012, Silbernagl and Despopoulos, 2009]

Multiple lead systems exist within the 12-lead ECG. These are theorized by Einthoven, Goldberger and Wilson where the leads from Goldberger are derived from Einthoven. [Bayés Bayes de Luna, 2012, Silbernagl and Despopoulos, 2009, Kramme et al., 2011] The leads and placements are illustrated in the following figure 1.2 and figure 1.3.

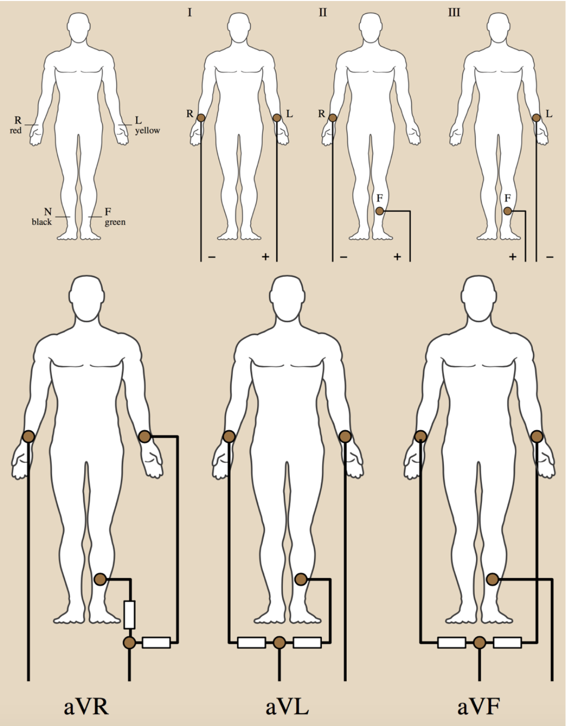

Figure 1.2: The Einthoven and Goldberger leads. The Einthoven leads I, II and III are bipolar leads with the reference to the central nerminal. The Goldberger leads are unipolar leads meaning that the potential are measured between an electrode and the reference consisting of the other two electrodes. [Bayés Bayes de Luna, 2012, Silbernagl and Despopoulos, 2009, Kramme et al., 2011] Modified from Kramme et al. [2011]

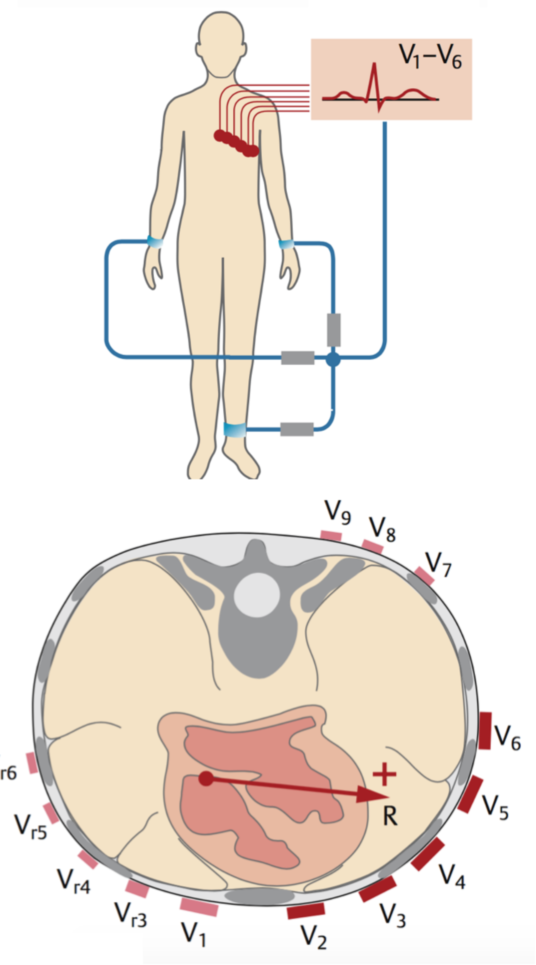

Figure 1.3: The leads as defined by Wilson. More deep coloured electrodes on the transverse plan are usually placed where the others are optional. The recordings performed are unipolar so the three limb electrodes are combined into a common reference and used with one of the V electrodes. [Bayés Bayes de Luna, 2012, Silbernagl and Despopoulos, 2009, Kramme et al., 2011] Modified from Silbernagl and Despopoulos [2009]

The lead to chose to investigate depends on what it is to be investigated since the strength of the electrical activity is reflected by the angle in which it is recorded and the different leads represent different angles. The following figure 1.4 represents different leads and the angle of the lead with respect to the heart. In the figure 1.4 the chest leads are not included. By using the chest lead different horizontal perspectives of the heart's electrical activity can be recorded and a 3D view of the heart also called the vector cardiogram (VCG) can be established [Silbernagl and Despopoulos, 2009, Bayés Bayes de Luna, 2012]. This VCG is however not as well studied as the ECG [Silbernagl and Despopoulos, 2009, Bayés Bayes de Luna, 2012].

The lead that has the biggest QRS complex is defined as the electrical axis and can itself have diagnostic value [Silbernagl and Despopoulos, 2009, Bayés Bayes de Luna, 2012]. The normal angle is between -30 to 120 [Silbernagl and Despopoulos, 2009].

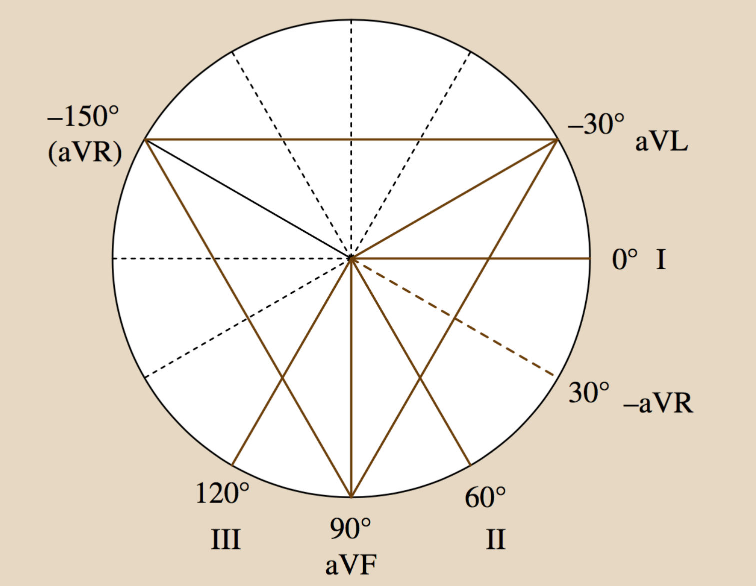

Figure 1.4: This figure is called Cabera's circle. It illustrates the different leads and their respective angle of view on the heart. Modified from [Kramme et al., 2011]

The recorded ECG contains what is known as a P-wave QRS complex and a T-wave [Bayés Bayes de Luna, 2012, Gilroy et al., 2008] as illustrated in figure 1.5. This recorded signal can be mapped to different stages of the contraction cycle of the heart as illustrated in figure 1.6. The initial contraction begins with a stimulation from the Sinuatrial node. From myocardial cell to myocardial cell the stimulation signal propagates along with inter-nodal connections creating the P-wave and resulting in a contraction of the atria. The cell to cell propagation stops at the cardiac skeleton. The signal continues through the Atrioventricular node and Purkinje fibers, but with a minor delay into the Arioventricular node because of the small size of nodal cells. This delay is important since it allows the atria to fully contract before contraction of the ventricles. The QRS complex is generated by the depolarization of the ventricles. The subsequent T-wave is generated by the repolarization of the ventricular muscle cells. [Silbernagl and Despopoulos, 2009, Martini et al., 2012]

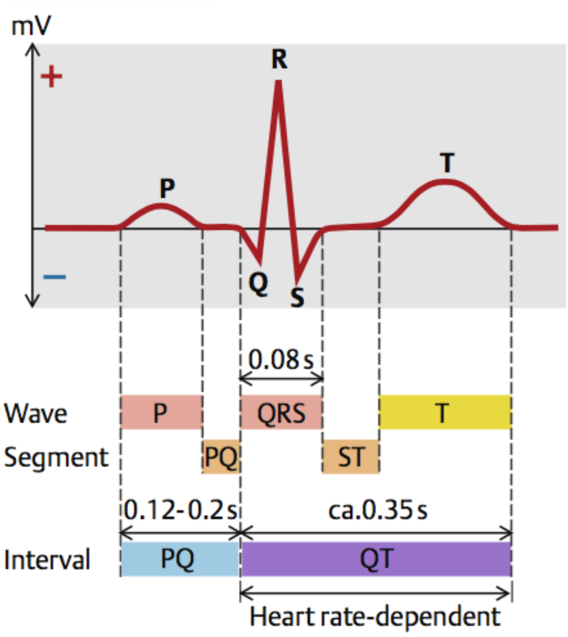

Figure 1.5: Example of one cycle ECG it is full length with P-wave QRS complex and T-wave. Image shows that the QT complex is varied by heart rate where the PQ interval remains somewhat unaffected. Modified from Silbernagl and Despopoulos [2009]

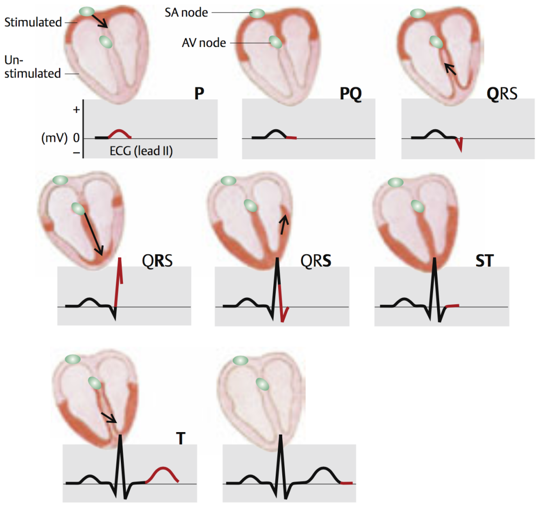

Figure 1.6: From top left to right bottom the order of the propagation of the polarisation and depolarization in the conducting system of the heart mapped to second lead ECG. SA node and AV is short for the Sinuatrial node and Atrioventricular node. Modified from Silbernagl and Despopoulos [2009]

1.3 Literature review

In order to investigate ECG and machine learning keywords were chosen. These keywords were based upon words related to machine learning and electrocardiography and its abbreviation ECG. These words were: artificial intelligence, recognition, data mining, memory, deep learning, algorithm. These are derived from Buchanan et al. [2017]. The search was carried out on aub.aau.dk since this tool supports multiple databases in the search.

1.3.1 Inclusion criteria and search strategy

The goal of this literature review was to find the most relevant information and to establish the state of the art and its limitation. An initial search was performed on the database for state of the art literature for the last 10 years, meaning from 2007. This search was further limited before any reading was done. The 10 years was chosen as an arbitrary number to be big enough to at least show the state of the art in the field. It was done in this order to minimize exclusion of relevant material. The limitation was done by adding extra keywords and filters to the search if the results showed more than 80 results. One of the filters were that the keywords should be in either title, abstract and or tags specified in the literature. The abstracts were then read and, based on the relevance of this project, they were excluded or included. The relevance was based upon redundancy in literature already found or whether the topic actually was machine learning or not. The full texts of the accepted articles were then read and their citations were reviewed and read if they also had relevance. The whole process and number of found articles is illustrated in figure 1.7. The keywords "ECG or electrocardiography" were always used and they alone gave 462.176 hits.

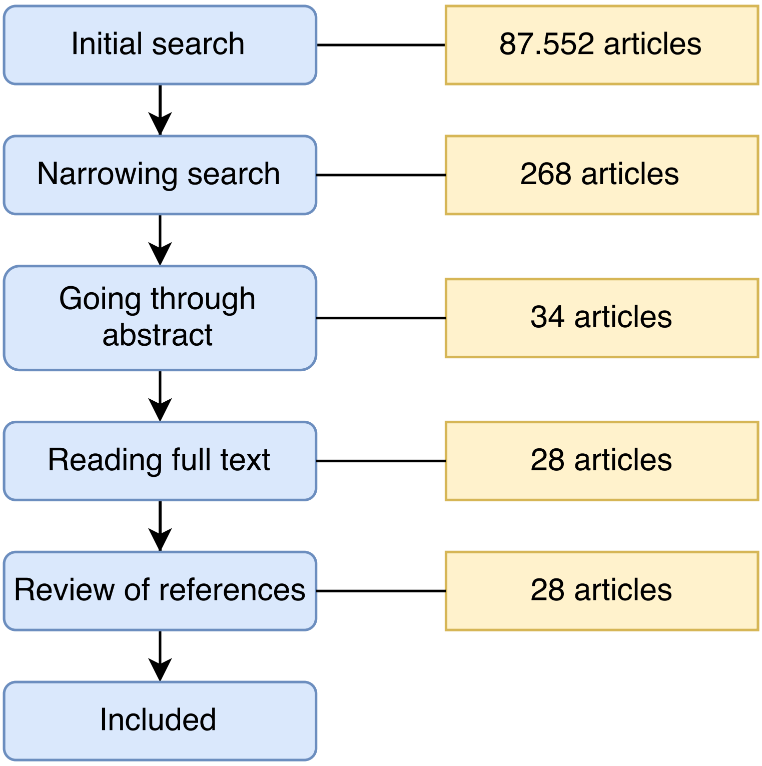

Figure 1.7: The steps of the literature review, from a broad search to a narrowed result

1.3.2 Literature included

The following table 1.1 shows the keywords used and the result from each of the keywords in every step of the search. The articles found by each keywords are hereafter presented with respect to what they investigated, machine learning methodologies, ECG source and findings.

Table 1.1: Keywords and their corresponding hits. The initial column hits are the unfiltered response from the initial searches. All the search were made with the requirements of keyword ECG or electrocardiography to be present as well. For filtering *′or′* means that the title, abstract or found material keywords should include the keyword used in the search. *′and′* means that title and abstract should include the keyword used in the search and *′all′* means that keywords, abstract and found material keywords should include the keyword used in the search

Keywords	Hits	Filtered	Abstract	Full text
Machine learning	4.240	(or) 239 (and) 79	13	9
Deep learning	144	(or) 1	1	1
Algorithm	36.267	(or) 3828 (and) 823 (all) 47	7	7
Recognition	25.204	(or) 1257 (and) 354 (all) 12	4	4
Artificial intelligence or AI	8	Not performed	0	0
Data mining	3.004	(or) 131 (and) 50	4	2
Memory	18.685	(or) 454 (and) 79	5	4

From the literature tables were created for each of the keywords that showed any results. The tables contain areas describing what was investigated, machine learning methods, the ECG data source and Results. The content of the text and table do not completely overlap but reading one or the other would be sufficient for the understanding of the text in the next section 1.4, Problem statement.

Findings for memory

The following table 1.2 is for key word “memory”.

Table 1.2: Literature found for the keyword “memory” along with a description of what was investigated, the machine learning methods used, the ECG source and findings.Aberrations used in this table: normal (NORM) right bundle branch block (RBBB), left bundle branch block (LBBB), premature ventricular contractions (PVC), atrial premature contraction (APC), paced beats (PB), ventricular couples (VC). Two papers are merged into one row since it is the same study

Literature	Investigated	Machine learning methods	ECG source	Findings
[Casarella M, 2012]	Diagnosis of ECG into the categories NORM, RBBB, LBBB on different kind and levels of noise	Hierarchical Temporal Memory (HTM)	Unprocessed ECG second lead beats was converted into images	99.8 % accuracy of classifier. Robust to noise that was not occlusion since accuracy of classification remained the same
[Chauhan and Vig, 2015]	Diagnosis of ECG into the categories NORM, PVC, APC, PB, VC	Long short-term memory	1 Minute ECG second lead beats recordings	precision for classification found for each class were: 0.96 VC, 0.99 PVC, 0.92 APC and 0.99 PB
[Chandrakar and Monisha, 2015, Chandrakar and Sharma, 2015]	Diagnosis of ECG into the categories NORM or abnormal with abnormal being LBBB, RBBB, VPC or APC	Network Intrusion Detection to perform profiling	ECG features extracted from PQRST	accuracy for classification 99.57 % correct classification. Abnormal values are not the same of each subject
[Rahhal et al., 2016]	Diagnosis of ECG into the categories normal, ventricular, supraventricular, fusion of normal and ventricular and unknown beats	Autoencoder and softmax for optimisation of features deep neural network (supervised)	Baseline corrected ECG beats with information about RR interval	100 % for both specificity and sensitivity of classification. They concluded than a deep neural network with more than two layers did not contribute to classification

In the thesis by [Casarella M, 2012] multiple algorithms were investigated, and it was concluded to use Hierarchical Temporal Memory Model (HTM) as the author argues and cite Gelernter [1991] for other algorithms to lack the ability of a by-the-model learned invariant pool rather than a pool hard coded by the investigator. The invariant pool is the pool of features which are invariant to for example scaling, shifting or rotation.

The novel approach to the ECG data set by [Casarella M, 2012] is to mimic the doctors so the ECG beats were given as an binary image rather than a set of numeric features are extracted from the ECG based on lead II. In this way no prior knowledge about ECG morphology and disease is needed. The amount of training data increased the classification percentage with a successful classification of up to close to a 99.8 %. The categories in which the beats were classified in were: NORM, RBBB or LBBB. Different levels of noise and different kind of noise were added to the image such as occlusion and salt and pepper noise. The occlusion noise was square boxes covering parts of the image.

In the article [Chauhan and Vig, 2015] the algorithm long-short term memory (LSTM) was used. It is a recurrent neural network. This algorithm was chosen since it limits the amount of preprocessing needed for the data, does not require features to be hard coded as these are derived of the algorithm itself and it is good at overcoming the vanishing gradient problem that can slow the training of a network. One minute ECG signals was used and classified to the following classes: NORM, PVC, APC, PB and VC. The classification was done for the individual beats in the individual classes with respect to a normal beat. The precision for classification found for each class was: 0.96 VC, 0.99 PVC, 0.92 APC and 0.99 PB.

The two publications by the same first author [Chandrakar and Monisha, 2015, Chandrakar and Sharma, 2015] investigated features and specify that often systems use these qualitative features for classification of ECG and a method for selecting a normal value span of these features is therefore needed. They investigate a method to identify this span so the features can be used in detecting the difference between normal and abnormal heart beats. Abnormal heartbeats were identified to be either: LBBB, RBBB, VPC or APC. They created a custom three stage repetition of pattern algorithm based on the Network Intrusion Detection field that classified ECG based on those features. They conclude that it is not possible to extract a normal feature span from the ECG that will cover all patient so profiling is needed when the ECG waveform is analysed. Using their suggested algorithm they achieved 99.57 % correct classification of the ECG.

In the article [Rahhal et al., 2016] the optimal feature representation to classify the ECG within the standard of Association for the Advancement of Medical Instrumentation (AAMI) was investigated. The use a combination of supervised active learning and unsupervised learning with autoencoder and softmax for feature learning and a deep neural network for classification according to the standards of the AAMI. Entropy is calculated for each beat in five minute ECG recordings and the ones with the highest entropy are given to an expert for labelling. Breaking tie beats that were not able to be correctly classified is also feed to an expert for labelling. The deep neural network is then trained with the new training set. This is done until successful classification converges. The number of features and labelled training data used increased the specificity and sensitivity with close to a 100 % for both. They tested and concluded that adding more layers than two to their deep network did not improve their results.

Findings for data mining

Only two articles were found for “data mining”. They are displayed in table 1.3.

Table 1.3: Literature found for the keyword “data mining” along with a description of what was investigated, the machine learning methods used, the ECG source and findings

Paper	Investigated	Machine learning methods	ECG source	Findings
[Tyagi and Thakur, 2015]	Different data mining techniques for data discovery	Techniques categorized under classification, regression and clustering	Not described	Classification is a better technique
[Sufi et al., 2009]	Compressed ECG for diagnosis of normal, premature ventricular contraction, atrial fibrillation, atrial premature beat	Expectation maximum clustering	unfiltered QRST features on individual heart beats	97 % accuracy

[Tyagi and Thakur, 2015] did a literature review of current data mining techniques on ECG and concluded from the different data mining methods (classification, clustering, prediction and association) the classification method was the best to analyse ECG data with. They did not specify a certain algorithm to be superior. The same authors have in an earlier article Sufi et al. [2009] and investigated clustering with 97 % accuracy. Their main focus in that article was however to investigate compressed ECG and abnormal vs. normal ECG beats so that a patient could seek treatment if something was wrong.

Findings for recognition

The literature that was found for "recognition" is displayed in following table 1.4.

Table 1.4: Literature found for the keyword "recognition" along with a description of what was investigated, the machine learning methods used, the ECG source and findings

Literature	Investigated	Machine learning methods	ECG source	Findings
[Xu et al., 2010]	Classification of emotions Joy, sadness and a combination of both	Binary Particle Swarm Optimization algorithm and K-Nearest Neighbours	Features of ECG based upon RR interval and PQRST	Accuracy of classification 86 %
[Khandoker et al., 2009]	Detection of the severity of obstructive sleep disorder	Support vector machine	Wavelet decomposition parameters from ECG beats, RR intervals	97.59 % accuracy was obtained using only 4 features. Using the full feature set of 28 features gave a trade off between true positive and false positive where these were close the same value
[Dong et al., 2014]	Temporal features of the ECG to classify between normal and myocardial infraction	Radial basis neural network	10 second ECG with baseline correction	Specificity of 78.8 % and sensitivity of 92.5 %
[Agrafioti, 2011]	Biometric application of ECG and correcting for emotion	Methods related to classification	Features from empirical mode decomposition of ECG signal	By classifying to a smaller group of people the error rate went down to 10 %

[Xu et al., 2010] differs from other found literature in that sense that it investigate emotions rather than diagnosis of diseases. They do however also investigate the selection of an optimal feature set for this task. They do this with with the Binary Particle Swarm Optimization algorithm and K-Nearest Neighbours. They successfully classified between the emotions joy, sadness and a combination with an average accuracy of 86 %.

As aforementioned in 1.1, Introduction ,not just the health of the heart can be investigated. In the case of Khandoker et al. [2009] obstructive sleep apnea syndrome was detected which is an obstruction of the airways in ones sleep. This is not a direct heart fault but can result as such. The authors of the literature describe that studies have already been made using heart rate variability derived from a wavelet transform and ECG-derived respiration to detect the problem although but not the severity of the problem. In order to do so the authors tried to investigate an optimal feature set and used a support vector machine on the features. They showed a Receiver operating characteristic (ROC) plot that illustrated a trade off between true positive and falls positive of the severity of disease. True positive rates would increase with falls positive rates with the values being about the same at any point in the plot.

As mentioned by Casarella M [2012] deterministic, statistic and synaptic learning exists and Dong et al. [2014] investigate deterministic learning. Deterministic decisions does not include information about the underlying probably destribution of a sample to be within different classes when a class is assigned to a sample. Statisticial methods are probebalistic based and synapitc based is neural networks. Casarella M [2012] In Dong et al. [2014] investigations of dynamic temporal features were made by including the entire duration of the ECG recording. This is investigated because of the problems with between and intra subject variations over time in the ECG [Dong et al., 2014]. They used a radial basis neural network to model the dynamics of the ECG. They compared healthy and myocardial infraction affected ECG. They managed to perform with a specificity of 78.8 % and sensitivity of 92.5 %.

In the doctoral thesis Agrafioti [2011] the biometric recognition of continous ECG signal is explored. This is not directly linked to medical diagnosis of ECG but investigates the challenges of temporal features of the ECG. Some of these mentioned are physical activity, noise and emotional activites as investigated also in Xu et al. [2010]. The physical stress can be accounted for [Agrafioti, 2011], but as investigated by Xu et al. [2010] the emotional stress cannot be done with high accuracy. Agrafioti [2011] use empirical mode decomposition to detect emotional patterns. Furthermore state features are subtracted from the ECG for state classification of the ECG beats to detect stages where biometric detection is not possible. The ECG state features are subtracted with a Autocorrelation-Linear Discriminant analysis and incorporates quality assessment of the ECG based on periodicity transform. The author concluded that a generalisation of their algorithm to a larger group of people was not feasible with error rates up to 45.5 %. By classifying to a smaller group of people the error rate went down to 10 %. They argue that the higher error rate is due to training on a general population rather than on each individual. They where however successful in classifying the emotional states of with an accuracy of 96.47 %.

Findings for keywords machine learning

The literature found for “machine learning” is split into two tables because of its size. The first found literature is in the following table 1.5.

Table 1.5: Literature found for the keyword "machine learning" along with a description of what was investigated, the machine learning methods used, the ECG source and findings. Abbreviations: Association of advancement of medical instrumentation (AAMI), normal (NORM) right bundle branch block (RBBB), left bundle branch block (LBBB), premature ventricular contractions (PVC), atrial premature contraction (APC)

Literature	Investigated	Machine learning methods	ECG source	Findings
[Kora and Sri Rama Krishna, 2016]	Optimisation of feature set and wavelet coherence on ECG beats	Levenberg Marquardt neural network	ECG beats baseline-corrected features from wavelet transform	Sensitivity of 96.97 % and specificity of 99.43 %
[Shadmand and Mashoufi, 2016]	Diagnosis according to AAMI with an individual tailored classifier to fit each patient	Block-based neural network and particle swarm	Temporal features and Hermit transfer features extracted from ECG beats	Accuracy of 97 % they highlight the importance of the patient specific classifier
[Kumari and Sadasivam, 2007]	Compression of ECG using Wavelet transform and compare it decompressed to the original source	Not given	Wavelet transformed ECG compared to non transformed	No data of diagnostic value lost in compression
[Balouchestani and Krishnan, 2016]	Rapid search in large ECG data sets with biomakers. Classification of the data set was done for classes normal, supra-ventricular problems, ventricular, fusion beat, paced beats and unknown	K-means algorithm and K-singular value decomposition for calcification and feature reduction	Probabilistic neural network	Accuracy of 99.98 %
[Yeh et al., 2010]	Selecting optimised features for diagnosis of NORM, LBBB, RBBB, VPC and APC	Range-Overlaps Method (ROM), clustering of fuzzy logic methods	Baseline corrected beats	Accuracy of roughly 93 %
[Lu et al., 2016]	Optimisation of feature set for diagnosis of NORM, LBBB, RBBB, APC and PVC	Support vector machine	empirical mode decomposition features of ECG beats	The accuracy for classification was NORM 96.25 %, LBBB 99.06 %, RBBB 95 %, APC 92.81 % and PVC 97.5 % respectively
[Elbuni et al., 2009]	More accurate estimation of ECG features	discrete wavelet transform	ECG beats baseline and noise reduced	sensitivity and specificity was 98 % and 96 % percent respectively

The article Kora and Sri Rama Krishna [2016] split abnormality detection in the ECG up into three steps: pre-processing, feature extraction and classification. They state that conventional time domain features are used for detection of abnormalities but suggest the usage of a wavelet coherence technique (WTC) on the ECG beats instead. They use a Bat algorithm to optimise features from the WTC and a Levenberg Marquardt neural network to classify atrial fibrillation. They compare the WTC with the often used auto regressive feature extraction [Kora and Sri Rama Krishna, 2016]. The results show a sensitivity of 96.97 % and specificity of 99.43 % where the auto regressive performs 92.5 % and 89.7 % respectively.

Just as investigated in Rahhal et al. [2016] the article Shadmand and Mashoufi [2016] investigate the classification of ECG heart beats according to AAMI standards. They use a block-based neural network (BBNN). They optimise the network with a Particle Swarm algorithm. The input to the BBNN is temporal features and Hermit transfer features extracted from ECG. Results of this approach showed an accuracy of 97 %. They argue that one of their successes is the profiling of a BBNN to each individual subject.

Kumari and Sadasivam [2007] did not investigate diagnostics but instead used Wavelet transform to extract features and compress the ECG. They conclude that all clinical information is still present in the compressed ECG making it ideal for portable heart monitoring systems with what they state as a low root mean square difference between decompressed signal compared to the original signal.

Going through large data sets of long-term ECG recording with clustering algorithms is a common source for detecting heart problems [Balouchestani and Krishnan, 2016]. Optimisation of this process was investigated by Balouchestani and Krishnan [2016] who wanted to investigate biomarkers (features) in the ECG with ease and rapidly. They used an advanced K-means algorithm and K-singular value decomposition. In order to validate the algorithm principal component analysis was used along with other dimensionality reduction methods. The data from these reductions were used in a probabilistic neural network. The proposed method was able to classify with an accuracy of 99.98 % thereby outperforming other clustering methods. The method also outperforms other methods in the sense of speed as it requires 13 % less computation. Classes were normal, supra-ventricular problems, ventricular, fusion beat, paced beats and unknown.

Yeh et al. [2010] investigate feature selection and especially selecting the optimal feature set from the ECG beats (QT length etc.) as to improve accuracy of classification. They investigate a fast, reliable method they name the Range-Overlaps Method (ROM) in order to select ideal feature sets and improve classification on mobile devices. They were able to classify with an accuracy of roughly 93 % using a clustering or fuzzy logic method by just using four features from the ECG heart beat. They classify for normal, LBBB, RBBB, VPC and APC.

Feature selection on ECG was performed by Lu et al. [2016] to select an optimal feature set. This was done with a genetic algorithm and an empirical mode decomposition (EMD) where the resulting features were fed to a support vector machine and classified. They argue that EMD is better than wavelet transform since it not requires prior knowledge of the ECG. Initially from the EMD output the generic algorithm used a variable-range encoding. Fisher criterion was then used to select the most dominant features. The classification was based on the classes normal, LBBB, RBBB, APC and PVC. The accuracy for classification was 96.25 %, 99.06 %, 95 %, 92.81 % and 97.5 % respectively. They compare their findings with other literature and conclude that they can use fewer features but perform the same as others. They also conclude that the most dominant features are in the first and sixth intrinsic mode function decomposed by the EMD and the fifth moment of these.

Elbuni et al. [2009] uses discrete wavelet transform to prepare the ECG signal for further use in statistical analysis. This was done by using the transform to remove baseline, noise and detect the QRS complex. They proposed their method so that that features could be more accurately estimated from the ECG. The sensitivity and specificity was 98 % and 96 % percent respectively.

The rest of the literature found for “machine learning” is given in the table 1.6.

Table 1.6: Literature found for the keyword "machine learning" along with a description of what was investigated, the machine learning methods used, the ECG source and findings

Literature	Investigated	Machine learning methods	ECG source	Findings
[Seera et al., 2015]	Evaluation of classifiers of types neural network, statistical, decision trees and others	Multiple different methods	ECG beats baseline and noise reduced	Most common used was neural networks but random forest had the highest accuracy
[Masetic and Subasi, 2016]	Diagnosis of congestive heart failure with multiple algorithms	Random forest method, Neural networks and others	Autoregressive burg method for extraction of features from long-term ECG signals	Random forest 100 % accuracy, the worst was 99.2 % accuracy
[Dubois et al., 2007]	Automatic extraction of P,Q,R,S,T from ECG	Orthogonal forward regression and an Gaussian mesa function	ECG beats	Use multiple leads to improve classification the same lead will not always be the best throughout a classification task
[Abdul-Kadir et al., 2016]	Classification of atrial fibrillation and selection of optimal feature set	neural networks and support vector machines	ECG beats second-order dynamic differential equation features	95.3 % accuracy
[Elgendi, 2013]	Faster processing of ECG to detect QRS complex	Their own developed technique	Baseline adjusted ECG	Was possible to detect faster but with less accuracy
[Jadhav et al., 2012]	Normal vs abnormal beats in incomplete data sets	Generalized feed forward neural network	Baseline adjusted ECG, features not mentioned	Accuracy of 82.35 %.
[Liu et al., 2015]	Performance of classification on imbalanced training data	Weighted extreme machine learning	Baseline adjusted ECG and PQRST features	Accuracy of 97.89 % with only 11 seconds training time
[Oster et al., 2015]	Heart rate turbulence following premature ventricular contraction to predict patient outcome	Clustering	Auto derived features from ECG beats	Accuracy of 99.50 %

Seera et al. [2015] made a review of state of the art classifiers and data pre processing techniques. They split the classifiers into the statistical, neural networks, decision trees and ensemble models, support vector machines and other methods. The most common used where neural networks, but they concluded that the random forest method performed the best overall and especially in high noise signals. Furthermore they concluded that they did not investigate an optimal feature set and that might have effected the results.

This random forest method was investigated by Masetic and Subasi [2016]. They used this classifier for congestive heart failure and an autoregressive burg method for extraction of features from long-term ECG signals. They compared the random forest method with neural networks, k-nearest neighbour and support vector machines. Only the random forest method had a 100 % accuracy but the other classifiers were, close with 99.2 % to be the worst accuracy.

Dubois et al. [2007] investigate the automatic extraction of the P, Q, R, S and T waves from ECG. Using a orthogonal forward regression and an Gaussian mesa function. They model each beat recorded with the ECG to be either normal or abnormal. They conclude that results from using just one lead are better than using multiple ECG leads. Detection of the parameters PQRST in the ECG should not necessarily be made from the same lead for each heartbeat [Dubois et al., 2007].

Classification of one problem (atrial fibrillation) was investigated in Abdul-Kadir et al. [2016] with a second-order dynamic differential equation (SODS) system. The natural frequency, damping coefficient amongst other features was derived from the SODS using different window lengths. The author used ANOVA and t-tests to determine the signification features and applied them to neural networks and support vector machines. With neural networks and a 4 second window for the SODS it was possible to obtain an accuracy of 95.3 %

Elgendi [2013] concluded that current QRS detection method Pan Tomkins very robustness regarding accuracy but elicit the problem of the heavy computational power it requires. They provide an alternative technique for fast detection with less computational power but also less accuracy than state-of-the-art standards.

[Jadhav et al., 2012] investigate arrhythmia patients into two classes: normal or abnormal using a generalized feed forward neural network. They tested the robustness of the classifier by using incomplete data sets from 452 different patients. It was possible to classify with an accuracy of 82.35 %.

Automatic decetion of problems from the ECG is important [Liu et al., 2015]. Liu et al. [2015] investigated one of the methods for machine learning on the data called extreme machine learning (ELM). They use a weighted ELM since they postulate it should have better time performance and generalization than traditional machine learning for imbalanced ECG samples. They derive features from the ECG such as ST segment and the QT interval for use en the weighted ELM neural network and manage to achieve an accuracy of 97.89 %. with only a 11 second training time. Other networks such as backward projection neural networks can perform similar results but requires more than 3050 seconds of training [Liu et al., 2015].

Oster et al. [2015] investigate a more challenging area of the automation of diagnosis via ECG. They investigate heart rate turbulence (HRT) following PVC as it has a indication of patient outcome after a PVC incidence [Oster et al., 2015]. They found that many systems today extract features manually from the ECG by the PQRST complex and then try to perform a feature optimisation with a following classification. Systems not relying on these features but instead relying only on beat labels by an expert outperforms these systems. They propose a supervised learning method with an uncertainty estimate of outcome. The idea is that in real life application it is better to not classify than misclassifying an ECG beat. The accuracy of their system was 99.50 %.

Findings for keywords deep learning

The one article was found for “deep learning” it is displayed in following table 1.7.

Table 1.7: Literature found for the keyword "deep learning" along with a description of what was investigated, the machine learning methods used, the ECG source and findings

Literature	Investigated	Machine learning methods	ECG source	Findings
[Kalkstein et al., 2011]	Data driven approach to detect quality of ECG instead of human interaction	K nearest and random forest	ECG beats baseline-corrected	Best accuracy for both classifiers where roughly 93 %

As identifyed by the Khunti [2013] the quality of the ECG can be a problem. An automatic solution to that problem is resolved by Kalkstein et al. [2011]. They suggest for a data driven approach to this assessment of quality rather than a model based one driving features from the ECG rather than human selected features. They use K nearest and random forest as classifiers. The best accuracy for both was roughly 93 %. They conclude that a more uniform classification of quality could be feasible.

1.4 Problem statement

In the application of machine learning of ECG, multiple kinds of machine learning have been applied Khunti [2013], Oster et al. [2015], Dubois et al. [2007]. Machine learning is however limited by the quality of data [Kalkstein et al., 2011, Oster et al., 2015] and as pointed out by [Khunti, 2013] quality can be a problem. This is especially the case where classifiers are trained to classify noise rather than the underlying ECG [Oster et al., 2015] and therefore also quality detection has been investigated Agrafioti [2011].

Common in the literature is that the authors apply their technique to different features [Rahhal et al., 2016, Chandrakar and Sharma, 2015] even in applications that are simmilar. This was done in the articles Oster et al. [2015], Dong et al. [2014], Liu et al. [2015] who all performed diagnosis of the heart but used a different set of features and different classes and number of classes in the classifiers. Not only diagnosis of the heart was investigated in the literature but also emotions [Xu et al., 2010, Agrafioti, 2011], ECG in biometrics [Agrafioti, 2011], between subject variations [Shadmand and Mashoufi, 2016, Chandrakar and Sharma, 2015] and time variation [Dong et al., 2014] has been investigated. The accuracy of classification for emotions and for identification of individuals was lower than of the diagnostic classifications. The studies investigating between subject variations concluded that profiling of ECG features had to be done for the individual patients to improve classification of disease. The study investigating time variation concluded that it is possible to diagnose the ECG heartbeats over time [Dong et al., 2014].

There is a lack of a gold standard feature set [Casarella M, 2012, Seera et al., 2015] as new feature sets are usually derived each time from disease annotated original ECG databases [Dubois et al., 2007, Yeh et al., 2010] instead of from a pool of already defined features. This is a problem for a fast query, low cost processing and storage as investigated by Elgendi [2013], Kumari and Sadasivam [2007], Balouchestani and Krishnan [2016], Yeh et al. [2010].

Therefore, behind much of ECG research there is a lot of work done to establish an underlying grammar of the ECG as the minimal ideal reduced dataset so that one can identify either the person, mood or diagnosis. The idea of using an experts for feature extraction does not seem to be the correct way but rather features should be synaptically or statistically extracted [Casarella M, 2012, Oster et al., 2015, Agrafioti, 2011, Lu et al., 2016].

These papers Casarella M [2012], Oster et al. [2015], Agrafioti [2011], Lu et al. [2016] investigating ECG with features not extracted by an expert show promising results thereby bypassing the feature extraction problem. Some of these results Casarella M [2012], Masetic and Subasi [2016], Seera et al. [2015] showed the accuracy of close to a 100 % where some algorithms showed to be more robust to noise than others. These results could be due to either overfitting, grammar in the data set is ideal for the application of the classifier or a general descriptive grammar of the ECG has been found. If such a general descriptive grammar is found then the ECG can be reduced to this grammar and allow for fast query, processing and storage.

Based on what was found in the literature review figure 1.8 have been generated. This figure illustrate what was investigated in the literature found. The figure contains three clusters *ECG*, *Application* and *Performance* marked with blue. Furthermore green boxes are marked. These green boxes illustrate what is researched in this project. The *ECG* cluster represents areas of ECG research where *Application* represents the application context of the research. The last cluster *Performance* is included since it has been found to exists within multiple areas of research. This were for example the case for the research field of unsupervised classification in Casarella M [2012], where the performance of the classification was researched in regards to noise.

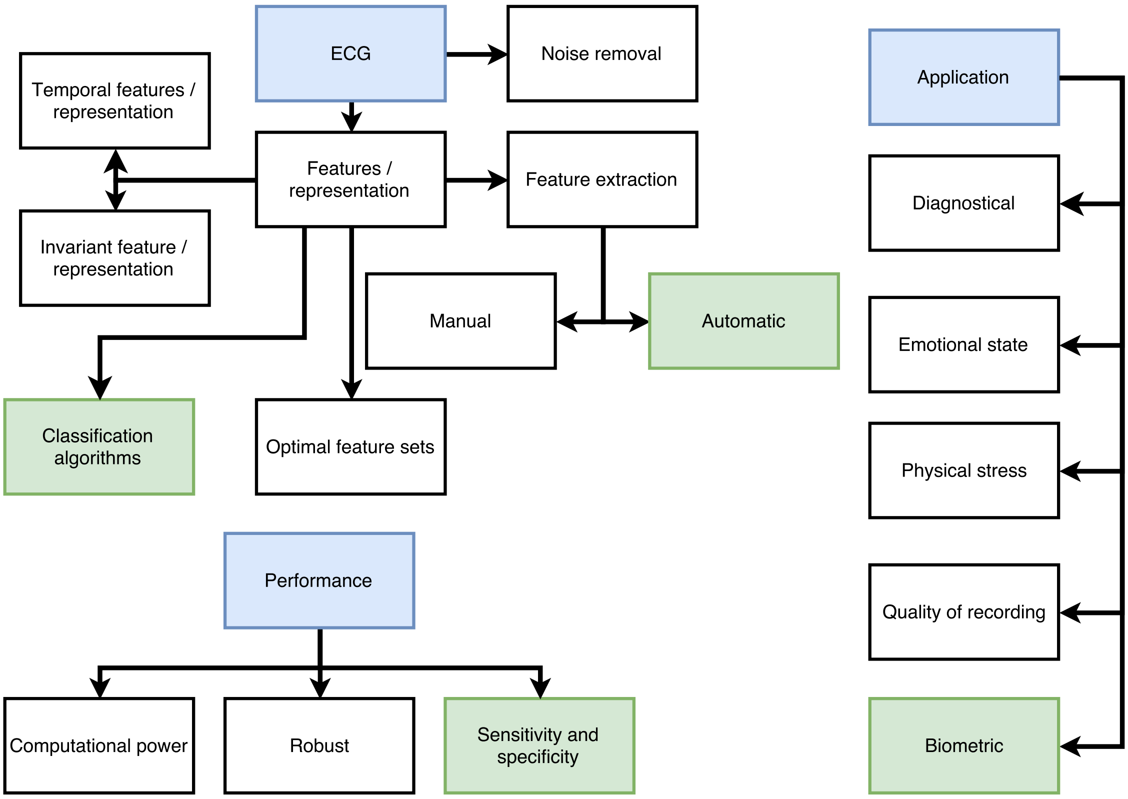

Figure 1.8: Illustration of the literature found. literature split into three clusters each illustrated by a blue box. Green boxes represents the area of what is being research in this project. Arrows are drawn between boxes to illustrate that these are sub research fields of the field where the arrow originates

Researched areas marked on figure 1.8 were selected since automatic extraction of features and classification would improve the quality of machine learning research and lower the amount of required human involvement. The biometric application was chosen since it had higher error rate than other application areas found, but was still found to be within an area that showed promise [Agrafioti, 2011]. Knowing the limitations of a solution a solution and the performance of the solution is evaluated according to the accuracy of the system.

This project therefore seeks to investigate if this general descriptive grammar can be found by using methods applied to another application than the biometric one. The HTM methodology used in Casarella M [2012] were chosen since the research field fit within the scope of this project and provide a novel feature extraction and classification that does not require an ECG expert.

A problem statement can be summarised to the following: *Till which extent can the underlying biometric ECG grammar be found with HTM on subjects heartbeat over multiple sessions?*

Chapter 2

Problem solution

In this section a more extensive description of the algorithm HTM is made along with a detailed description of the applied methods including the real life application used, software tools, data preparation and the configuration of the HTM model.

2.1 Hierarchical temporal memory overview

The goal of the HTM model is to mimic the human neocortex columnar structure and algorithmic property as these properties are a big part of human intelligence. HTM is essentially a distributed memory model where the model learns and recalls a pattern based upon prior encounters thereby utilising temporal information and requiring less training than conventional neural networks. One other difference between neural networks and HTM is that various parameters are adjusted in HTM rather than just the weight between adjacent nodes as of neural networks. These parameters are such as threshold for establishing new connections between "neurons" and "weight" of a change in weight between neurons. [Hawkins et al., 2011, Casarella M, 2012]

HTM is hierarchical with a single node layer (classifier) in the top and a multi-node data input layer at the bottom (sensory layer) with multiple layers in between containing multiple nodes. Information exchange can happen both internally within a region of a layer and both ways between layers. Information exchanged between layers will gain a sparse distributed representation (SDR) in the receiving layer meaning only a small percentage (2 %) of the nodes will be activated following an input. One of the mechanisms to achieve the SDR is that highly activated nodes will deactivate nodes with less input. Implemented in HTM is also a tiebreaker which randomly selects what to deactivate if there is no clear indication of what columns to deactivate. [Hawkins et al., 2011, Casarella M, 2012] The flow of information is illustrated in figure 2.1.

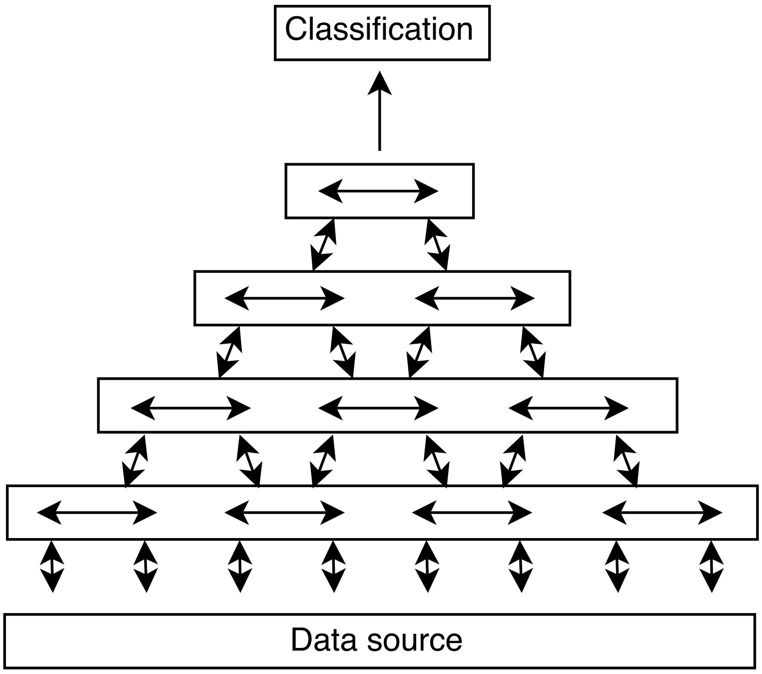

Figure 2.1: Simple illustration of the exchange pathways of information in a multiple layered HTM. Top layer is called classification layer and bottom layer is called a sensory layer. Connections are made within a layer and between layers. Figure is recreated from [Hawkins et al., 2011]

The layers are not as simple as illustrated in figure 2.1. In each of the layers the nodes are actually columns of what is annotated cells which are interconnected with other cells from other columns. Essentially each cell gets activated via a input to its column. [Hawkins et al., 2011]. There are more advanced requirements before a cell can get active. This is illustrated in figure 2.2.

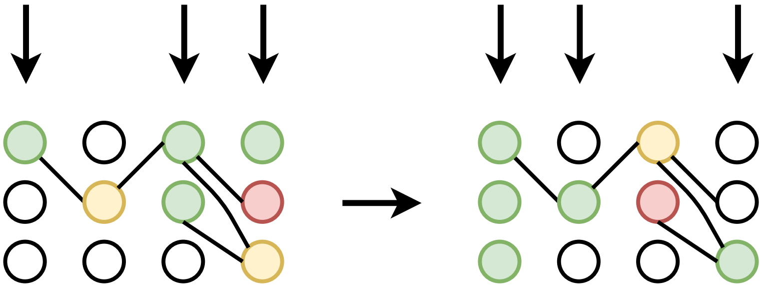

Figure 2.2: Illustration of four columns with 3 cells in each column. On the left some cells are active due to a input to the columns. These active cells are marked green. The yellow cells are connected to two other cells that are active and are therefore in a predictive sate. The red cell is only connected to one cell so it is not in an predictive state. On the right a new input arrived. Every cell in the first column is active since non of the cells were in a predictive state before. In the other columns the cells that were predictive before are now active

The bottom layers of an HTM deals with minor components of the input where the higher levels learn more complex patterns. This is also seen in the neocortex where higher level neurons deals with more complex patterns such as words and lower levels neurons with for example syllables of the words. [Hawkins et al., 2011]

Basic functions exist within the HTM algorithm: encoding, learning, inference, prediction, behaviour and classification. [Hawkins et al., 2011, Casarella M, 2012]

The learning phase is not needed but does improve latter classification. The reason for why it is not needed is that online learning can be used with HTM or disabled for each layer in the model. [Hawkins et al., 2011, Casarella M, 2012]

2.1.1 Encoding

At the sensory level an encoding should be made of the input space. The encoding is a representation of the data set and corresponding class. HTM only finds patterns, so in order to provide essential information about the input additional data needs to be added so the total input space will be an encoding contain both original data and meta data. [Casarella M, 2012] Two ECG's are encoded in figure 2.3 and figure 2.4.

```
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 1 0 0 0 0
1 1 1 0 0 0 1 1 1 0 0 0 1 0 0 0 0 0 0 1 0 1 1 1 0 0 1 1 1 1
0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

Figure 2.3: Person one heartbeat

```
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0
0 0 0 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 1 0 0 0
1 1 1 0 0 0 0 1 1 0 0 0 1 0 0 0 0 0 0 1 0 1 1 1 0 0 0 1 1 1
0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

Figure 2.4: Person two heartbeat

What columns initially revive the encoded data are usually random but multiple other methods exists including a topology method that will allow for more complex patterns to be learned of the dataset. This however mean that more data is needed to learn patterns. [Hawkins et al., 2011]

2.1.2 Learning

Each layer learns in unsupervised manner by finding patterns in the input and the sequence of patterns over time. The two things happening is called spacial and temporal pooling, where spacial pooling is adjustment of connections between columns and the input space and temporal pooling is between cells of columns. [Hawkins et al., 2011]

[Casarella M, 2012] argues that some supervised learning is happening for the classification layer and sensory layer as the data needs to be labelled in the classification layer and data needs to be encoded in the sensory layer.

Cells in each column are as earlier described connected to cells in multiple other columns just as columns are connected to the output of multiple columns in their input spaces. Their initial connections can be set in multiple ways but is usually done randomly. [Hawkins et al., 2011, Casarella M, 2012] In figure 2.5 an illustration is made of connections between columns in each layer of an example HTM implementation.

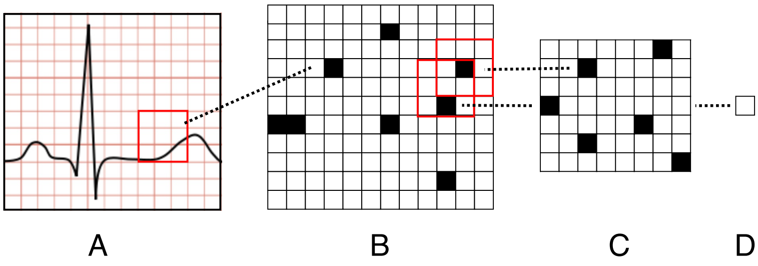

Figure 2.5: From A, a subset of the original ECG image is connected to a column in B. In the connections between B and C an overlap in the data source B is illustrated. In the connection between C and D the the entire set of C is used as a feature set for classification

Both the spacial and temporal pooling works by cells or columns that gets activated together strengthens their connections (permanence), cells or columns that don't, they get a weaker connection. These trained connections functions as learned patterns of HTM. [Hawkins et al., 2011]

Inhibition was mentioned earlier as a way to obtain a SDR during spatial pooling. Besides inhibition another mechanism called boosting also exists and can be enabled [Hawkins et al., 2011]. In homoeostasis inhibition and boosting mechanisms exists [Martini et al., 2012], and that idea is expressed with inhibition and boosting in HTM.

Boosting allows less activated columns to be more activated and over activated neurons to be activated less and thereby allow more complex patterns [Hawkins et al., 2011].

After a manually defined number of iterations, where these connections are adjusted, temporal pooling is performed. [Hawkins et al., 2011, Casarella M, 2012].

The fundamentals for temporal pooling is a cells three states: active, predicting or deactivated. Each cell will look at its connections and if a certain amount of these is active then it will go to a predictive state. [Hawkins et al., 2011]

When a column gets an input only the neurons in a predicting mode will then be activated. If none is in that stage then all the cells in the column will be activated as the input is unexpected. [Hawkins et al., 2011, Casarella M, 2012]

As a cell becomes activated it will look to other cells from other columns that have been active earlier and strengthen the connection where connections to cells that were inactive will be weakened. [Hawkins et al., 2011]

Essentially what is learned under temporal pooling is a series of Markov models representing the transitions between activated nodes where these models are grouped together in a Agglomerative Hierarchical Clustering fashion. [Casarella M, 2012] A Markov model is a way to model statistical system where only knowledge about the current state of the system is needed to make predictions of the future (temporal memory of HTM). Agglomerative Hierarchical Clustering is illustrated in the following figure 2.6.

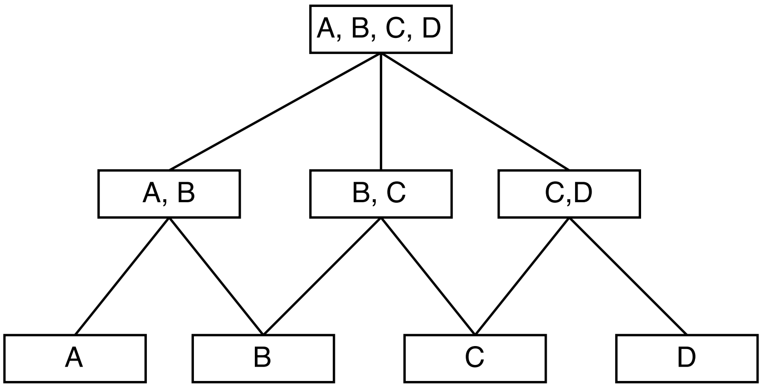

Figure 2.6: This algorithm works by clustering subcomponents according to a given similarity measure into bigger components that eventually will result in a big cluster expanding the entire dataset. [Casarella M, 2012] The figure is created from [Casarella M, 2012]. This figure illustrates that event A, B are not similar, but A,B is more similar than B,C. Eventually the cluster converge as the similarity measure expands and a single group containing A,B,C,D is created

The different level of the cluster is the different levels of the HTM model. [Hawkins et al., 2011]

2.1.3 Inference

Once learning is finished, educated guesses (inference) can be made about the underlying patterns of novel inputs. Each input will be somewhat novel in that sense that variance will occur in inputs [Hawkins et al., 2011, Casarella M, 2012] such as the same abnormality will not look identical in each ECG image. HTM is robust to these variations because of the aforementioned generation of a SDR set under learning. With the SDR just a partial match of the input with learned patterns is required to make an educated guess of the underlying patterns of the input. [Hawkins et al., 2011, Casarella M, 2012]

2.1.4 Prediction

HTM memories transition between SDR and their ability to predict in a sense that it predicts a following sequence of input based upon the prior input. This makes HTM very robust to noise and missing data in a sense that prior input will help to generate a context of what is to be predicted so that its more likely that a correct inference is made. The predictions gets more stable at higher layers as underlying minor components of an input shifts more rapidly than higher level components. [Hawkins et al., 2011, Casarella M, 2012]

Inference and prediction is somewhat intertwined. Output from a layer is the resulting inference and prediction(s) made on its input and act as input to the adjacent layer. Basically the output is a vector containing activity of all the cells in the layer. If any cell is in an active or predictive state then the output of that column is 1, otherwise the output of that column will be 0. Multiple predictions can be made and this help to increase the stability of the system at higher levels as predictions change less than the input. [Hawkins et al., 2011, Casarella M, 2012]

2.1.5 Behaviour

Behaviour is very different from other functions of the HTM. Instead of recognition, behaviour references to an actual reaction of a system. This reaction is best described as a neural motor reaction to input to sensory input. [Hawkins et al., 2011] An reaction could be call grant access in a biometric application or call an ambulance in a medical application.

Behaviour functionality is however not necessary in the first stages of implementing HTM and is therefore not yet implemented in many systems. [Hawkins et al., 2011]

2.1.6 Classification

The last node in HTM can be multiple kinds such as an SVM or K-Nearest classifier. Natively to HTM implementations are the classifier is the Naive Bayes (NBA) algorithm. [Casarella M, 2012] The NBA classifier is also called simple Bayes and independence Bayes.

Essentially this classification algorithm is naive in that sense that it assumes the indices of the input vector is independent. This means that correlation of features do not contribute to information about class. [Caruana and Niculescu-Mizil, 2006]

This algorithm learns in a supervised matter. The output from the layer below the classification layer acts as a feature set. The feature set gets labelled by an investigator with a class where multiple of theses act as the training set for the NBA. [Casarella M, 2012]

2.1.7 Configuration

Naturally multiple parameters can be altered in an implementation of the algorithm such as number of layers, number of columns, number of cells in each of the columns and their connections. Further more configurations can be set for the depth of the temporal memory (size of the context). [Hawkins et al., 2011, Casarella M, 2012] Not every parameter is covered in this section because of the excessive amount of parameters.

Chaining parameters in any form such as increased number of cells, layers or increased context size will allow for more complex patterns to be learned (increased learning capacity). [Hawkins et al., 2011, Casarella M, 2012]

Selection of values for memory size, boosting or any other parameter do not have a gold standard yet. Instead one should rely on former literature on the topic or whatever is required by the current application of HTM. [Hawkins et al., 2011, Casarella M, 2012]

2.2 Applied methods

In the current study parts of the methodologies applied in Casarella M [2012] are used, that is a dataset is prepared and a Hierarchical Temporal Memory model (HTM) is built and configured. Multiple experiments were run in different configurations in order to investigate the influence of number of subjects and number of iterations through the dataset. The non-shared configuration for each of these experiments is described under the subsection experiments.

2.2.1 Application

The real life application used in this study is based on the literature Agrafioti [2011]. Described in Agrafioti [2011] are three biometric applications. The best results in form of lower error rate were in the application where heartbeats derived from a finite known population's ECG were to be identified over time. In that application a continuous classification a non overlapping dataset of five seconds ECG where used. Temporal information can be learned without the overlap in HTM. Casarella M [2012] That application is therefore mimiced in the current study so that only single consecutive heartbeats are used to identify a subject over time.

2.2.2 Software

Some prepossessing had to be done to the data since Images need to be generated from the original data set. This was done in MatLab® 9.1.0.441655 (R2016b). The actual HTM model is however not supported in MatLab® and software by Numenta NuPIC 0.7.0-dev Hawkins et al. [2011] was therefore used. Run on 2015 MacBook Pro 15 Inch 2.8 GHz i7 with 16 of ram. MacOS version 10.12.4. Python 2.7.10. Further more a MongoDB 3.04 database was created with the data to avoid memory leaks.

2.2.3 Dataset

The origin of the dataset is a database from the general practitioners of Copenhagen, Denmark. It is 500 Hz sampled unfiltered ECG for 25.000 patients with two sessions each. There were 14.225 female and 10.775 male with a total mean age of 64 years and a standard deviation of 16.64. From each session the full length of 10 seconds was chosen. In the application being mimiced from Agrafioti [2011] they used multiple leads but in this study the lead II was chosen to match the methodology by Casarella M [2012]. The dataset contained both healthy and unhealthy patients which is a broader spectrum than both Agrafioti [2011], Casarella M [2012] who only used either one of the two patient groups.

Subsampling of the dataset

Two subsampled dataset called optimisation and control were substracted from the entire dataset. Samples were assigned to the two subsets in a random manner automatically by the file system of the computer running the software. 1.250 subjects (5 % of entire dataset) were included in each of the subsets. This number was chosen in order to reduce the runtime of the software, since initial investigations indicated that it would not be able to finish within time scope of the project otherwise.

Segmentation of heartbeats

In order to segment the heartbeats from the ECG the Pan Tompkins algorithm was used as described by Elgendi [2013], Agrafioti [2011], Dubois et al. [2007]. The Pan Tompkins algorithm were developed in Pan and Tompkins [1985], provided an accuracy of 99.3 % on 12-lead ECG arrhythmic data and can be used on the unfiltered ECG signal. The algorithm contains four stages: 5-15 Hz filtering including subtraction of mean ECG, differentiation, squaring and 150 ms width integration. A refractory period of 200 ms is included in the detection algorithm to avoid false detection of the

QRS complex. If no certain complex has been found within 360 ms of the previous the highest slope of the processed ECG is selected.

The width of a heartbeat were defined to be 740 ms in this current study. A fixed length was chosen to match with the methodologies by Casarella M [2012]. The exact length of a heartbeat is affected by both disease and heart rate, but was calculated based upon it normally being within; 80-100 ms for the P-wave, 120-200 ms for PR-interval and 300-440 ms for QT-interval resulting in a maximum total length of 740 ms. Casarella M [2012].

Encoding images

Images were encoded so that each sample corresponded to a pixel. Since the length of a heartbeat were calculated to be 740 ms, and each sample corresponds to 2 ms, a width of 370 samples (370 pixels) were selected for the image. This correspond to the length of the heartbeat normalized to the sampling frequency.

In Casarella M [2012] the height of images were 96 pixels so the same height for images were selected for the current study.

In order to determine the vertical pixel a sample belonged to, a grid was created with 96 values each representing a pixel. The distance between the value of a sample and the 96 pixels in the grid were calculated and the pixel with the minimum distance was selected to be the corresponding pixel of that sample.

In case of a tie between distance of pixels, the upper pixel was selected to be the corresponding one.

Grids were created for each individual recording so that the maximum value of the grid corresponded to the maximum value of the session and the minimum to the minimum value. Each value in the grid were calculated as described in equation 2.2.

$$value(pixel) = minValue + pixel * (maxValue - minValue)/96 \tag{2.1}$$

Training and test groups

The very essence of machine learning is the ability to predict future values based upon previous experience. Subjects data from each of their respective subsets were therefore split into training and test groups.

In Casarella M [2012] different ratios between the two groups are described. They describe that the implementers of Numenta NuPIC suggest a 8:1 training to test data ratio but that they usually used a 1:1 ratio. Casarella M [2012] uses roughly a 1:2 ratio for their own study but the ratio is not the same for each class they investigated.

In the current study the 4:1 ratio was chosen. The data amount is limited for each class in the current study compared to what was used in Casarella M [2012] and increasing the ratio gives more training data and thereby better accuracy might be achieved.

All data on each subject were used as a combined dataset where a 4:1 ratio of data was extracted for each of the subjects. This gave roughly 25 beats for training and 6 beats for test for each subject.

2.2.4 Configuration of algorithm

There are not yet any gold standards for selecting the values for parameters for an implementation of HTM as described in 2.1.7, Configuration. In this section the high level parameters are selected such as number of levels before the low level parameters such as number of cells in each column.

In the software used by Casarella M [2012] the parameters *Maximum Coincidence Count*, *Equalize Group Size* amongst other were mentioned. These parameters were however changed in the newer version of the software that is described in Numenta [2017]. It was therefore not possible to do a direct mapping of the parameters from Casarella M [2012] and parameters of HTM were therefore as a starting point left at their default value.

As described by Casarella M [2012] whose methodology is mimicked in the current study, parameters had to be tweaked to fit within the problem domain. Since the application or problem domain in the current study is biometrics, values in this current study would not necessarily be the same as of Casarella M [2012].

Amount of layers

The number of levels to use is not directly specified by Numenta, but experiments have been made with both three and four levels. More levels increased the robustness of the system but decreased the accuracy. [Casarella M, 2012] For image classification problems three levels seemed to produce the best results [Casarella M, 2012], and a three level structure is therefore used in the current study.

Configuration of layers

The configuration of the two bottom layers was made with the following considerations from Casarella M [2012]: The image size, patterns within the image, the correlation between number of levels and number of columns in each level. Following the methodology described in that literature, two prime factors were derived based upon the image size and used to calculate the configuration of columns. These calculations are made in equation 2.2 and 2.3.

$$96 \Rightarrow 3 \times 2 \times 2 \times 2 \times 2 \times 2 \tag{2.2}$$

$$370 \Rightarrow 2 \times 5 \times 37 \tag{2.3}$$

A pyramidal regression was found in Casarella M [2012] to be required for the proportion of the images. It is specified that the first layer should be an eighth of the original data set and the second layer should be a half of the first layer. The last layer is the classifier. It is however not possible to fulfil these requirements with a width of 370 pixels. Therefore the images were adjusted to a width of 384 pixels (by chaining the width of the window) and new primes were calculated in equation 2.4.

$$384 \Rightarrow 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 5 \tag{2.4}$$

By using the primaries it was possible to calculate the following properties of the layers in equation 2.5 and 2.6. This means 48 × 12 columns (width × height) in the first layer and 12 × 3 (width × height) in the second layer.

$$96/8 = 12 \| 12/4 = 3 \tag{2.5}$$

$$384/8 = 48 \| 48/4 = 12 \tag{2.6}$$

Capacity of learning

The capacity of a layer defines how many underlying patterns can be stored. Essentially this is the configuration of the SDR. The equation for the capacity in a layer is given in equation 2.7. n is number of columns where k is number of active columns. By selecting a rounded 2% suggested in Hawkins et al. [2011] the number of columns for layer one is $48 \times 12 = 576$, the number of active columns is 12 and the capacity is $2.4810e + 24$. In the second layer the number of columns for layer one is $12 \times 3 = 36$, active columns is only 1 and the capacity is $1.2517e + 09$.

$$\frac{n!}{k!(n-k)!} = \binom{n}{k} \tag{2.7}$$

Selecting number of cells

The number of temporary contexts that can be represented is given in 2.8 and limited by c number of active columns and n cells in each of the columns. Examples are given for a layer of 4 cells in each columns in Hawkins et al. [2011]. This number allows for a big temporal memory and it is argued in that literature that a layer of above the 4 is usually not needed. The temporal patterns (context) that can be learned in this implementation is therefore 16.777.216 and 4 in the first and second layer respectively. This guideline is also followed in Casarella M [2012].

$$n^c \tag{2.8}$$

Classifier and performance measures

The top layer (classification algorithm) was chosen to be Naive Bayes Classifier (NBA) since it was also used in Casarella M [2012] and a class was defined for each subject. This classifier was set to output classes and a samples corresponding probability to belong to a class. The three most likely classes and corresponding probability for each of those classes were used in further analysis. The further away from the most likely class a prediction is, the less likely it is for the system to be able to classify such a sample and an arbitrary cut-off of three classes was therefore chosen.

2.2.5 Experiments

Experiments were conducted with different amount of iterations in this present study as of Casarella M [2012]. The span of iterations were set to be between 9.000 to 60.000 times in the study by Casarella M [2012] and was only done on the spatial pooler. No improvement in classification was seen for over 21.000 iterations. Essentially it was theorised that the more times an image is exposed to the spatial pooler the better the underlying patterns can be detected. For this application the ideal amount of iterations needed were not known and multiple experiments were therefore conducted with an arbitrary number of 1, 100 and 1.000 iterations respectively for the control dataset. The trial with 1 and 100 iterations were then run again but with the optimisation dataset and an experiment with only 1 iteration but with a combined dataset of the control and optimisation dataset was also run. This last experiment was conducted to investigate the influence of the size of the dataset on accuracy of classification.

Chapter 3

Results

The results are represented in table 3.1 for the different number of iterations 1, 1.000 and 10.000 and the different optimisation and control datasets.

Table 3.1: The results of using HTM with the different datasets are given in this table. The accuracy of the systems ability to correctly classify a sample as the most likely or within the top three most likely is given

Iterations	Dataset	Accuracy (1)	Accuracy (3)
1	Optimisation	0.07 %	0.24 %
100	Optimisation	0.07 %	0.24 %
1000	Optimisation	0.07 %	0.24 %
1	Control	0.09 %	0.27 %
100	Control	0.09 %	0.27 %
1	Optimisation and Control	0.04 %	0.014 %

The results did not improve with the number of iterations. By creating a dataset consisting of both of the subsets a drop in accuracy dropped to half of the accuracy of each of the groups individually. Since the number of subjects influenced the results a lower number of subjects were investigated to further enlighten the limitation of the implemented configuration of HTM. Different number of iterations were also investigated as of the original methods. The results are presented in table 3.2.

Table 3.2: The results of using HTM with the different number of subjects are given in this table. The accuracy of the systems ability to correctly classify a sample as the most likely or within the top three most likely is given

Iterations	Num subjects	Accuracy (1)	Accuracy (3)
1	10	8.89 %	31.3 %
10	10	8.89 %	31.3 %
100	10	8.89 %	31.3 %
100	100	4.47 %	18.5 %

The results of these showed an improvement compared to the former results of this study with the number of subject influenced the accuracy of the system. The number of iterations did not change the accuracy. These findings aligns with what was already found earlier in this study. Two classifications made by the classifier has been included in figure 3.1 and 3.2 respectively. Figure 3.1 is a correct match were figure 3.2 is not. These figures illustrates how similar both the correctly and wrongly classified images look with their matches.

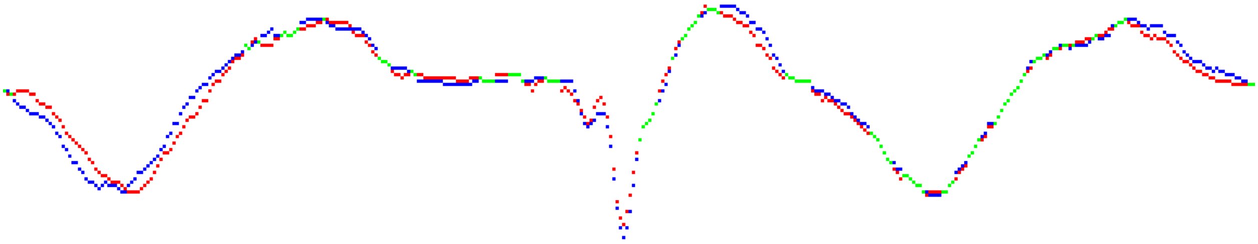

Figure 3.1: Correct match from random subject's encoded heartbeat. The shared pixels are marked with green, the non-shared are marked with red and blue where each of these colours represents two different beats

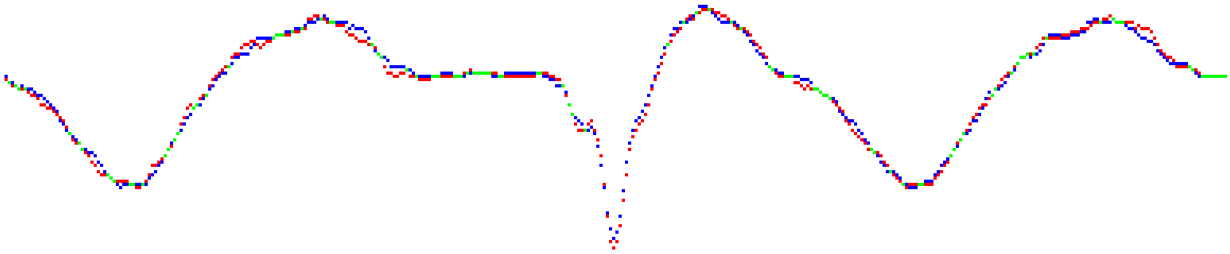

Figure 3.2: Wrong match from random subject's encoded heartbeat. The shared pixels are marked with green, the non-shared are marked with red and blue where each of these colours represents two different beats

Chapter 4

Discussion

A literature review was conducted and it was found that there are still an extensive amount of work that can be done within the field of ECG machine learning research. Pitfalls were identified of machine learning such as its ability to learn depends on the quality of the ECG recording where this quality shown to be problematic. Furthermore it was identified that even though that research has been made within the same applications such as diagnosis of a specific heart disease, the same features are not used in these and there seem to be a lack of a gold standard for the meaning and limitations of the underlying grammar of ECG. HTM showed promising results and was said by Casarella M [2012] to be robust to some of the problems that can be caused by the aforementioned problems with ECG quality as it was used on unfiltered ECG.

In order to test if the underlying grammar of ECG could be found with HTM, another application than the originally tested in Casarella M [2012] was chosen. This application was identification of a patient based on the heartbeats derived from the ECG.

In Matlab Pan Tompkins were used to segment heartbeats and from those heartbeats images were generated. These images were used in a implementation of Numenta's Nupic HTM. ECG of a total 50.000 session each 10 seconds long was used where data from 2.500 subjects were subsampled from the entire dataset. Multiple experiments were run to control the effect of the subset, number of subjects and iterations.

In the present study it was found that the accuracy of the system to classify the a heartbeat derived from the ECG were up to 31.3 &. When the number of subjects increased the accuracy of the system decreased and eventually dropped below 1 %. These results persisted with different subsets and amount of iterations.

Results

The results found in this study were less accurate than what was found by Agrafioti [2011], Gregg et al. [2016] and Casarella M [2012]. In Casarella M [2012] the application were not biometric but HTM was also used and an almost 100 % accuracy were found. This study did however only differentiate between four classes and in this study it was found that increasing the number of classes lowed accuracy of the system. Furthermore these results were achieved using 60.000 iterations. When only 3.000 iterations were used the results were 70 %. These number of iterations is higher than what used in the current study that was chosen due to the limited processing time.

In the biometric application in Agrafioti [2011], mimiced in the current study, the lowest equal error rate were 10 %. This was however only on a database with up to 52 subjects where this current study used up to 2500 subjects. More data were used for classification in Agrafioti [2011] as five seconds ECG recordings was used for classification providing more data for the classifier to make its decision than in the current study where single heartbeats were used. Using more data should improve the accuracy of the system [Numenta, 2017, Holzinger, 2016, Daume, 2012]. Furthermore filtered ECG were used in Agrafioti [2011] as of also recomended in literature Sufi et al. [2009], Rahhal et al. [2016], Chandrakar and Sharma [2015] but not done in the current study. Noises might therefore be so strong that they suppress the information about the identify of the subject in this current study.

The study Gregg et al. [2016] reported, just as the findings in the current study, that sensitivity decreased with an increasing amount of subjects. Training was made in this study on 294 subjects but included more data in form of 15 leads instead of one lead usd in the current study. Gregg et al. [2016] reported an sensitivity of 37 % but also included 8369 subjects for test that were not trained on. Since Gregg et al. [2016] tested the system be matching a trained subjects ECG with and untrained subjects ECG no match should be found thereby increasing the sensitivity of the system without increasing the actual accuracy. In Gregg et al. [2016] the entire ECG were used to determine if a match were found of a subject where only single heartbeats were used in this current study. By using the entire ECG instead of a single heartbeat the subject who have been suggested most frequently for any of the heartbeats could be selected as the subject the ECG belonged to and therefore improve the results.

Application

The current study mimiced the application described in Agrafioti [2011] but did have some differences. In Agrafioti [2011] the application was biometric identification of subjects on ECG signals. In the current study only single beats where used instead of five seconds of ECG. Chaining the application in the current study to also contain a consecutive five seconds of ECG would allow HTM to learn more of the underlying patterns as the number of columns to handle the input data would increase. By introducing an overlap in the images it would also be possible to strengthen the temporal information in the images in HTM and thereby potentially increase the accuracy of the system. In this current study only 10 seconds ECG were available and a five seconds dataset would therefore result in a very limited size of a training and test set for each subject. This issue could have been overcome by having multiple either increased length of sessions or more sessions for each subject.

The application in the current study were adapted to also fit Casarella M [2012] who used unfiltered ECG. In a real world application there might still be some kind of filtering because of the issues with ECG quality as described by Khunti [2013], Kalkstein et al. [2011]. Good results using unfiltered ECG does however show a more robust solution.

Software

It was not possible to use the same software version as used in the Casarella M [2012] even though attempts were made to contact the author. Changes in software can influence both function and stability of the software [Numenta, 2017] and therefore it might have affected the outcome of this study. This did limit the possibilities of using and replicate the experience gained by Casarella M [2012].

Dataset

One lead was used in this study with two sessions were each session was 10 seconds long. In the cases of Casarella M [2012], Agrafioti [2011] they both used ECG recordings with a length more than a 100 times greater than that of the recordings used in study.

Since Casarella M [2012] showed that an increase amount of data showed higher accuracy it might be worth investigating the effect of an increased data size for outcome.

Some of the ways to increase the data size could be to introduce more sessions since the data size for each session in this study is limited to 10 seconds. This will also give the possibility of gaining more complex temporal information from the ECG since only two sessions are available. Weather or not this would be feasible with more sessions can be discussed since it would require resources from healthcare professionals and patients to collect the extra amount of data.

Different leads might provide information different from each other about the heart Bayés Bayes de Luna [2012], Silbernagl and Despopoulos [2009]. In HTM no additional knowledge about what the input represents (such as its diagnostical value) is required according to Casarella M [2012], meaning that even thought the exact value of these leads is not fully understood in a biometric application it would indeed be possible to use multiple leads to increase the dataset.

The dataset contained a broader spektrum of patients than both Agrafioti [2011] and Casarella M [2012] since both unhealthy and health subjects were included in this current study. In Agrafioti

[2011] it was shown that the ability to perform the biometric application was affected by emotions. Weather or not accuracy of the system could also be affected in the same way but for health status of the subject is not known but could be controlled for by using either only healthy subjects or for subjects of a specific disease.

Subsampling of the dataset

The subsets from the entire dataset were subtracted in a random manner. This does however also mean that a balance between male/female, age and different diagnosis were not known for these subsets. Introducing a pesudo randomisation could ensure an equal distribution of these. It is however not known how these demographics actually influence the accuracy of classification. The size of the subset did however have an influence of the accuracy of the system as an increased number of subjects decreased the accuracy of the system. Databases can however require even lager group sizes than used in this current study if comparisons would to be made for the entire database.

Segmentation of heartbeats

Pan Tompkins were used to segment heartbeats. This method has proven to be accurate and robust Pan and Tompkins [1985]. A specific width of a heartbeat were selected. This is even though that some variations in a heartbeat length will be present due to factors such as the heart rate Casarella M [2012] and thereby an overlap between heartbeats might occur. The purpose of segmenting heartbeats might therefore be lost and a fixed length of time could therefore be used instead.

Encoding images

The height of the image where chosen to follow Casarella M [2012]. Even though it matched with that literature increasing the amount of pixels would however allow for more columns in the HTM layers and thereby increase the the capacity of learning for the layers.

Training and test groups

Recommendations from Numenta [2017], Casarella M [2012] were not directly followed as the study Casarella M [2012] had more data available and the training data was maximised. In order to avoid over-fitting of the entire dataset and investigating the effect of different subsets an optimisation and control dataset was included. This was done according to what is described by Holzinger [2016], Daume [2012]. They argue that this is a robust method to overcome over-fitting issues when optimisation of algorithm has to be done.

The amount of beats for testing were however very limited for each subjects as only 2-3 beats were available. Since temporal information is stored as consecutive beats and only a few are available the change from one subject to another might disrupt the ability to classify. Using one session as training and another one as test would increase the amount of data available for temporal information.

Configuration of algorithm

Only Casarella M [2012] were found to use HTM on ECG from the literature review. There were not done any further investigations of literature to find out how to optimise this HTM model other than reading through the software documentation of HTM. Important knowledge could therefore have been lost that could help optimise the HTM configuration.

Numenta [2017] has created a new tool since the publication of Casarella M [2012] to help optimise the model, called swarming. This tool was however not used since it would move the methodology of this study further away from Casarella M [2012] but could potentially improve the results according to Numenta [2017]. Casarella M [2012] did also notice that a correct configuration of the algorithm to the problem domain would significantly improve the results and a correct configuration was not necessarily found in this study even though attempts were made. It is also worth noticing that the HTM algorithm is not completely implemented in the software tool [Numenta, 2017] used in this project.

Amount of layers

The amount of layers did follow the guidelines by Numenta [2017] and Casarella M [2012]. It is therefore not likely that another amount would change the outcome of this current study.

Configuration of layers

Each of the layers were configured to follow what was found in Casarella M [2012]. These recommendations were however made where more temporal information were available. Increase the amount of columns while still maintaing the pyramidal form, would increase the amount of spatial information that could be deducted from the images available.

Capacity of learning

It was selected that 2 % of columns should be active. This did follow all recommendation set forward by Casarella M [2012] and Numenta [2017]. Due to the limited amount of columns this value however did increase the amount of temporal information that could be derived from the already limited amount of temporal information due to the few beats. Increase this number might therefore in this case prove to be beneficial for the accuracy of the system. This could be tested with the new swarming methods by Numenta [2017].

Selecting number of cells

The number of cells did not have big effect of the amount of temporal information that could be stored due to the limited amount of active columns selected. This number should therefore potentially be much higher as examples given by Numenta [2017] normally include thousands of temporal patterns.

Classifier and performance measure

Other classifier types used than the one in software for this project have shown better results on other datasets [Caruana and Niculescu-Mizil, 2006] and this might therefore limit results. Furthermore the amount of classes (subjects) to differentiate between in this study was up to more than a 100 times higher than in studies compared under results. The amount of information about the underlying ECG grammar from the layers before the classifier might not be sufficient to differentiate between that amount of subjects. It was found that increasing the amount of subjects decrease the accuracy of the classifier.

The actual value of the probabilities of a sample belonging to a subject were not further investigated in the current study. With a high accuracy of the system these might carry more information beneficial for the understanding of underlying grammar of the ECG but were not used here due to the low accuracy when the higher number of classes were used.

Experiments

The amount of iterations chosen were much lower and below the amount of iterations tested in Casarella M [2012]. They show a somewhat sigmoid form of the relationship between number of iterations and classification accuracy thereby using just a 1.000 iterations in the current study would might have resulted in the same accuracy for Casarella M [2012]. In other words the amount of iterations used in this current study might have been insufficient, but were however limited due to the time frame of the current study. Furthermore the amount of iterations and subsets were not great enough to actually perform any statistics and thereby prove the effect of the subset size and number of iterations. This would require a greater amount of iterations of more subsets with different sizes.

Technical issues

In this project a MongoDB server had to be created to distribute the memory load as the software had problems crashing midway throughout the calculations. This was due to the very large dataset above 30 GB. Further more it took roughly a full week for the computer to calculate results of just 100 iterations above for the respective dataset. This massive requirement of hardware resources might limit the real life feasibility for usage of HTM especially in the initial training phase. After the training phase less than a minute was required to perform the actual classifications.

One solution could therefore be to apply online learning or train on just a specific subject with abnornamlity detection enabled Numenta [2017]. Only training on a specific person and then detect similiar ECGs from that persons was done by Gregg et al. [2016]. Here time of training could be reduced since the amount of training data would be limited and only two classes would be needed: the specific subject and others (abnormal).

Another solution could be to create multiple HTM instances. Calculations could then be distributed and each HTM instance could in parallel detect and classify different underlying parts of the ECG that then can be used in a HTM instance that combine these results to classify the subject.

Future perspective

Finding the underlying ECG grammar and especially in an automated manner with machine learning can be proven to be an essential tool in both biometrics but also in diagnosis. For example could a specific part of the grammar of the ECG help diagnose a specific disease. Casarella M [2012] points out that one of the benefits of HTM is that no prior knowledge is required about the underlying parameters of ECG and their origin in the leads. This means more leads could be included as dataset for HTM even though their representation of the ECG grammar is not yet known and that representation could be automatically investigated. Indeed this current study show that even though no prior knowledge is known about the grammar representing biometric information from ECG lead II some matches could still be found. Finding the underlying grammar would also help illuminate the limitation of ECG due to for example inter- and intra-subject variability within the grammar. This current study's findings show that increasing that increasing the number of subjects and thereby introducing more variability challenges the classifier as the accuracy drops. The study did however also find that new configurations and methods as been introduced to counter some of these issues as numenta introduced a tool called swarming for HTM. Finding the right configuration for HTM rather than completely discarding it as a tool due to the initial low accuracy might therefore be more feasible for the research of underlying grammar of ECG.

Chapter 5

Conclusion

In order to test if the underlying grammar of ECG could be found with HTM, another application than the originally tested in Casarella M [2012] was chosen. This application was identification of a patient based on the heartbeats derived from the ECG.

It was possible to use underlying biometric grammar of the ECG to classify heartbeats with an accuracy up to 31.3 &. When the number of subjects increased the accuracy of the system decreased and eventually dropped below 1 %. These results persisted with different amount of iterations.

The results of this study are somewhat novel because of the limited dataset available for each of the subjects compared to other studies of ECG in a biometric application, and the high amount of subjects compared to amount of data for the individual subject. Whether or not the results are due to the limitations of HTM or the inter-subject variability of ECG, would require a more similar dataset with other biometric applications.

Furthermore, the application of ECG in HTM for a biometric application is also novel so the right configuration for a biometric application of HTM is therefore not necessarily used in this project. It is worth investigating how a different configuration of HTM influences the results where Numenta's newly developed Swarming tool might help.

References

Abdul-Kadir, Mat Safri, and Othman, 2016. Nurul Ashikin Abdul-Kadir, Norlaili Mat Safri, and Mohd Afzan Othman. *Dynamic ECG features for atrial fibrillation recognition.* Computer Methods and Programs in Biomedicine, 136, 143–150, 2016. ISSN 18727565. doi: 10.1016/j.cmpb.2016.08.021. URL http://dx.doi.org/10.1016/j.cmpb.2016.08.021.

Foteini Agrafioti. ECG in Biometric Recognition : Time Dependency and Application Challenges by Foteini Agrafioti A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy Graduate Department of Electrical and Computer Engineering Univers. Doctorate, University of Toronto, 2011.

Balakrishnan, Srinivasa Chakravarthy, and Guhathakurta, 2015. Minimol Balakrishnan, V. Srinivasa Chakravarthy, and Soma Guhathakurta. *Simulation of cardiac arrhythmias using a 2D heterogeneous whole heart model.* Frontiers in Physiology, 6(DEC), 2015. ISSN 1664042X. doi: 10.3389/fphys.2015.00374.

Balouchestani and Krishnan, 2016. Mohammadreza Balouchestani and Sridhar Krishnan. *Advanced K-means clustering algorithm for large ECG data sets based on a collaboration of compressed sensing theory and K-SVD approach.* Signal, Image and Video Processing, 10(1), 113–120, 2016. ISSN 18631711. doi: 10.1007/s11760-014-0709-5. URL http://dx.doi.org/10.1007/s11760-014-0709-5.

Bayés Bayes de Luna, 2012. Antoni Bayés Bayes de Luna. *Clinical Electrocardiography.* John Wiley & Sons, Incorporated, 2012. ISBN 9781118392027.

Denis MD Berdajs and Marko PROF MD I. Tuna. The Heart Conduction System. In *Operative Anatomy of the Heart*, pages 427–454. 2011. ISBN 978-3-540-69229-4.

Buchanan, Smith, and Eckroth, 2017. Bruce Buchanan, Reid Smith, and Joshua Eckroth. *Machine learning overview*, 2017. URL http://aitopics.org/topic/machine-learning.

Caruana and Niculescu-Mizil, 2006. Rich Caruana and Alexandru Niculescu-Mizil. *An empirical comparison of supervised learning algorithms.* Proceedings of the 23th International Conference on Machine Learning, pages 161–168, 2006. ISSN 1595933832. doi: 10.1145/1143844.1143865. URL http://doi.acm.org/10.1145/1143844.1143865.

John Casarella M. The Classification of Un-preprocessed ECG Waveforms through the Application of the Hierarchical Temporal Memory Model. Doctor thesis, Pace University, 2012.

Chandrakar and Monisha, 2015. Chinmay Chandrakar and Sharma Monisha. *Qualitative features selection techniques by profiling statistical features of ECG.* EEE India Conference (INDICON), 2015. doi: 10.1109/INDICON.2015.7443145.

Chandrakar and Sharma, 2015. Chinmay Chandrakar and Monisha Sharma. *A real time approach for classification of ECG beats using repetition-based pattern detection and cardiac profiling scheme.* International Conference on Computing, Communication and Security, ICCCS 2015, 2015. doi: 10.1109/IC4.2015.7375607.

Chauhan and Vig, 2015. Sucheta Chauhan and Lovekesh Vig. *Anomaly detection in ECG time signals via deep long short-term memory networks.* Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics, DSAA 2015, 2015. doi: 10.1109/DSAA.2015.7344872.

Christoffels and Moorman, 2009. Vincent M. Christoffels and Antoon F M Moorman. *Development of the cardiac conduction system why are some regions of the heart more arrhythmogenic than others?* Circulation: Arrhythmia and Electrophysiology, 2(2), 195–207, 2009. ISSN 19413149. doi: 10.1161/CIRCEP.108.829341.

Daume, 2012. Hal Daume. *A course in machine learning Ch3.* Self Published, 2012. URL http://ciml.info.

Luca, 1997. C J de Luca. *The use of surface electromyography in biomechanics.* Journal of Applied Biomechanics, 13, 135–163, 1997. ISSN 1065-8483. doi: citeulike-article-id:2515246.

Dong, Wang, Hu, and Ou, 2014. Xunde Dong, Cong Wang, Junmin Hu, and Shanxing Ou. *Electrocardiogram (ECG) pattern modeling and recognition via deterministic learning.* Control Theory and Technology, 12(4), 333–344, 2014. ISSN 21980942. doi: 10.1007/s11768-014-4056-4.

Dubois, Maison-Blanche, Quenet, and Dreyfus, 2007. Rémi Dubois, Pierre Maison-Blanche, Brigitte Quenet, and Gérard Dreyfus. *Automatic ECG wave extraction in long-term recordings using Gaussian mesa function models and nonlinear probability estimators.* Computer Methods and Programs in Biomedicine, 88(3), 217–233, 2007. ISSN 01692607. doi: 10.1016/j.cmpb.2007.09.005.

Eftestol, Aase, and Husoy, 1998. T. Eftestol, S.O. Aase, and J.H. Husoy. *A flexible pattern recognition system for analysis of ECG and related demographics and annotations.* Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol.20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No.98CH36286), 20(1), 135–138, 1998. doi: 10.1109/IEMBS.1998.745848. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=745848.

Elbuni, Kanoun, Elbuni, and Ali, 2009. Abdulrhman Elbuni, Salah Kanoun, Mohamed Elbuni, and Nasser Ali. *ECG parameter extraction algorithm using (DWTAE) algorithm.* ICCTD 2009 - 2009 International Conference on Computer Technology and Development, 2, 57–62, 2009. doi: 10.1109/ICCTD.2009.103.

Elgendi, 2013. Mohamed Elgendi. *Fast QRS Detection with an Optimized Knowledge-Based Method: Evaluation on 11 Standard ECG Databases.* PLoS ONE, 8(9), 2013. ISSN 19326203. doi: 10.1371/journal.pone.0073557.

Gelernter, 1991. David Gelernter. *Mirror Worlds or The Day Software put the Universe in a Shoebox.* 1991. ISBN 019507906X.

Gilroy, Ross, and MacPerason, 2008. Anne M. Gilroy, Lawrence M. Ross, and Brian R. MacPerason. *Atlas of anatomy.* 2008. ISBN 978-1-60406-099-7.

Gregg, Zhou, and Babaeizadeh, 2016. Richard E Gregg, Sophia H Zhou, and Saeed Babaeizadeh. *Finding Similar ECGs in a Large 12-lead ECG Database Introduction.* 43, 2016. doi: 10.22489/CinC.2016.086-171.

Hawkins, Ahmad, and Dubinsky, 2011. Jeff Hawkins, Subutai Ahmad, and Donna Dubinsky. *Cortical Learning Algorithm and Hierarchical Temporal Memory.* Numenta Whitepaper, pages 1–68, 2011. URL http://numenta.org/resources/HTM{_}CorticalLearningAlgorithms.pdf.

He, Clifford, and Tarassenko, 2001. Taigang He, Gari Clifford, and Lionel Tarassenko. *Application of ICA in Removing Artefacts from the ECG.* Neural Networks, pages 1–18, 2001. ISSN 0941-0643. doi: 10.1007/s00521-005-0013-y.

Holzinger, 2016. Andreas Holzinger. *Machine Learning for Health InformaticsState-of-the-Art and Future Challenges.* Springer, 2016. ISBN 978-3-319-50477-3. doi: 10.1007/978-3-319-50478-0.

Jacobson and Dalianis, 2016. Olof Jacobson and Hercules Dalianis. *Applying deep learning on electronic health records in Swedish to predict healthcare-associated infections.* Acl 2016, page 191, 2016.

Jadhav, Nalbalwar, and Ghatol, 2012. Shivajirao Jadhav, Sanjay Nalbalwar, and Ashok Ghatol. *Performance Evaluation of Multilayer Perceptron Neural Network Based Cardiac Arrhythmia Classifier.* International Journal of Computer Engineering and Technology (IJCET), 3(2), 1–11, 2012.

Kalkstein, Kinar, Na'aman, Neumark, and Akiva, 2011. Nir Kalkstein, Yaron Kinar, Michael Na'aman, Nir Neumark, and Pini Akiva. *Using machine learning to detect problems in ECG data collection.* 2011 Computing in Cardiology (CinC), pages 437–440, 2011. ISSN 0276-6547.

Khandoker, Palaniswami, and Karmakar, 2009. Ahsan H. Khandoker, Marimuthu Palaniswami, and Chanda K. Karmakar. *Support vector machines for automated recognition of obstructive sleep apnea syndrome from ECG recordings*, 2009. ISSN 10897771.

Khunti, 2013. K. Khunti. *Accurate interpretation of the 12-lead ECG electrode placement: A systematic review.* Health Education Journal, 73(5), 610–623, 2013. ISSN 0017-8969. doi: 10.1177/0017896912472328. URL http://hej.sagepub.com/cgi/doi/10.1177/0017896912472328.

Kora and Sri Rama Krishna, 2016. Padmavathi Kora and K. Sri Rama Krishna. *ECG Based Heart Arrhythmia Detection Using Wavelet Coherence and Bat Algorithm.* Sensing and Imaging, 17(1), 1–16, 2016. ISSN 15739317. doi: 10.1007/s11220-016-0136-5.

Rüdiger Kramme, Klaus-Peter Hoffmann, and Robert Pozos. Functional diagnostics devices. In *Springer Handbook of Medical Technology.* 2011.

Kumari and Sadasivam, 2007. R. Shantha Selva Kumari and V. Sadasivam. *A novel algorithm for wavelet based ECG signal coding.* Computers and Electrical Engineering, 33(3), 186–194, 2007. ISSN 00457906. doi: 10.1016/j.compeleceng.2006.02.001.

Liu, Jin, and Li, 2015. Hui Liu, Xin Yu Jin, and Lan Juan Li. *Practical Method for ECG Classification Using Weighted ELM.* Applied Mechanics and Materials, 738-739, 832–835, 2015. ISSN 1662-7482. doi: 10.4028/www.scientific.net/AMM.738-739.832. URL http://www.scientific.net/AMM.738-739.832.

Lu, Yan, and Silva, 2016. Lei Lu, Jihong Yan, and Clarence W. de Silva. *Feature Selection for ECG Signal Processing Using Improved Genetic Algorithm and Empirical Mode Decomposition.* Measurement, 94, 372–381, 2016. ISSN 02632241. doi: 10.1016/j.measurement.2016.07.043. URL http://dx.doi.org/10.1016/j.measurement.2016.07.043.

Martini, Nath, and Bartholomew, 2012. Frederich H Martini, Judi L. Nath, and Ewdin F. Bartholomew. *Fundamentals of Anatomy & Physiology.* Pearson, 2012. ISBN 0321735536.

Masetic and Subasi, 2016. Zerina Masetic and Abdulhamit Subasi. *Congestive heart failure detection using random forest classifier.* Computer Methods and Programs in Biomedicine, 130, 54–64, 2016. ISSN 18727565. doi: 10.1016/j.cmpb.2016.03.020. URL http://dx.doi.org/10.1016/j.cmpb.2016.03.020.

Numenta, 2017. Numenta. *NuPIC 0.7.0.dev0 API Documentation*, 2017. URL http://nupic.docs.numenta.org/index.html.

Oster, Behar, Sayadi, Nemati, Johnson, and Clifford, 2015. Julien Oster, Joachim Behar, Omid Sayadi, Shamim Nemati, Alistair E W Johnson, and Gari D. Clifford. *Semisupervised ECG Ventricular Beat Classification with Novelty Detection Based on Switching Kalman Filters*, 2015. ISSN 15582531.

Pan and Tompkins, 1985. Jiapu Pan and Willis J. Tompkins. *A Real-Yime QRS Detection Algorithm.* IEEE transactions on bio-medical engineeringBiomedical Engineering, BME-32(3), 230–236, 1985. ISSN 0018-9294. doi: 10.1109/TBME.1985.325532. URL http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=4122029.

Rahhal, Bazi, Alhichri, Alajlan, Melgani, and Yager, 2016. M. M Al Rahhal, Yakoub Bazi, Haikel Alhichri, Naif Alajlan, Farid Melgani, and R. R. Yager. *Deep learning approach for active classification of electrocardiogram signals.* Information Sciences, 345, 340–354, 2016. ISSN 00200255. doi: 10.1016/j.ins.2016.01.082. URL http://dx.doi.org/10.1016/j.ins.2016.01.082.

Seera, Lim, Liew, Lim, and Loo, 2015. Manjeevan Seera, Chee Peng Lim, Wei Shiung Liew, Einly Lim, and Chu Kiong Loo. *Classification of electrocardiogram and auscultatory blood pressure signals using machine learning models.* Expert Systems with Applications, 42(7), 3643–3652, 2015. ISSN 09574174. doi: 10.1016/j.eswa.2014.12.023. URL http://dx.doi.org/10.1016/j.eswa.2014.12.023.

Shadmand and Mashoufi, 2016. Shirin Shadmand and Behbood Mashoufi. *A new personalized ECG signal classification algorithm using Block-based Neural Network and Particle Swarm Optimization.* Biomedical Signal Processing and Control, 25, 12–23, 2016. ISSN 17468108. doi: 10.1016/j.bspc.2015.10.008. URL http://dx.doi.org/10.1016/j.bspc.2015.10.008.

Shin, Kim, Song, Lee, Kim, and Jeung, 2015. Seung Won Shin, Kyeong Seop Kim, Chul Gyu Song, Jeong Whan Lee, Jeong Hwan Kim, and Gyeo Wun Jeung. *Removal of baseline wandering in ECG signal by improved detrending method.* Bio-medical materials and engineering, 26, S1087–S1093, 2015. ISSN 18783619. doi: 10.3233/BME-151405.

Silbernagl and Despopoulos, 2009. Stefan Silbernagl and Agamemnon Despopoulos. *Color Atlas of Physiology.* 2009. ISBN 9783135450063.

Strasburger, Huhta, Carpenter, Garson, and McNamara, 1986. J F Strasburger, J C Huhta, R J Carpenter, A Garson, and D G McNamara. *Doppler echocardiography in the diagnosis and management of persistent fetal arrhythmias.* Journal of the American College of Cardiology, 7(6), 1386–91, 1986. ISSN 0735-1097. doi: 10.1016/S0735-1097(86)80161-9. URL http://www.ncbi.nlm.nih.gov/pubmed/3711496.

Sufi, Mahmood, and Khalil, 2009. Fahim Sufi, Abdun Mahmood, and Ibrahim Khalil. *Diagnosis of cardiovascular abnormalities from compressed ECG: A data mining based approach.* 2009 9th International Conference on Information Technology and Applications in Biomedicine, 15(1), 1–4, 2009. doi: 10.1109/ITAB.2009.5394350. URL http://www.scopus.com/inward/record.url?eid=2-s2.0-77949636123{&}partnerID=tZOtx3y1.

Tyagi and Thakur, 2015. Kratika Tyagi and Sanjeev Thakur. *A Survey On Various Data Mining Techniques for ECG Meta Analysis.* International Journal of Science, Engineering and Technology Researc, 4(5), 2015. URL http://ijsetr.org/wp-content/uploads/2015/05/IJSETR-VOL-4-ISSUE-5-1241-1245.pdf.

Van Boxtel, 2001. Anton Van Boxtel. *Optimal signal bandwidth for the recording of surface EMG activity of facial, jaw, oral, and neck muscles.* Psychophysiology, 2001. doi: 10.1111/1469-8986.381002.

WHO, 2014. WHO. *Global status report on noncommunicable diseases 2014.* World Health, page 176, 2014. ISSN 1475-5785. doi: ISBN9789241564854.

Wick, Inan, Mcclellan, and Cardiac-computed, 2015. Carson A Wick, Omer T Inan, James H Mcclellan, and Abstract Cardiac-computed. *Seismocardiography-Based Detection of Cardiac Quiescence.* 62(8), 2025–2032, 2015.

Xu, Liu, Hao, Wen, and Huang, 2010. Ya Xu, Guangyuan Liu, Min Hao, Wanhui Wen, and Xiting Huang. *Analysis of affective ECG signals toward emotion recognition.* Journal of Electronics, 27(1), 8–14, 2010. ISSN 02179822. doi: 10.1007/s11767-009-0094-3.

Yeh, Wang, and Chiou, 2010. Yun-Chi Yeh, Wen-June Wang, and Che Wun Chiou. *Feature selection algorithm for ECG signals using Range-Overlaps Method.* Expert Systems with Applications, 37(4), 3499–3512, 2010. ISSN 09574174. doi: 10.1016/j.eswa.2009.10.037. URL http://dx.doi.org/10.1016/j.eswa.2009.10.037.

Zygmunt and Stanczyk, 2010. Agnieszka Zygmunt and Jerzy Stanczyk. *Methods of evaluation of autonomic nervous system function.* Archives of Medical Science, 6(1), 11–18, 2010. ISSN 17341922. doi: 10.5114/aoms.2010.13500.