Tampere University

Manu Joronen

PERSONALIZED ECG MONITORING

by sparse representation based feature extraction

Faculty of Information Technology and Communication Sciences
Bachelor's thesis
December 2019

ABSTRACT

Manu Joronen: Personalized ECG monitoring
Bachelor's thesis
Tampere University
Degree Programme in Computing and Electrical Engineering, BSc (Tech)
December 2019

Arrhythmia is a condition that may appear at any-time and anywhere. Unfortunately, the consequences are as well sometimes very serious for the suffering person. Therefore, the studies about automated ECG monitoring have gained a lot of attraction in the literature. This study focuses on the personalized monitoring aspect, which has been rarely investigated. In this thesis, we propose a novel sparse representation based feature extraction methods to be used with one class classification, which has potential to outperform the competing methods. To evaluate the performance, we run experiments with data from the MIT-BIH database and find that the proposed methods occasionally surpass the performance of the method that does not use the proposed feature extraction procedures and one class classification.

Keywords: anomaly detection, one-class classification, personalized ECG monitoring, ECG anomaly detection

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

TIIVISTELMÄ

Manu Joronen: Henkilökohtainen EKG-monitorointi
Kanditaatintyö
Tampereen yliopisto
Tieto- ja sähkötekniikan TkK-tutkinto-ohjelma
Joulukuu 2019

Sydämen rytmihäiriö voi kehittyä missä ja milloin vain. Rythmihäiriön seuraamukset voivat olla vakavia. Automatisoitu EKG-monitorointi onkin saanut paljon huomiota kirjallisuudessa. Tämä työ keskittyy automatisoituun henkilökohtaiseen EKG-monitorointiin, jota ei ole käsitelty kirjallisuudessa läheskään niin paljon kuin ei-henkilökohtaista EKG-monitorointia. Esitämme tässä työssä ratkaisuksi menetelmiä, jotka perustuvat uusiin harvaan esitykseen perustuviin ominaisuuksiin. Pyrimme ehdotetuilla menetelmillä parantamaan jo olemassa olevan menetelmän tehokkuutta. Vertailemme kaikkia menetelmiä tutkimuksissa, joissa menetelmien tehokkuutta mitataan MIT-BIH -tietokannan datalla. Tutkimuksissa ilmeni ehdotettujen menetelmien voivan osittain parantaa jo olemassa olevan menetelmän tehokkuutta.

Avainsanat: anomalian tunnistus, yhden luokan luokittelu, henkilökohtainen EKG-monitorointi, EKG-anomalian tunnistus

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -ohjelmalla.

PREFACE

This thesis was written in Tampere University during the fall semester of 2019. The topic was provided by my supervisor Mehmet Yamac and I made the decision to pick it based on my interests in anomaly detection and automated ECG monitoring. I would like to thank Mehmet Yamac for being my supervisor and also for providing this interesting topic besides the great ideas on the methods used in this thesis. Additionally, I would like to thank Fahad Sohrab for teaching me about one-class classifiers and my companion for mental support.

Tampere, 18th December 2019

Manu Joronen

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

AAMI	Association for the Advancement of Medical Instrumentation
ACC	Accuracy
AD	Anomaly detection
AUC	Area under the curve
ECG	Electrocardiogram
F1	F1 score
FN	False negative
FP	False positive
FPR	False positive rate
K-SVD	K-singular value decomposition
MIT-BIH	Physionet's open arrhythmia database
OCSVM	One-class support vector machine
OMP	Orthogonal matching pursuit
PRE	Precision
RBF	Radial basis function
ROC	Receiver operating characteristic
SEN	Sensitivity
SPE	Specificity
TN	True negative
TP	True positive
TPR	True positive rate

1 INTRODUCTION

Automated electrocardiogram (ECG) monitoring by machine learning methods is covered widely in the literature. Though, most of the study focuses on the ECG classification problem, which requires annotated data with different types of arrhythmias for the training phase. As a consequence, ECG classification is only practical with both non-person specific settings and hospital environments. Instead, this research focuses on the personalized ECG anomaly detection (AD) problem with wearable devices. Moreover, methods used in this study enables person-specific monitoring since AD algorithms require only data of normal heartbeats for the training phase. The proposed methods in this research are not computationally intensive and so are reasonable to be computed by a wearable device.

In the prior related work, a lot of different machine learning methods have been applied for solving the automated ECG monitoring problem. Accordingly, there are many different approaches for solving the ECG classification problem. These methods include hand crafted feature extraction based methods [1][2] as well as methods based on the data-driven feature extraction [3][4]. Additionally, the ECG AD problem is also covered in the literature. [5][6] approaches the problem with neural networks while [7][8] studies sparse representation based methods. Similarly to our research, [9] uses K-singular value decomposition (K-SVD) not only for the AD but also for the feature extraction. Certain issues may also arise during long term ECG monitoring by wearable devices. However, some solutions for long term monitoring exists, i.e., solutions for K-SVD based methods [10].

This thesis aims to improve the ECG AD method proposed in [7] by using more extensive feature extraction methods. A Somewhat similar way of feature extraction compared to our methods is proposed in [9]. Consequently, K-SVD and orthogonal matching pursuit (OMP) [11] algorithms are used for the person specific feature extraction. Then we try to detect anomalous heartbeats by using either single or multiple extracted features. The detection step is known as an AD problem in the literature, although AD is often also referred as a one-class classification or as a novelty detection problem. Fortunately, many solutions have been discovered by the research. In this study, we adopt a famous one-class support vector machine (OCSVM) with a Gaussian radial basis function (RBF) kernel [12] to solve the AD problem. Both are chosen for our study due to their performance and reasonable computational requirements [13]. Finally, to compare the method [7] and the proposed methods in this study, we evaluate the performance of each method in several experiments.

In Chapter 2, we explain ECG briefly, introduce how we utilize the MIT-BIH arrhythmia database [14], and define our methods for the ECG signal preprocessing and hearbeat segmentation. Following that, in Chapter 3, we address the theory of the methods used in this study. The methods consist of sparse representation, feature extraction and one-class classification. Next, in Chapter 4, the figures of merit, experimental setup and the results of the experiment are presented. Finally, in Chapter 5 we draw the conclusions of this study and discuss the future work.

2 ECG DATA

Electrocardiogram, more commonly known as ECG, is a record of the electrical potential difference between determined points as a function of time. Such a record can be measured in practice by placing electrodes on the skin of the patient. One commonly used configuration is a setup of 10 electrodes, which allows recording 12 different types of leads. One of the leads named as lead II is essential for the automatic arrhythmia detection and classification problem. [15]

ECG monitoring is mostly utilized in a hospital environment. However, ECG monitoring elsewhere than in the hospital environment is, in fact, possible with mobile devices. Though automatic heartbeat classification methods designed for the hospital environment can not be utilized straightforwardly for mobile device use cases since using such devices for long term monitoring raises issues to overcome. The morphology of the ECG signal recorded with such a mobile device might reshape through time, for example, in cases if the user misplaces the device or if the device gets misplaced by the user's movement. [7]

2.1 MIT-BIH

In this research, the arrhythmia database MIT-BIH [14] will be used for the experiment. It consists of 48 two-channel ECG records, each from different patients. All the records are approximately 30-minutes-long and digitized with a 360 Hz sampling rate. For most of the records these channels are modified lead V1 and modified limb lead II.

Table 2.1. *Comparison of labeling of MIT-BIH and AAMI heartbeat classes and how they corresbond to classes used in this study*

MIT-BIH	AAMI	Normal (N) / Abnormal (A)
N	N	N
L	N	N
R	N	N
e	N	N
j	N	N
A	SVEB	A
a	SVEB	A
J	SVEB	A
S	SVEB	A
V	VEB	A
E	VEB	A
F	F	A
P	Q	A
f	Q	A
U	Q	A

For the ECG classification problem the AAMI standard [16] recommends dividing heartbeats into five different classes. Instead, for purposes of this research those five classes will be again divided into two classes: normal and abnormal. Table 2.1 illustrates how those two classes correspond to the MIT-BIH and AAMI classes.

Table 2.2. *All the patients from the MIT-BIH database mapped to the datasets 1 and 2. One patient may either be part of the validation set (VAL), part of the test set (TES) or excluded from the dataset (EXC). The columns are named accordingly Patient (PAT), dataset 1 (DS1) and dataset 2 (DS2).*

PAT	DS1	DS2	PAT	DS1	DS2	PAT	DS1	DS2	PAT	DS1	DS2
100	EXC	VAL	113	EXC	VAL	201	TES	TES	217	EXC	EXC
101	EXC	VAL	114	VAL	VAL	202	TES	TES	219	TES	TES
102	EXC	EXC	115	EXC	VAL	203	TES	TES	220	TES	TES
103	EXC	VAL	116	VAL	VAL	205	TES	TES	221	TES	TES
104	EXC	EXC	117	EXC	VAL	207	TES	TES	222	TES	TES
105	EXC	VAL	118	VAL	VAL	208	TES	TES	223	TES	TES
106	VAL	VAL	119	VAL	VAL	209	TES	TES	228	TES	TES
107	EXC	EXC	121	EXC	VAL	210	TES	TES	230	EXC	TES
108	EXC	VAL	122	EXC	VAL	212	EXC	TES	231	EXC	TES
109	EXC	VAL	123	EXC	VAL	213	TES	TES	232	EXC	EXC
111	EXC	VAL	124	VAL	VAL	214	TES	TES	233	TES	TES
112	EXC	VAL	200	TES	TES	215	TES	TES	234	TES	TES

In this research we divide the MIT-BIH database patients into two different datasets, Dataset 1 and Dataset 2, where Dataset 1 is a subset of Dataset 2. As suggested in [2], patients 102, 104, 107 and 217 will be excluded from both datasets due to the paced heartbeats. Moreover, all the patients with less than 550 normal heartbeats or 50 abnormal heartbeats are excluded from Dataset 1. The patient 232 is excluded from both datasets since its record does not contain 500 normal heartbeats which is the size of the patient specific training set used in our experiments. Thus, the two explained datasets meet all the requirements set by the evaluation metrics explained in Chapter 4. Table 2.2 illustrates the described setting and also presents how the patients are divided into the validation–test subsets.

2.2 Preprocessing

In experiments the preprocessing step will be the same as in [2], as also suggested in [15]. In total, it consists of three steps, two median filters and one low-pass filter. The signal is filtered by the median filter with a width of 200 ms following by another one 600 ms wide. The resulting signal is then extracted from the original signal. Finally, the signal is filtered with finite 12-tap low-pass filter. More precisely, the filter is designed to have a 3-dB cutoff frequency at 35 Hz and an equal ripple for both pass and stopband.

Figure 2.1. *The signal before (blue) and after (orange) the preprocessing*

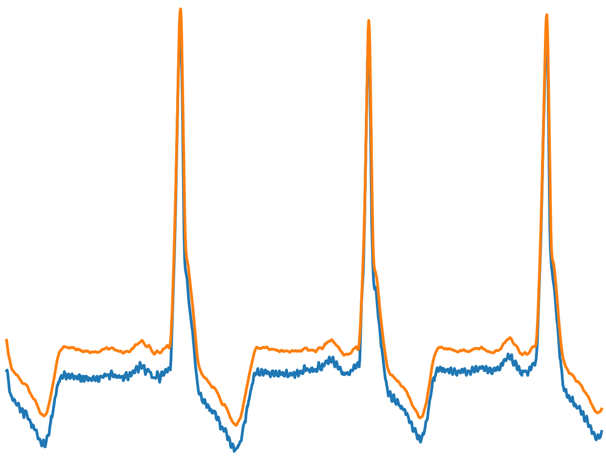


Figure 2.1 demonstrates the effect of the explained preprocessing step.

2.3 Heartbeat segmentation

After preprocessing, the signals are split into segments. Each segment being R-peak centered with a length of 0.6 seconds, there are 216 samples per each segment since the sampling frequency is 360 Hz.

MIT-BIH database contains handcrafted location annotations for the R-peaks. However, such handcrafted location marks are practically not available. Fortunately, this problem is already solved and some of the methods are listed in [15]. Although, these kinds of algorithms always suffer from some level of error. In addition to the R-peak annotations provided by the MIT-BIH, we also run experiments in which additional artificial jitter is applied to the R-peak annotations. Thus, we also consider the robustness of the anomaly detection methods in this experiment. As suggested in [1][15], the jitter will be generated randomly with zero mean and standard deviation of five samples.

3 METHODS

In this chapter the proposed methods for detecting the anomalous heartbeats will be explained. All the methods studied in this thesis consist of either two or three different steps. The steps are sparse representation, feature extraction and one-class classification. The latter step may not be necessary if only a single feature is utilized for detecting anomalous heartbeats.

3.1 Sparse representation

The heartbeat signal after the preprocssing and segmentation $s \in \mathbf{R}^p$ will be classified as either normal s_n or abnormal s_a. Dictionary $D \in \mathbf{R}^{p \times n}$ is a matrix which consists of n so-called atoms. With s and D we may produce a sparse code vector $x \in \mathbf{R}^n$ for sparse representation. Consequently, if the D and x suits well for the s, we obtain:

$$s \approx Dx \tag{3.1}$$

In this research we measure the sparsity of the x by l_0 norm. Thus, we can limit the sparsity:

$$\|x\|_0 \leq \kappa \tag{3.2}$$

where the $\kappa \in \mathbf{N}$. The reconstruction error may be measured by using the l_2 norm, denoted by:

$$r = \|s - Dx\|_2 \tag{3.3}$$

Obtaining such an x for s and D, which minimizes the r is known as sparse coding. Thus, an optimization problem may be formulated in which the sparsity of x is also limited accordingly to the equation 3.2:

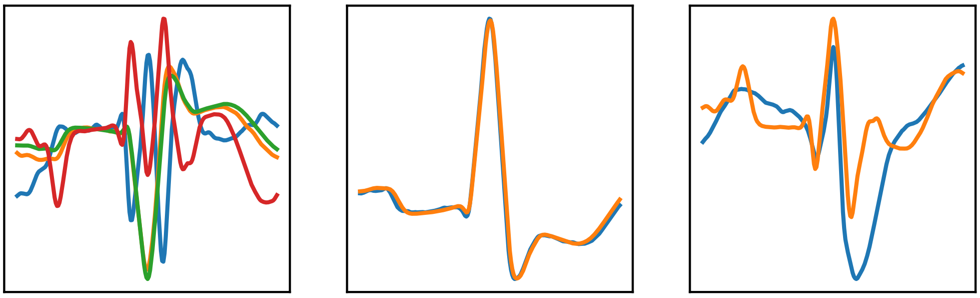
$$\min_x \|s - Dx\|_2, s.t.\|x\|_0 \leq \kappa \tag{3.4}$$

Additionally, to be able to find an optimal solution for the equation 3.4, finding a suitable D is as well necessary. In this study we produce the optimal D by a process which is known as dictionary learning. Similarly to the equation 3.4, the dictionary learning is denoted by an optimization problem:

$$\min_{D,X} \|S - DX\|_2, s.t. \|x_i\|_0 \leq \kappa, i = 1, ..., m \tag{3.5}$$

where the $S \in \mathbf{R}^{p\times m}$ and $X \in \mathbf{R}^{n\times m}$. In this study, the S refers to the training set which consists of m normal heartbeat signals s_n of a specific user. Therefore, we expect the r to be smaller for s_n than for s_a if the x and person specific dictionary D are produced by the described sparse coding and dictionary learning. The figure 3.1 demonstrates how the sparse approximation is more similar to normal heartbeat than to abnormal one for the learned dictionary.

Figure 3.1. *At left, all the dictionary atoms are plotted. In the middle, a normal heartbeat (blue) is plotted beside the according sparse approximation (orange). The right plot is similar to the middle one with the exception that the heartbeat is abnormal instead. The chosen parameters were* $n = 4$ *and* $\kappa = 2$.



Similarly to the [7], we chose the K-SVD and orthogonal matching pursuit (OMP) algorithms for the dictionary learning and sparse coding tasks. The algorithms are explained precisely in [17]. Basically, the K-SVD algorithm iterates k times through a given initial dictionary D_0 and the training set S. For our experiments, we use the random uniform distribution to generate the D_0 and then normalize its atoms to have unit l_2-length. Then, we chose the first 500 normal heartbeats from the patient as the S and set the k to 50. The best performing κ and n at the validation step are then chosen for the final evaluation. The validation of κ and n are discussed more in Chapter 4.

3.2 Feature extraction

In this section, we introduce all the feature extraction methods. Nonetheless, the feature combinations chosen for the experiments are not discussed here. Those combinations are explained in Section 4. For convenience, all the considered features extraction methods are marked into a list:

i. The reconstruction error r from the equation 3.3.

ii. The l_1 norm of the sparse code vector: $\|x\|_1$ where the x is obtained from the equation 3.4.

iii. Kurtosis of x which is obtained form the equation 3.4. Denoted by: μ_4/σ^4, where μ_4 and σ^4 are the fourth central moment and the standard deviation of x.

iv. The sparse code vector x itself from the equation 3.4.

v. First, the x is obtained from the equation 3.4. Then the s and D are divided into α equal size segments and the reconstruction error from equation 3.3 is calculated per segment: $r_i = \|s_i - D_i x\|_2, i = 1, ..., \alpha$. Unlike with the feature extraction method 1, α amount of features are received.

Depending on the amount of atoms n and the value of α, features extraction methods iv and v may result more than just a single feature. Instead, the methods i, ii and iii will always result a single feature.

The methods i and ii have been already discovered in the literature for the AD problem. In the prior studies, the method i has been the most usual way to detect anomalies [7][5]. The method ii was proposed in [9], although, they measured the sparsity of the sparse vector x by l_1 norm. Additionally, the method iv was proposed to solve the ECG classification problem in [4].

3.3 One-class classification

Detection of anomalous heartbeats based only on a single feature is very straightforward. Only a threshold τ value has to be set when the heartbeats may be classified as a normal if $\leq \tau$ and as an abnormal otherwise. However, in our experiments we also study the effect of utilizing more than just a single feature for the AD task. Classifying such multivariate feature sets is possible by one-class classification algorithms. For our experiments, we chose the OCSVM algorithm with the Gaussian RBF kernel [12].

The OCSVM is very similar algorithm to the support vector machine (SVM) [18]. While the SVM aims to find a hyperplane which maximizes the margin between two different classes, the OCSVM aims to find a hyperplane which separates the single class from the origin maximizing the distance between the origin and the hyperplane. The Gaussian RBF is denoted by:

$$K(x, x') = exp(-\gamma \|x - x'\|^2) \tag{3.6}$$

where the $\gamma \in \mathbf{R}$ is situation specific parameter. The OCSVM algorithm finds an optimal hyperplane when it is given: training data, kernel function K and parameter $\nu \in (0, 1)$. The ν parameter controls two things: it sets an upper bound for the amount of outliers in the training data and a lower bound for the amount of chosen support vectors. Then with the obtained hyperplane it is possible to classify the sample as a normal (1) or as an abnormal (-1) by:

$$f(x) = sgn(\sum_i \alpha_i K(x_i, x) - \rho) \tag{3.7}$$

In the equation 3.7 the α_i with the ρ characterizes the hyperplane. However, the learned hyperplane may not always be the most optimal one for the intended task. Thus, we define an anomaly score function which may then be used to classify the anomalies:

$$s(x) = -(\sum_i \alpha_i K(x_i, x) - \rho) \tag{3.8}$$

The bigger the anomaly socre, the more anomal the x is considered. With the defined anomaly score function it is also possible to do the classification in a similar fashion compared to the methods based on a single feature, only the threshold τ being required.

4 EXPERIMENTS

First, we cover the figures of merit used in our experiments. Then two different experiments are explained, followed by their results.

4.1 Figures of merit

For performance evaluation, the following metrics are introduced. True positive (TP) means classifying abnormal class correctly, whereas false positive (FP) is about classifying abnormal class incorrectly. Similarly, true negative (TN) and false negative (FN) are about correct and incorrect predictions for the normal class.

Accordingly, the true positive rate (TPR) is defined as

$$TPR = TP/(TP + FN) \tag{4.1}$$

and false positive rate (FPR) is:

$$FPR = FP/(FP + TN), \tag{4.2}$$

The receiver operating characteristic (ROC) curve may thus be explained. To obtain the curve, TPR and FPR have to be plotted as a function of threshold which varies from the smallest possible threshold to the biggest one. Using the FPR as a x-coordinate and TPR as a y-coordinate the curve will settle between points (0,0) and (1,1).

Finally, we define the area under the curve (AUC). Basically the AUC is the area between the curve and the x-axis. Thus, the best possible AUC score is 1 while the worst is 0. Although, it should be noted that a random classifier would achieve AUC score of 0.5. In this experiment the AUC score is always taken from the area under the ROC.

The AUC score depends on how well the method's anomaly indicator separates the normal class from the abnormal class. Unlike the AUC score, the rest of the metrics to be introduced depend on the actual class predictions. Accuracy (ACC) is defined as

$$ACC = (TP + TN)/(TP + FP + TN + FN) \tag{4.3}$$

whereas precision (PRE) is

$$PRE = TP/(TP + FP), \tag{4.4}$$

specificity (SPE) is

$$SPE = TN/(TN + FP), \tag{4.5}$$

and F1 score (F1)

$$F_1 = 2TP/(2TP + FP + FN). \tag{4.6}$$

Also, the TPR from the equation 4.1 is also called as sensitivity (SEN).

4.2 Experimental setup

This research will cover 9 different methods in 2 different experiments. The methods mostly differ on their feature extraction part. Thus, for the convenience we list all the feature extraction combinations:

1. i
2. iv
3. i and ii
4. i and iii
5. i and iv
6. i , ii and iii
7. i , ii , iii and iv
8. v where the α is set to 4
9. v where the α is set to 8

Method i [7] produces only a single feature which will be considered as the anomaly score. The rest of the methods produce multivariate feature sets where the proposed OCSVM algorithm can be applied. For the multivariate methods we also obtain the anomaly score but from Equation 3.8 instead. Therefore, a simple thresholding method suits for all the proposed methods. Lets also remark that with the threshold τ, we consider heartbeats with anomaly scores $\leq \tau$ as normal ones and the rest as anomalous.

Next, we introduce the 2 different experiments where the 9 different methods are evaluated. For both experiments, the training procedure for the dictionary and the for OCSVM are the same. Since the main focus on this study is on the personalized setting, we use patient specific training datasets. Each training set consists of the first 500 normal heartbeats of the particular patient. Hence, the rest of the patient's heartbeats are for the evaluation. Also, the covered parameter combinations for the sparse representation and the OCSVM are the same in both of the experiments. Parameters $n = \{4, 8, 16, 32\}$ and $\kappa = \{2, 4, 8\}$ where $\kappa \leq n$ are covered for the sparse representation step. Similarly for the methods which utilize the OCSVM, we cover parameters $\nu = \{0.1, 0.01\}$ and $\gamma = \{2^{-6}, 2^{-5}, ..., 2^{1}\}$. The best proposed parameter combination is found for each method by evaluating the performance against the validation dataset. Then the final performance is measured based on the test dataset by using the parameter combination found from the validation step. Finally, we take the random nature and the tolerance against the error of the R-peak detection algorithm into consideration. So, all of the experiments are proceeded seven times and the average and standard deviation over all the results are reported. Moreover, to test the tolerance against the jitter, we report two different results for both of the experiments, with and without the described artificial jitter.

In the first experiment, instead of classifying the heartbeats, we obtain the anomaly scores for each method. This enables comparison of the methods based on their AUC score. All the anomaly scores are user-specific, therefore we measure the AUC score patient specifically. To be able to evaluate the performance, we then take the average over the patient-specific AUC scores. Dataset 1 defined in Chapter 2 is utilized in this experiment. All the patients in Dataset 1 have at least 50 normal and abnormal heartbeats in their records after the 500 normal heartbeats for the training are excluded. In this study the minimum requirement for the patient's heartbeats is decided to exclude the biased ROC curves.

The second experiment measures how the proposed methods perform based on the F1, ACC, PRE, SPE and SEN metrics. Moreover, we use the F1 score in the validation step to find the best parameter combination. Unlike in the first experiment, Dataset 2 is used and the metrics are not measured patient specifically. Since the metrics covered in this experiment are based on the class predictions instead of the anomaly score, a suitable threshold τ has to be found for each patient. To find the optimal threshold, we gathered 500 abnormal heartbeats randomly from the whole dataset excluding the targeted patient's record. Then the optimal threshold is settled to be the one that maximizes the F1 score across the gathered 500 abnormal heartbeats and the 500 normal from the patient's training set.

4.3 Results

Table 4.1 presents the results for both experiments where there was no artificial jitter applied to the heartbeat segmentations.

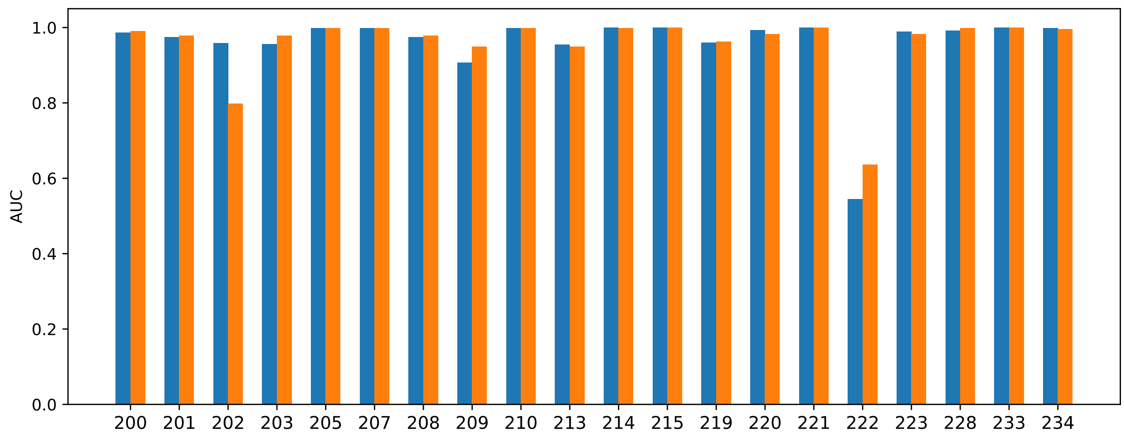
Table 4.1. *All the introduced metrics for all the methods from the experiments without artificial jitter. The figures are the average and standard deviation of 7 repetition.*

Method	AUC	F1	ACC	SEN	SPE	PRE
1 [7]	$\mathbf{0.959 \pm 0.001}$	0.733 ± 0.015	0.885 ± 0.008	0.944 ± 0.010	0.873 ± 0.009	0.599 ± 0.018
2	0.957 ± 0.003	0.762 ± 0.006	0.904 ± 0.004	0.918 ± 0.009	0.901 ± 0.006	0.651 ± 0.011
3	0.954 ± 0.003	0.784 ± 0.007	0.916 ± 0.003	0.912 ± 0.006	0.917 ± 0.004	0.688 ± 0.010
4	0.952 ± 0.002	0.718 ± 0.032	0.890 ± 0.018	0.836 ± 0.006	0.900 ± 0.022	0.632 ± 0.054
5	$\mathbf{0.959 \pm 0.003}$	$\mathbf{0.792 \pm 0.008}$	$\mathbf{0.920 \pm 0.004}$	0.913 ± 0.008	$\mathbf{0.922 \pm 0.005}$	$\mathbf{0.700 \pm 0.013}$
6	0.956 ± 0.002	0.778 ± 0.006	0.913 ± 0.003	0.915 ± 0.005	0.912 ± 0.004	0.676 ± 0.001
7	0.958 ± 0.003	0.789 ± 0.009	0.918 ± 0.004	0.916 ± 0.011	0.919 ± 0.006	0.693 ± 0.014
8	0.945 ± 0.002	0.725 ± 0.007	0.890 ± 0.004	0.868 ± 0.004	0.894 ± 0.004	0.622 ± 0.001
9	0.950 ± 0.004	0.755 ± 0.012	0.897 ± 0.006	$\mathbf{0.949 \pm 0.008}$	0.886 ± 0.007	0.626 ± 0.015

It seems that the method 5 specially outperforms the other methods on the second experiment.

To compare the method 1 [7] to the one that achieves the highest AUC-score from the others, we also present the figure 4.1. In the figure we plot the patient-specific AUC-scores for the methods 1 [7] and 5.

Figure 4.1. *Person-specific AUC-scores from the experiment without artificial jitter. The blue bars present the method 1 [7] while the orange bars present the method 5. The scores are the average of 7 repetition.*



According to the figure 4.1, it should be noted that the best method may vary among the patients.

Finally, to address the robustness against the errors from the heartbeat segmentation methods, we also present the results of experiments where there is the described artificial jitter applied. The table 4.2 has the results in the same format as in the table 4.1.

Table 4.2. *The figures are produced similarly to the table 4.1 except from the experiments with the artificial jitter.*

Method	AUC	F1	ACC	SEN	SPE	PRE
1 [7]	0.953 ± 0.001	0.706 ± 0.009	0.868 ± 0.006	$\mathbf{0.950 \pm 0.005}$	0.852 ± 0.008	0.562 ± 0.012
2	0.956 ± 0.003	0.709 ± 0.012	0.878 ± 0.008	0.891 ± 0.010	0.875 ± 0.011	0.589 ± 0.019
3	0.940 ± 0.005	$\mathbf{0.712 \pm 0.032}$	$\mathbf{0.885 \pm 0.018}$	0.844 ± 0.009	$\mathbf{0.894 \pm 0.022}$	$\mathbf{0.617 \pm 0.049}$
4	0.942 ± 0.003	0.633 ± 0.032	0.844 ± 0.022	0.804 ± 0.012	0.851 ± 0.028	0.524 ± 0.045
5	$\mathbf{0.957 \pm 0.002}$	0.656 ± 0.065	0.849 ± 0.044	0.839 ± 0.018	0.851 ± 0.055	0.546 ± 0.094
6	0.943 ± 0.005	0.690 ± 0.019	0.873 ± 0.012	0.849 ± 0.011	0.877 ± 0.015	0.583 ± 0.032
7	$\mathbf{0.957 \pm 0.003}$	0.697 ± 0.042	0.876 ± 0.025	0.842 ± 0.009	0.883 ± 0.003	0.597 ± 0.061
8	0.945 ± 0.003	0.665 ± 0.025	0.862 ± 0.019	0.818 ± 0.020	0.870 ± 0.026	0.563 ± 0.044
9	0.949 ± 0.003	0.639 ± 0.025	0.841 ± 0.018	0.837 ± 0.008	0.842 ± 0.023	0.518 ± 0.036

According to the tables 4.1 and 4.2, the artificial jitter has an effect on how the methods perform and on which methods perform better.

5 CONCLUSIONS

Many different methods for personalized ECG monitoring by anomaly detection methods were proposed. Novel proposed methods covered the impact of the sparse representation based feature extraction with OCSVM algorithm. Also, all the considered methods were chosen not to be computationally intensive since the personalized monitoring methods are very important specifically for wearable devices.

Eventually, all the methods were evaluated in the experiments using MIT-BIH data. The proposed methods were partially able to surpass the performance of the method without the suggested more extensive ways of feature extraction. Although, none of the methods was the best across all the covered metrics in the experiments. Neither, the best method for a certain person may not always be the best for everyone. The artificial jitter also had a significant effect on the performance. So, the robustness of the methods against the heartbeat segmentation error may also be wise to take into consideration.

In the future work, different sparse coding and dictionary learning methods could as well be discovered. This study did not either utilize any data actually from the wearable devices. The MIT-BIH data is from hospital ECG so the figures in this study are only directive when these methods are considered to be used with the wearable devices.

REFERENCES

[1] Ye, C., Kumar, B. V. and Coimbra, M. T. Heartbeat classification using morphological and dynamic features of ECG signals. *IEEE Transactions on Biomedical Engineering* 59.10 (2012), 2930–2941.

[2] De Chazal, P., O'Dwyer, M. and Reilly, R. B. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE transactions on biomedical engineering* 51.7 (2004), 1196–1206.

[3] Kiranyaz, S., Ince, T. and Gabbouj, M. Real-time patient-specific ECG classification by 1-D convolutional neural networks. *IEEE Transactions on Biomedical Engineering* 63.3 (2015), 664–675.

[4] Ceylan, R. The effect of feature extraction based on dictionary learning on ECG signal classification. *International Journal of Intelligent Systems and Applications in Engineering* 6.1 (2018), 40–46.

[5] Chauhan, S. and Vig, L. Anomaly detection in ECG time signals via deep long short-term memory networks. *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE. 2015, 1–7.

[6] Chuah, M. C. and Fu, F. ECG anomaly detection via time series analysis. *International Symposium on Parallel and Distributed Processing and Applications*. Springer. 2007, 123–135.

[7] Carrera, D., Rossi, B., Zambon, D., Fragneto, P. and Boracchi, G. Ecg monitoring in wearable devices by sparse models. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer. 2016, 145–160.

[8] Adler, A., Elad, M., Hel-Or, Y. and Rivlin, E. Sparse coding with anomaly detection. *Journal of Signal Processing Systems* 79.2 (2015), 179–188.

[9] Boracchi, G., Carrera, D. and Wohlberg, B. Novelty detection in images by sparse representations. *2014 IEEE Symposium on Intelligent Embedded Systems (IES)*. IEEE. 2014, 47–54.

[10] Carrera, D., Rossi, B., Fragneto, P. and Boracchi, G. Online anomaly detection for long-term ECG monitoring using wearable devices. *Pattern Recognition* 88 (2019), 482–492.

[11] Aharon, M., Elad, M. and Bruckstein, A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing* 54.11 (2006), 4311–4322.

[12] Schölkopf, B., Williamson, R. C., Smola, A. J., Shawe-Taylor, J. and Platt, J. C. Support vector method for novelty detection. *Advances in neural information processing systems*. 2000, 582–588.

[13] Abdiansah, A. and Wardoyo, R. Time complexity analysis of support vector machines (SVM) in LibSVM. *International journal computer and application* 128.3 (2015), 28–34.

[14] Moody, G. B. and Mark, R. G. The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine* 20.3 (2001), 45–50.

[15] Luz, E. J. d. S., Schwartz, W. R., Cámara-Chávez, G. and Menotti, D. ECG-based heartbeat classification for arrhythmia detection: A survey. *Computer methods and programs in biomedicine* 127 (2016), 144–164.

[16] AAMI, A. and EC57, A. *(R) 2008-Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms*. 2008.

[17] Rubinstein, R., Zibulevsky, M. and Elad, M. Efficient implementation of the K-SVD algorithm and the Batch-OMP method. *Department of Computer Science, Technion, Israel, Tech. Rep* (2008).

[18] Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J. and Scholkopf, B. Support vector machines. *IEEE Intelligent Systems and their applications* 13.4 (1998), 18–28.