PATIENT-ADAPTED AND INTER-PATIENT ECG CLASSIFICATION USING NEURAL NETWORK AND GRADIENT BOOSTING

L. Zaorálek, J. Platos, V. Snašel

Abstract: Heart disease diagnosis is an important non-invasive technique. Therefore, there exists an effort to increase the accuracy of arrhythmia classification based on ECG signals. In this work, we present a novel approach of heart arrhythmia detection. The model consists of two parts. The first part extracts important features from raw ECG signal using Auto-Encoder Neural Network. Extracted features obtained by Auto-Encoder represent an input for the second part of the model, the Gradient Boosting and Feedforward Neural Network classifiers. For comparison purposes, we evaluated our approach by using MIT-BIH ECG database and also following recommendations of the Association for the Advancement of Medical Instrumentation (AAMI) for ECG class labeling. We divided our experiment into two scenarios. The first scenario represents the classification task for the patient-adapted paradigm and the second one was dedicated to the inter-patient paradigm. We compared the measured results to the state-of-the-art methods and it shows that our method outperforms the state-of-the-art methods in the Ventricular Ectopic (VEB) class for both paradigms and Supraventricular Ectopic (SVEB) class in the inter-patient paradigm.

Key words: ECG, AAMI, VEB, SVEB, GBM, ANN

Received: April 24, 2018 DOI: 10.14311/NNW.2018.28.015
Revised and accepted: June 25, 2018

1. Introduction

The electrocardiographic (ECG) recordings offer a non-invasive method to analyse heart activity [5, 36]. The analysis can reveal cardiovascular diseases based on heart abnormalities. The ECG analysis brings the necessary manual review of each heartbeat and classify it. For that reason the analysis process is a very tedious and time-consuming task for human beings. Also there is a chance of human errors because the heart beat activity obtained by an ECG device, such as the ECG holter, has hours or even days of the ECG record. Over the last two decades, several approaches have been developed in order to automatically classify each beat and
detect heart abnormalities in the ECG recordings. The advantage of the automated approach is reducing human errors, low cost of the analysis and significantly decreased time to analyse. The heart beat can be classified as normal or abnormal depending on the heart activity. In the case of the abnormal heart beat activity, the heart has irregular rhythm. These rhythm disturbances are known as arrhythmias. There are several different irregular heart rhythms or arrhythmias that exist, for example extra beats, supraventricular tachycardias, ventricular arrhythmias, and bradyarrhythmias. In this paper, we present a novel approach in the classification task followed by [1]. There are three paradigms that exist on how to evaluate the automated system for classification of the ECG recordings:

- intra-patient
- inter-patient
- patient-adapted/patient-specific

The first one is the intra-patient paradigm. The intra-patient paradigm is focused on the classification of the ECG label. It means that the ECG recordings are randomly divided into training and test subsets. Therefore in this case, the ECG record belonging to one patient can occur in both of subsets. The evaluation of such an automated system has optimistic results, see [29]. The second paradigm is called inter-patient and brings out a more realistic scenario on how to evaluate the automated system [17]. Specifically, the ECG recording of one patient can be in training or testing subset but not in both subsets simultaneously. The last one is patient-specific or patient-adapted paradigm [8, 16]. In this case, the first 2-5 minutes of each ECG test recording are evaluated manually by cardiologists.

In recent years, several classification methods have been used in the task of arrhythmia detection, for example [9, 19, 27] or [34]. The classification methods are composed of four basic steps [2, 13]: ECG signal preprocessing, heartbeat segmentation, feature extraction and classification. Also in this paper we follow these steps, which are described in the Section 3. The preprocessing and heartbeat segmentation were described in [25, 30, 35]. The feature extraction step was widely explored by multiple sources. Almost all works in the context of the feature extraction used R-R interval (distance between two R-peaks of heartbeats), for example [11]. Another commonly used technique for the features extraction is QRS interval [22], wavelet transformation [15, 24] or reduction of dimensionality [6, 21, 40]. There are four popular algorithms that exist which are used for the classification step: Support Vector Machines, Linear Discriminant Analysis, Artificial Neural Network and Decision Tree, such as [3, 23, 28, 31, 37, 39].

2. Dataset

In this work we used the MIT-BIH arrhythmia database [32]. This database is publicly available on Physionet. There are 48 two-lead recordings obtained from 25 men aged 32 to 89 years, and 22 women aged 23 to 89 years. There is a total of 109 000 annotated beats in the database. There are 15 possible annotations in MIT-BIH database. The annotation of each beat was corrected by two cardiologists.
who worked independently on their corrections. For the purpose of comparison of our proposed methods with other works we used recommendations from the Association for the Advancement of Medical Instrumentation (AAMI). The AAMI designs the standard approach on how to evaluate methods and algorithms for ECG classification task. Specifically, we selected four to five annotations based on AAMI recommendations as shown in the Tab. I. In addition, the work [7] follows the AAMI recommendations and applies them to the MIT-BIH database. They divided MIT-BIH database into two subsets: DS1 and DS2. Specifically, the DS1 subset has following records: 101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223, 230. The DS2 includes 100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, 234 records. The following Tab. II summarizes the AAMI scheme for the MIT-BIH database. The recordings from DS1 are used for the training phase of the classification and the DS2 subset is dedicated to test and evaluate the performance of the proposed approach. Also other publications such as [8,12,13,16,18,19,27,28] and our own methods used the MIT-BIH database, AAMI scheme and de Chazal et al.’s concept in order to compare performance of the ECG classifiers.

<table>
<thead>
<tr>
<th>Group name</th>
<th>Shortcut</th>
<th>Symbol</th>
<th>Number of beats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>N</td>
<td>N, e, j, L, R</td>
<td>90125</td>
</tr>
<tr>
<td>Ventricular Ectopic</td>
<td>V</td>
<td>V, E</td>
<td>2781</td>
</tr>
<tr>
<td>Supraventricular Ectopic</td>
<td>S</td>
<td>A, a, J, S</td>
<td>7009</td>
</tr>
<tr>
<td>Fusion</td>
<td>F</td>
<td>F</td>
<td>803</td>
</tr>
<tr>
<td>Unknown</td>
<td>Q</td>
<td>U</td>
<td>15</td>
</tr>
</tbody>
</table>

Tab. I AAMI annotation groups for MIT-BIH database.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>N</th>
<th>S</th>
<th>V</th>
<th>F</th>
<th>Q</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS1</td>
<td>45866</td>
<td>944</td>
<td>3788</td>
<td>415</td>
<td>8</td>
<td>51021</td>
</tr>
<tr>
<td>DS2</td>
<td>44259</td>
<td>1837</td>
<td>3221</td>
<td>388</td>
<td>7</td>
<td>49712</td>
</tr>
<tr>
<td>Total</td>
<td>90125</td>
<td>2781</td>
<td>7009</td>
<td>803</td>
<td>15</td>
<td>100733</td>
</tr>
</tbody>
</table>

Tab. II The AAMI scheme for MIT-BIH database.

3. Methodology

Fig. 1 shows the scheme of our proposed approach. First, we removed noise from the ECG signal by using wavelet decomposition. After reconstructing the ECG signal without noise we normalized the signal. Next, we segmented each heartbeat and associated its annotation. In order to select the most important features from each ECG heartbeat, we performed a feature selection by using a type of Neural Network, namely the Auto-Encoder model. Finally, we trained proposed methods and measured the performance compared to other works. The following statements describe preprocessing, feature selection and classifier models in more detail.
3.1 ECG Preprocessing

In the preprocessing phase of the experiments, we removed a noise from each ECG signal. There are exists numerous approach that exist on how to denoise the ECG signal, such as an impulse response (FIR), a wavelet transformation or (Extended) a Kalman Filter. We chose discrete wavelet transformation as a denoising method based on promising results from [26, 33]. First of all, we calculated the discrete wavelet transformation (DWT) just as described by the Eq. (1),

$$\text{DWT}(m, k) = \frac{1}{\sqrt{a_0^{m}}} \sum_n x(n)\Psi\left(\frac{k - nb_0a_0^m}{s}\right),$$  \hspace{1cm} (1)

where $\Psi(\cdot)$ is the mother wavelet, $a = a_0^m$ and $b = nb_0a_0^m$ are the scaling and translation parameters, respectively [20, 33]. As the mother wave for DWT we selected “db4”. After the DWT computation we performed thresholding in the wavelet domain. Next, we shrank the wavelet coefficients by thresholding. Finally, we reconstructed the ECG signal from thresholded wavelet coefficients using inverse DWT transformation. See Fig. 2 for details on how the ECG signal looks before and after removing high frequency noise. After the signal denoising procedure it is necessary to rescale the ECG recordings to normal range from $-1$ to $1$ because we used neural network (NN) for the next step. The neural network cannot work properly if the range of the ECG signal is different than the normal range from $-1$ to $1$.

3.2 Segmentation

For the classification purpose we divided each preprocessed ECG waveform into segments and created a segmented dataset, such as $(X_1, Y_1), \ldots, (X_p, Y_p)$, where $p$ is the number of all segments created from preprocessed recordings, and $X_i$ represents the vector of samples belonging to one segment and $Y_i$ is associated annotation located nearby each R-peak in the original MIT-BIH database. For
example, [14] uses 90 samples on either side of the R-peak. However, based on our experiences and results it is sufficient to take 70 samples before and 20 samples after the R-peak location in the case of the MIT-BIH database. Note that, the last segment of a record is avoided from the dataset if it does not have 20 samples after the R-peak.

### 3.3 Feature selection

There are many approaches that exist on how to select important features from the ECG signals for the classification task. The feature selection can be divided into three general groups: temporal, morphological and statistical. In our methods, we employed the AE model for extracting important features from the ECG waveform as morphology features. Also we included temporal features for the following classification step: R-R intervals from last 30 heartbeats (including the current one).
Therefore the first 30 heartbeats of the dataset are avoided. The training algorithm of the AE model is the same as described in the next section, Neural-network. The algorithm remains the same, but the output of the AE model is a reflection of the input vector. Therefore, the AE model is a self-supervised regression technique in which the desired output vector is the input vector. The efficiency of the AE model in the task of the feature selection has been proved in several works, such as [2, 4, 10, 38]. We utilized grid search method to find optimal hyperparameters of the AE model. Concretely, we built the AE model with 3 dense layers. The first and last dense layer has 90 neurons and the second dense layer (“latent layer”) has 32 neurons. The Tab. III shows details of the AE model. Note that, \( I(n) \) means an input layer, \( D(n) \) is a dense layer, \( BN \) is a batch normalization layer, \( O(n) \) is an output layer and \( n \) is the number of output neurons from each layer. The AE transforms each segment \( X_i \) into 32 dimensional latent spaces.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value/Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topology</td>
<td>([I(90), D(90), BN, D(32), BN, D(90), O(90)])</td>
</tr>
<tr>
<td>Loss function</td>
<td>RMSE</td>
</tr>
<tr>
<td>Optimizer</td>
<td>RMSPROP</td>
</tr>
<tr>
<td>Learning rate</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Tab. III The AE setup and hyperparameters.

3.4 Classification

After feature selection the dataset was prepared to train the classifiers. The training phase includes two independent scenarios. In the first scenario, we performed classification tasks for patient-adapted paradigm. This means that the first 5 minutes of each DS2 record were added to DS1 subset. For the second scenario, based on the performance evaluation obtained from first scenario we chose better classifiers and measured performance within inter-patient paradigm. There were two classifiers in our experiments. The first was the Gradient Boosting Model (GBM) and the second one was the Feedforward Neural Network (FNN) model. See Tabs. IV and V for concrete setup and hyperparameters of the GBM and FNN models. The following statements describe how FNN and GBM model work in more detail.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value/Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topology</td>
<td>([I(62), D(32), BN, D(16), BN, O(5)])</td>
</tr>
<tr>
<td>Loss function</td>
<td>Cross-Entropy</td>
</tr>
<tr>
<td>Optimizer</td>
<td>RMSPROP</td>
</tr>
<tr>
<td>Learning Rate</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Tab. IV The FNN setup and hyperparameters.
Parameter | Value/Type
---|---
Depth | 10
Eta | 0.01
Estimators | 200
Boost Round | 200
Min child weight | 1
Subsample ratio | 1

Tab. V The Gradient Boosting setup and hyperparameters.

3.4.1 Neural network

The Artificial Neural Network (ANN) is a model inspired by biological neural network. The ANN model is composed in layers. Each layer has a number of neurons. The topology of the ANN model defines how neurons in the layer are connected with neurons from the next layer. Note that it is possible to create connections with neurons from previous layer. This ANN model is called the recurrent NN model. However, in our experiment, we focused only on the forward ANN model without recurrent connections between neurons. The one connection between $k$-th neuron from $(l - 1)$ layer and $j$-th neuron from $l$-th layer is defined as weight $w_{jk}$. Each neuron has internal potential $z_j^l$ defined as weighted input for the activation function $\sigma$. The internal potential and the activation function for the $j$-th neuron in the $l$-th layer is shown as (4) and (3) where $b_j^l$ is a neuron bias.

Let’s suppose ECG dataset $(X_1, Y_1), \ldots, (X_p, Y_p)$, where $X_j \in \mathbb{R}^n$ is a vector for multilead ECG signal and $Y_j \in \mathbb{R}^m$ is a vector for associated annotation label. We would like to train the ANN model to predict label $Y_j$ based on the input vector $X_j$. In other words, the goal is to minimize a loss function, such as a root mean square error (RMSE), between the actual output of the ANN model $a^L(X_j)$ based on the sample $X_j$, and the desired output vector $Y_j$ as shown in Eq. (2), where $L$ is the number of layers

$$C = \frac{1}{2n} \sum_{j=1}^{p} (Y_j - a^L(X_j))^2,$$

(2)

$$a_j^l = \sigma(z_j^l),$$

(3)

$$z_j^l = \sum_k w_{jk}^l a_{jk}^{l-1} + b_j^l.$$  

(4)

The way to improve neuron weights in order to minimize loss function is through the backpropagation algorithm. The backpropagation algorithm calculates the value of the loss function $C$ and is based on the gradient composed by Eqs. (7), (8), (9), (10) updates each weight and neuron bias of the ANN model as shown (5), (6)

\footnote{Annotation should be represented by one-hot encoding.}
\[ w_{jk}^l = w_{jk}^l - \eta \frac{\partial C}{\partial w_{jk}^l}, \]  
(5)

\[ b_j^l = b_j^l - \eta \frac{\partial C}{\partial b_j^l}, \]  
(6)

\[ \frac{\partial C}{\partial w_{jk}^l} = a_k^{l-1}\delta_j^l, \]  
(7)

\[ \frac{\partial C}{\partial b_j^l} = \delta_j^l, \]  
(8)

\[ \delta_L^j = \frac{\partial C}{\partial a_j^L}\sigma'(z_j^l), \]  
(9)

\[ \delta_j^l = \sum_k w_{jk}^{l+1}\delta_k^{l+1}\sigma'(z_j^l). \]  
(10)

### 3.4.2 Gradient Boosting Model

The GBM is a type of machine learning technique, where the model is composed from weak predictors (models). Typically, a weak predictor is a Decision Tree (like CART tree) but it could be Ordinary Linear Regression (OLR), Ridge Penalized Linear Regression (RPLR) or other predictors. The idea behind the GBM is to minimize loss function by iteratively adding additional weak predictors into the model based on expected values of the training dataset. Note that a previously added weak predictor does not change in the next learning iteration. Let’s suppose that \( N \) datapoints pair \( X_j, Y_j \) of a dataset, where \( X_j = \{x_1, \ldots, x_d\} \) is a multivariate input variable of the dataset and \( Y_j \) is an univariate output variable of the dataset. The following Eqs. (11) describe the learning phase of GBM classifier. The function \( \bar{f} : \mathbb{R}^d \to \mathbb{R} \) is a minimization of expectations of the loss function as it shows in Eq. (11):

\[ \bar{f}(X) = \arg\min_{f} E_{y,x} [L(Y, f(X))], \]  
(11)

where \( \bar{f} \) is the additive model with minimized loss function for the multivariate input variable \( X \) and univariate output variable \( Y \), \( L \) is the loss function and \( f \) is the additive model. Following Eq. (12) shows how GBM model represents by the function \( \bar{f}(X) \) is composed from weaker models \( f_i \):

\[ \bar{f}(X) = \sum_{i=1}^{M} \beta_i f_i(X) = \sum_{i=1}^{M} \beta_i h_i(X; \theta_i), \]  
(12)

where \( M \) is the number of algorithm iterations and \( f_i \) is the weak model and \( \beta_i \) is the expansion coefficient. The function \( h_i \) represents parameterized decision tree of weak model \( f_i \), where \( \theta_i = \{A_i, \gamma_i\} \) is parameter set of the tree. The parameter set \( \theta_i \) has the region set \( A_i \) and the coefficients \( \gamma_i \). The \( \gamma_i \) represents all outputs for all examples \( X \) associated to the region set \( A_i \). The Eq. (11) does not have closed-form solution. Instead of finding a solution in function space the training
algorithm can use the steepest gradient descent approach because the loss function $L$ is differentiable. Eq. (13) shows the first step of the gradient descent update:

$$g_{ij}(x_j, y_j, f) = \frac{\partial L[y_j, f(x_j)]}{\partial f(x_j)},$$  \hspace{1cm} (13)

where $x_j$ is an $j$-th mutlivariate input variable, $y_j$ is an univariate output variable associated with the $x_j$ variable, $L$ is loss function and $f$ is the function of additive model at step $i-1$. The second step of the gradient descent update is the gradient originating from Eq. (13) transforms to construct $i$-th decision tree. The step seems to very hard task because it means find the solution for the boost increment in function space. The solution of this particular problem is a “proxy”, where instead of working in function space, one can construct new additive model that is most correlated with $-g_i(x)$ with step-size $\rho$:

$$(\theta_i, \rho_i) = \arg\min_{\theta, \rho} \sum_{j=1}^{N} (\rho h(X_j; \theta) - g_{ij}(x_j, y_j, \bar{f}_{i-1}))^2,$$  \hspace{1cm} (14)

where $N$ is number of samples in the dataset, $\theta_i$ is found new tree parameters at $i$-th iteration, $h(X_j, \theta)$ is parameterized decision tree. Also the expansion coefficients can be estimated by Eq. (15):

$$\beta_j = \arg\min_{\beta} \sum_{j=1}^{N} L(y_j, f_{i-1}(x_j) + \beta h(X_i; \theta_i)).$$  \hspace{1cm} (15)

Finally, Eq. (16) defines construction of the new model with $(\beta_i, \theta_i)$ by using least squares fit method obtained from Eqs. (14) and (15) as follows:

$$\bar{f}_i = \bar{f}_{i-1} + \beta_i h(X, \theta_i).$$  \hspace{1cm} (16)

4. Results

The following statements present the results for patient-adapted and inter-patient paradigms followed by the AAMI standard. First, we introduce the results for patient-adapted paradigm, where we used all five classes described in Section 2. Next we show the classification performance of three classes for inter-patient paradigm. In both cases, we compared our proposed methods with other state-of-the-art algorithms. Note that the classifiers were always trained on DS1 subset and performance was measured on DS2 subset as it was defined in Section 2 and mentioned in Section 3. The performance evaluation was calculated based on the sensitivity ($Se$), positive predictivity ($Pp$), specificity ($Sp$), and overall accuracy ($Oa$) for the class VEB (V) and the class SVEB (S) in the same way as it was proposed by [7].

4.1 Patient-adapted paradigm

Tab. VI shows summary beat-by-beat results of the GBM classifier for DS2 subset. There are 5 classes (N, S, V, F, and Q) recommended by AAMI standard. The
algorithms achieved higher true positive ratio for the class V in comparison to the class S. On the other hand, the proposed algorithm has problems recognizing class Q. The reason why Q was not recognized by the algorithm is that the class Q has very low occurrence (see Tab. I). The total accuracy of the proposed method was 97.9%.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>N</th>
<th>S</th>
<th>V</th>
<th>F</th>
<th>Q</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>43886</td>
<td>140</td>
<td>39</td>
<td>123</td>
<td>0</td>
<td>44188</td>
</tr>
<tr>
<td>S</td>
<td>525</td>
<td>1297</td>
<td>13</td>
<td>2</td>
<td>0</td>
<td>1837</td>
</tr>
<tr>
<td>V</td>
<td>172</td>
<td>10</td>
<td>3013</td>
<td>25</td>
<td>0</td>
<td>3220</td>
</tr>
<tr>
<td>F</td>
<td>70</td>
<td>1</td>
<td>32</td>
<td>285</td>
<td>0</td>
<td>388</td>
</tr>
<tr>
<td>Q</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>44657</td>
<td>1448</td>
<td>3100</td>
<td>435</td>
<td>0</td>
<td>49640</td>
</tr>
</tbody>
</table>

**Tab. VI** The confusion matrix of the GBM classifier for DS2 dataset.

In case of the FNN classifier, the algorithm has worse accuracy than the GBM classifier. Concretely, the FNN classifier correctly recognized only 249 S beats of the total 1837 S beats and 1898 V beats of total 3220 V beats. The classifier was unable to recognize any Q beat due to a problem with the number of Q beats.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>N</th>
<th>S</th>
<th>V</th>
<th>F</th>
<th>Q</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>44158</td>
<td>1</td>
<td>28</td>
<td>1</td>
<td>0</td>
<td>44188</td>
</tr>
<tr>
<td>S</td>
<td>1433</td>
<td>249</td>
<td>155</td>
<td>0</td>
<td>0</td>
<td>1837</td>
</tr>
<tr>
<td>V</td>
<td>1320</td>
<td>2</td>
<td>1898</td>
<td>0</td>
<td>0</td>
<td>3220</td>
</tr>
<tr>
<td>F</td>
<td>308</td>
<td>0</td>
<td>48</td>
<td>32</td>
<td>0</td>
<td>388</td>
</tr>
<tr>
<td>Q</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>47224</td>
<td>252</td>
<td>2131</td>
<td>33</td>
<td>0</td>
<td>49640</td>
</tr>
</tbody>
</table>

**Tab. VII** The confusion matrix of the FNN classifier for DS2 dataset.

Also we compare our proposed algorithms with three existing methods [8,16,18, 19]. Note that these methods also comply with AAMI standards. See Tab. VIII, where Se is sensitivity, Pp is positive predictivity, Sp is specificity and Oa is overall accuracy. Our methods clearly outperformed other mentioned methods. As we can see, our method has a higher positive predictive ratio and specificity for both classes S and V. Also, the sensitivity of the class V has the highest value, 94.0%. Only in the case of the S class, the proposed method with GBM classifier has a lower sensitivity ratio. But, in the whole context of the measured statistics, such as sensitivity, positive predictivity and specificity, the GBM classifier for class S also achieved better performance than other methods. Mainly, the positive predictivity of the GBM classifier was 89.6%. The second classifier which we proposed, the FNN classifier, achieved lower performance than the GBM classifier in both classes S and V.
In this section, we present performance evaluation and comparison of the best proposed method from Patient-adapted paradigm. The chosen classifier based on the results from Tab. VIII was the GBM classifier. The GBM classifier we compared with other state-of-the-art methods within the class S and V. Note that, we omit measurement of the class Q and F in this measurement evaluation due to very low occurrence of the Q and F beat samples. Tab. IX shows confusion matrix of the GBM classifier on the DS2 subset. The GBM classifier recognized 43876 N beats and 2950 V beats but it had a lower predictivity rate of the S beats. Concretely, it predicted only 113 S beats of total 1836 S beats.

### Tab. VIII
**Comparison patient-specific results for DS2 (In percent).**

<table>
<thead>
<tr>
<th>Method</th>
<th>Se</th>
<th>Pp</th>
<th>Sp</th>
<th>Oa</th>
<th>Se</th>
<th>Pp</th>
<th>Sp</th>
<th>Oa</th>
</tr>
</thead>
<tbody>
<tr>
<td>de Chazal and Reilly [8]</td>
<td>87.7</td>
<td>47</td>
<td>N/A</td>
<td>95.9</td>
<td>94.3</td>
<td>96.2</td>
<td>N/A</td>
<td>99.4</td>
</tr>
<tr>
<td>Hu et al. [16]</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>78.9</td>
<td>75.8</td>
<td>96.8</td>
<td>94.8</td>
</tr>
<tr>
<td>Ince et al. [18]</td>
<td>81.8</td>
<td>63.4</td>
<td>98.5</td>
<td>96.1</td>
<td>90.3</td>
<td>92.2</td>
<td>98.8</td>
<td>97.9</td>
</tr>
<tr>
<td>Jiang and Kong [19]</td>
<td>74.9</td>
<td>78.8</td>
<td>98.8</td>
<td>97.5</td>
<td>94.3</td>
<td>95.8</td>
<td>99.4</td>
<td>98.8</td>
</tr>
<tr>
<td>Proposed (GBM)</td>
<td><strong>70</strong></td>
<td><strong>89.6</strong></td>
<td><strong>99.7</strong></td>
<td><strong>98.6</strong></td>
<td><strong>94.3</strong></td>
<td><strong>98.3</strong></td>
<td><strong>99.9</strong></td>
<td><strong>99.5</strong></td>
</tr>
<tr>
<td>Proposed (FNN)</td>
<td>13.6</td>
<td>98.8</td>
<td>99.9</td>
<td>96.8</td>
<td>58.9</td>
<td>91.2</td>
<td>99.6</td>
<td>96.9</td>
</tr>
</tbody>
</table>

4.2 **Inter-patient paradigm**

In this section, we present performance evaluation and comparison of the best proposed method from Patient-adapted paradigm. The chosen classifier based on the results from Tab. VIII was the GBM classifier. The GBM classifier we compared with other state-of-the-art methods within the class S and V. Note that, we omit measurement of the class Q and F in this measurement evaluation due to very low occurrence of the Q and F beat samples. Tab. IX shows confusion matrix of the GBM classifier on the DS2 subset. The GBM classifier recognized 43876 N beats and 2950 V beats but it had a lower predictivity rate of the S beats. Concretely, it predicted only 113 S beats of total 1836 S beats.

### Tab. IX
**The confusion matrix of the GBM classifier for inter-patient paradigm and DS2 subset.**

<table>
<thead>
<tr>
<th>Classifier</th>
<th>N</th>
<th>S</th>
<th>V</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truth</td>
<td>43876</td>
<td>209</td>
<td>107</td>
<td>44192</td>
</tr>
<tr>
<td>S</td>
<td>1616</td>
<td>113</td>
<td>107</td>
<td>1836</td>
</tr>
<tr>
<td>V</td>
<td>255</td>
<td>15</td>
<td>2950</td>
<td>3220</td>
</tr>
<tr>
<td>Total</td>
<td>45747</td>
<td>337</td>
<td>3164</td>
<td>49248</td>
</tr>
</tbody>
</table>

See also Tab. X where there is a comparison of the GBM classifier with other state-of-the-art methods, where Se is sensitivity, Pp is positive predictivity, Sp is specificity and Oa is overall accuracy. The GBM model has significantly outperformed other methods in case of the class V. Concretely, the GBM classifier achieved 91.6% for sensitivity, 93.2% for positive predictivity, 99.5% for speci-
ficity and 99.0 % accuracy for the class V. But in case of the class S, the GBM classifier has only 6 % sensitivity because it predicted only 113 S beats of total 1836 S beats.

<table>
<thead>
<tr>
<th>Method</th>
<th>SVEB Se</th>
<th>SVEB Pp</th>
<th>SVEB Sp</th>
<th>Oa</th>
<th>VEB Se</th>
<th>VEB Pp</th>
<th>VEB Sp</th>
<th>Oa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lin &amp; Yang [27]</td>
<td>81.0</td>
<td>31.0</td>
<td>N/A</td>
<td>N/A</td>
<td>86.0</td>
<td>73.0</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Llamedo &amp; Martínez [28]</td>
<td>77.0</td>
<td>39.0</td>
<td>95.3</td>
<td>94.7</td>
<td>81.0</td>
<td>87.0</td>
<td>99.0</td>
<td>97.7</td>
</tr>
<tr>
<td>Garcia et. al, 2016 [12]</td>
<td>30</td>
<td>26</td>
<td>98.3</td>
<td>95.0</td>
<td>85</td>
<td>66</td>
<td>96.2</td>
<td>95.4</td>
</tr>
<tr>
<td>Garcia et. al, 2017 [13]</td>
<td>62</td>
<td>53</td>
<td>97.9</td>
<td>96.6</td>
<td>87.3</td>
<td>59.4</td>
<td>95.9</td>
<td>95.4</td>
</tr>
<tr>
<td>Proposed (GBM)</td>
<td>6.0</td>
<td>33.5</td>
<td>99.5</td>
<td>96.0</td>
<td>91.6</td>
<td>93.2</td>
<td>99.5</td>
<td>99.0</td>
</tr>
</tbody>
</table>

Tab. X Compare inter-patient results for DS2 (In percent).

5. Conclusion

In this paper, we proposed a novel approach to the task of arrhythmia classification for patient-specific and inter-patient paradigms. We measured the performance of our approach on the MIT-BIH database and also compare the results with other state-of-the-art methods. Concretely, the denoised and normalized ECG signals we transformed to reduced feature space, and we show the performance of trained NN and GBM models in the reduced space. For the purpose of comparison, results obtained by our approach with other works, we used five classes defined by the AAMI standard for patient-adapted paradigm and three classes for inter-patient paradigm. We compared our approach to state-of-the-art methods based on statistical and machine learning methods. The results show that our proposed method outperforms state-of-the-art methods for VEB heartbeats in both paradigms, and outperforms the prediction of the SVEB heartbeat in case of the patient-adapted paradigm. The focus of the future work will be increasing the accuracy of Supraventricular Ectopic heartbeats, mainly in the context of the inter-patient paradigm.

References


