INTER- AND INTRA- PATIENT ECG HEARTBEAT CLASSIFICATION FOR ARRHYTHMIA DETECTION: A SEQUENCE TO SEQUENCE DEEP LEARNING APPROACH

Sajad Mousavi, Fatemeh Afghah

School of Informatics, Computing and Cyber Systems, Northern Arizona University, USA

ABSTRACT
Electrocardiogram (ECG) signal is a common and powerful tool to study heart function and diagnose several abnormal arrhythmias. While there have been remarkable improvements in cardiac arrhythmia classification methods, they still cannot offer acceptable performance in detecting different heart conditions, especially when dealing with imbalanced datasets. In this paper, we propose a solution to address this limitation of current classification approaches by developing an automatic heartbeat classification method using deep convolutional neural networks and sequence to sequence models. We evaluated the proposed method on the MIT-BIH arrhythmia database, considering the intra-patient and inter-patient paradigms, and the AAMI EC57 standard. The evaluation results for both paradigms show that our method achieves the best performance in the literature (a positive predictive value of 96.46% and sensitivity of 100% for the category S, and a positive predictive value of 98.68% and sensitivity of 97.40% for the category F for the intra-patient scheme; a positive predictive value of 92.57% and sensitivity of 88.94% for the category S, and a positive predictive value of 99.50% and sensitivity of 99.94% for the category V for the inter-patient scheme). The source code is available at https://github.com/SajadMo/ECG-Heartbeat-Classification-seq2seq-model.

Index Terms— ECG analysis, heartbeat classification, deep learning, sequence to sequence model, RNNs.

1. INTRODUCTION
An electrocardiogram (ECG) is a common non-invasive tool to record heart activities and detect different abnormalities in heart functionality. Classification of the arrhythmic heartbeats in the ECG signal can be a challenging and time-consuming task for a physician, therefore, such heartbeat hand-annotating is often prone to error. This problem calls for automatic heartbeat classification methods that are able to diagnose arrhythmic heartbeats in real-time with high accuracy. Several machine learning algorithms such as support vector machines (SVM), multilayer perceptron (MLP), reservoir computing with logistic regression (RC) and decision trees have been utilized for arrhythmia detection [1, 2, 3, 4, 5, 6]. These shallow machine learning methods for ECG processing usually follow three main steps, including 1) signal pre-processing, which includes noise removal methods, heartbeat segmentation, etc.; 2) feature extraction; and 3) learning/classification. Even though these methods with hand-engineered features and applying noise removal techniques have achieved acceptable performances, deep learning approaches (i.e., automated feature extractions) have shown impressive results in various domains ranging from computer vision and reinforcement learning to natural language processing [7, 8, 9] as well as more applicable outcomes in biomedical signal processing [10, 11, 12].

One of the main limitations of the current heartbeat classification methods including shallow and deep machine learning methods is their poor performance when dealing with imbalanced datasets. In particular, they attain a low positive predictive value and sensitivity for the classes with lower sample size in the dataset. For instance, the majority of existing ECG analysis techniques achieve a low sensitivity in the MIT-BIH arrhythmia database for ventricular escape beat (S) and fusion of ventricular, and normal beat (F) classes. Furthermore, most previously reported works in the literature have been evaluated based on intra-patient paradigm rather than the inter-patient scheme, which is an obviously more realistic scenario to prevent training and test the model using the samples from the same patients. Therefore, although some of these methods achieved good accuracies using the intra-patient scheme, their results are not reliable as their evaluation process was biased [13].

As mentioned above, the conventional arrhythmia classification systems can be generally divided into two categories of inter-patient and intra-patient paradigms in terms of their evaluation mechanism. In intra-patient paradigm, the training and evaluation datasets can include heartbeats from the same patients, while in inter-patient paradigm, a more realistic evaluation mechanism is used where the heartbeat sets for test and training come from different individuals. One of our aims in this paper is to evaluate the proposed method with both the paradigms.

Inspired by the aforementioned issues with the previous works, this paper proposes a novel and effective approach for automatic ECG-based heartbeat classification by leveraging a sequence to sequence deep learning method and an oversampling method named Synthetic Minority Over-sampling Technique (SMOTE) to address the aforementioned challenge with minority classes. The proposed model is evaluated using inter-patient and intra-patient paradigms where it achieves the best results compared to the existing works in the literature.

The rest of this paper is organized as follows. Section 2 introduces the database utilized in this study. Section 3 describes the proposed method. Section 4 presents the experimental setup and shows the achieved results by the proposed method along with a performance comparison to the state-of-the-art algorithms. Finally, Section 5 concludes the paper.

2. DATASET
In this study, we used the PhysioNet MIT-BIH Arrhythmia database to evaluate the performance of our proposed method [14, 15]. The MIT-BIH dataset includes the ECG signals for 48 different subjects recorded at the sampling rate of 360Hz.
Each record contains two ECG leads; ECG lead II and lead V1. Usually, the lead II signal is used to detect heartbeats in the literature. Similarly, here in all experiments, we have applied ECG lead II. This database is recommended by the American association of medical instrumentation (AAMI) [16], since it includes the five essential arrhythmia groups as described in Table 1.

We considered two main paradigms of inter-patient and intra-patient to evaluate the proposed model. In the inter-patient paradigm, two sets of data samples (beats) are chosen randomly as training and test samples in which there may be the heartbeat samples of the same patient in the training and test sets. While, in the inter-patient paradigm, the training and test set are constructed from different patients, following the protocol proposed by de Chazal et al. [13]. In this method, the heartbeats from the MIT-BIH database (44 records based on AAMI) are divided into two sets of records: DS1 = {101, 101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223, 230} and DS2 = {100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, 234}. DS1 is used to build the classification model and DS2 is utilized to test the model. Using this division approach, there is no concern about including the heartbeats from the same patient in both training and test sets.

3. METHODOLOGY

In the following sections, we present a detailed description of our proposed novel model to automatically classify each heartbeat of a given ECG signal.

3.1. Pre-processing

The input of this method is a sequence of ECG beats. In order to extract the ECG beats from a given ECG signal, we follow a few simple steps:

1. Normalizing the given ECG signal to the range of between zero and one.
2. Finding the set of $t$ waves regarding the ECG waves (R-peaks) of its corresponding annotation file in the MIT-BIH Arrhythmia database.
3. Splitting the continuous ECG signal to a sequence of heartbeats based on the extracted $t$ waves and assigning a label to each heartbeat based on the annotation file.
4. Resizing each heartbeat to a predefined fixed length (280 samples).

We would like to note that these pre-processing steps for beat extraction are very simple and do not involve any form of filtering or noise removal methods.

3.2. The architecture

The sequence to sequence models have shown very impas-sive results in neural machine translation applications, nearly similar to human-level performance [17]. The architecture of sequence to sequence networks is usually composed of two main parts of the recurrent neural network (RNN) encoder and decoder. In this study, we leverage a RNN sequence to sequence model along with a convolutional neural network (CNN) to perform a heartbeat detection task.

Fig. 1 illustrates the proposed network architecture for automatic beat classification. The CNN consists of three consecutive one-dimensional convolutional layers. The first layer is composed of 32 1-D convolution filters with a kernel size of $2 \times 1$ and a stride 1, followed by a Rectified Linear Unit (ReLU) non-linearity. The second layer consists of 64 1-D convolution filters with a kernel size of $2 \times 1$ and a stride 1, again followed by an ReLU. Finally, the third layer is comprised of 128 1-D convolution filters with a kernel size of $2 \times 1$ and a stride 1, followed by a rectifier non-linearity. Each convolutional layer except the last layer is also followed by a max pooling layer of pooling region of size $2 \times 1$ with a stride 1. At each time-step of training/testing the model, a sequence (size of maxtime) of ECG beats is fed into the CNN in order for feature extraction. The last convolutional layer outputs the maxtime of $F$ feature maps of size $k \times 1$ (e.g., here, we reached 128 feature maps $3 \times 1$). In the end, each beat of the input sequence is associated with a vector $c \in \mathbb{R}^d$.

We have utilized the bidirectional recurrent neural network (BiRNN) units in the network architecture instead of the standard LSTM (i.e., RNN). Standard RNNs are unidirectional, hence they are restricted to the use of the previous input state. To address this limitation, the BiRNN have been proposed [19], which can process data in both forward and backward directions. Thus, the current state has access to previous and future input information simultaneously. The BiRNN consists of a forward network and a backward network. The input sequence is fed in normal time order, $t = 1, \ldots, T$ for the forward network, and in reverse time order, $t = T, \ldots, 1$ for the backward network. Finally, the weighted sum of the outputs of the two networks is computed as the output of the BiRNN. This mechanism can be formulated as follow:

\[ h_t = \tanh(W_x x_t + V h_{t-1} + b) \]
\[ h_t = \tanh(W_x x_t + V h_{t+1} + b) \]
\[ y_t = (U[h_t; h_t] + b_y) \]
where \((\overrightarrow{h_t}, \overrightarrow{b})\) are the hidden state and the bias of the forward network, and \((\overleftarrow{h_t}, \overleftarrow{b})\) are the hidden state and the bias of the backward network. Also, \(x_t\) and \(y_t\) are the input and the output of the BiRNN, respectively. The decoder is used to generate the target sequence beat by beat. Similar to the encoder, the building block of the decoder is an LSTM, but a many-to-many LSTM. The decoder gets a new representation of the input sequence generated by the encoder to initialize its hidden state. It also takes the same given target shifted by one and starts with a special feature vector \(< GO >\) as input. We should note that the input (the shifted target) is just used during the training phase and is not applied for the testing phase. Then, a softmax is applied to the output of the LSTM to convert it to a vector of probabilities \(p \in \mathbb{R}^C\), where \(C\) represents the number of categories (i.e., the heartbeat types) and each element of \(p\) indicates the probability of each class in the category.

4. EXPERIMENTAL EVALUATION

4.1. Experimental setup

The performance of the proposed heartbeat detection method has been evaluated using the MIT-BIH arrhythmia database using both inter-patient and intra-patient evaluation methods. Heartbeat category distribution of extracted beats from the MIT-BIH arrhythmia is not uniform and the numbers of normal beats are much greater than other categories. Machine learning approaches usually have trouble learning when one class dominates the others. To cope with this, the dataset has been over-sampled to nearly reach a balanced number of beats in each category (group). To this end, we used the synthetic minority over-sampling technique (SMOTE) which generates the synthetic data points by considering the similarities between existing minority samples [23].

Similar to several works in the literature, we evaluated the arrhythmia classifier on four cardiac cycles including N (normal and bundle branch block beats), S (supraventricular ectopic beats), V (ventricular ectopic beats), and F (fusion of N and V beats) (for more details refer to Table 1). For the intra-patient scheme, we extracted a total of 101,290 heartbeats, including 90,494 beats for the category N, 2,777 for the category S, 7,217 for the category V and 802 for the category F. To reduce the impact of class imbalance problem, the SMOTE technique was applied to upsample the categories with the small factions of heartbeats. We have trained the proposed model with 80% of the dataset and evaluated it with the remaining 20%. While for the inter-patient scheme, we had 45,796 beats of the category N, 941 beats of the category S, and 3,780 beats of the category V in DS1 (i.e., training set), and 44,196 beats for the category N, 1,836 beats for the category S, and 3,216 beats for the category V in DS2 (i.e., test set). Again, the SMOTE approach was utilized to compensate the number of heartbeats of the categories with the lower heartbeats.

The augmented data (i.e., synthetic data generated based on the original available data) should not be considered in the validation phase as the information of the validation set may exist in the training set. For instance, although the work reported in [11], followed this approach, we believe that the results are not quite accurate. It is worth mentioning that we generated new data samples by the SMOTE algorithm after splitting the heartbeats into the training and test datasets. Indeed, the oversampling was only performed on the training dataset for the intra-patient paradigm (and on DS1 for the inter-patient paradigm) and the information in the test dataset was not used to create synthetic data points. Therefore, the evaluation process is reliable and generalizable.

Four main measures usually were considered in the literature to evaluate the performance of heartbeat classification techniques including the sensitivity (\(SE\)), positive predictive value (PPV), specificity (\(SP\)) (1 - False Positive Rate (FPR)) and accuracy (\(Acc\)) as defined in following:

\[
SE = \frac{TP}{TP + FN}
\]

\[
PPV = \frac{TP}{TP + FP}
\]

\[
SP = \frac{TN}{TN + FP}
\]

\[
Acc = \frac{(TP + TN)}{(TN + FP + FN)}
\]

where TP (True Positive), TN (True Negative), FP (False Pos-
4.2. Results and Discussion

The results are presented for two evaluation scenarios of inter-patient paradigm, in which the training and test sets were randomly chosen from all available patients’ heartbeats, and the inter-patient paradigm, in which the training and test have been performed on the heartbeats of extracted from DS1 and DS2, respectively (i.e., no common individual in test and training sets). Table 2 presents a comparison of heartbeat classification results for the proposed method and the existing algorithms, considering intra-patient scheme. As it is clear, our sequence to sequence model outperforms all state-of-the-art algorithms significantly in terms of all evaluation metrics including overall accuracy, sensitivity, positive predictive value and specificity for the considered groups, N, S, V, F. In addition, it is worth noticing that the work done by Acharya et al. [11] used artificially augmented dataset to build the model, and their model evaluation was performed using augmented data, while we evaluated our model on real data samples without including any augmentations in the test set.

As confirmed by the results, our proposed method can provide a robust solution for class imbalance problem as one of the key challenges in dealing with medical data, which is due to the limited availability of abnormal classes compared to the normal classes in biomedical datasets. It is shown in Table 2 that our model achieves remarkable outcomes for the category F with only 802 heartbeats and the category S with 2,777 heartbeats.

We also validated our method using the more realistic evaluation method of inter-patient for heartbeat classifiers based on using DS1 set for training and DS2 set for testing [13, 24]. Table 3 presents the performance comparison between our proposed method and several state-of-the-art methods using MIT-BIH arrhythmia database where the inter-patient paradigm is considered. As it can be seen from the table, overall the proposed method has better performance for classifying all heartbeat categories compared to works listed in Table 3. In spite of the low number of S heartbeats in the training set, our proposed method obtained significant evaluation results.

The proposed method is generic in nature and it is expected to achieve a promising performance in several biomedical applications dealing with class imbalance problem. In the proposed classification structure, first, the CNN extracts a set of meaningful features of the given ECG heartbeats (Note, the model was trained by added synthetic data generated by SMOTE algorithm to the available samples to compensate the number of small categories such as F). Then, the encoder maps the features to new feature representations, capturing temporal patterns, and finally, the decoder takes the feature representations and produces the outputs (i.e., the labels for each heartbeat of the input sequence), considering complex context dependencies between the inputs and the targets.

5. CONCLUSION

In this study, we presented a novel and effective automatic heartbeat classification/annotation, considering inter- and inter-patient schemes and validated its performance using the MIT-BIH arrhythmia database. The proposed method leverages the ability of deep convolutional neural network and encoder-decoder network in which we have used a bidirectional recurrent neural network as its building blocks. According to the results, the suggested method significantly outperforms the existing algorithms in the literature for both inter-patient paradigm and inter-patient paradigm. Furthermore, the proposed method can be applied to several biomedical applications such as sleep staging where there are strong dependencies between each stage and sufficient data are not available. Also, the proposed network with a low number of parameters (i.e., with a maximum size of 5.5MB) can be used with wearable devices.
6. REFERENCES


