ECG Beat Arrhythmia Classification by using 1-D CNN in case of Class Imbalance

Çağla SARVAN and Nalan ÖZKURT
Department of Electrical and Electronics Engineering
Yaşar University
İzmir, Turkey
cagla.sarvan@yasar.edu.tr, nalan.ozkurt@yasar.edu.tr

***Abstract*— In this study, ECG arrhythmia types of non-ectopic (N), ventricular ectopic (V), unknown (Q), supraventricular ectopic (S) and fusion (F) were classified by using the convolutional neural network (CNN) architecture. QRS detection was performed on these ECG arrhythmias that downloaded from MIT-BIH database. An imbalanced number of beats was obtained for 5 different arrhythmia types. In order to reduce the effect of imbalance in statistical performance metrics, data mining techniques, such as recall of data, were applied. It was aimed to increase the positive predictive value (PPV) rates of the classes which consist of a few instances.**

Keywords —CNN, ECG, Arrhythmia, Imbalanced, Data Mining

I. INTRODUCTION

Nowadays, many methods are used for the diagnosis of diseases in the human body. Electrocardiography (ECG) enables the monitoring and recording of electrical activity during the operation of the heart. ECG signals are important in disease diagnosis. The individuals with cardiac arrhythmia can be identified by experts by examining the ECG records. Therefore, the analysis of heart signals and the development of automatic methods, that can help the experts, have great importance in the diagnosis of cardiac cycle disorders.

To understand the signal characteristics, it is important to obtain different features on the signal. Usually, feature extraction methods have been used on ECG signals that require human experimentation to find best features which differentiate ECG arrhythmia types. The suitability of the extracted features from signals has a significant impact on the reliability and performance of a particular classification algorithm. Different methods were proposed in the literature for the feature extraction from ECG signals such as a general and patient-specific classification system by using morphological wavelet transform features of ECG data and multidimensional particle swarm optimization technique are used [1-4]. The information about how well the features reflect the signal characteristics is measured by the performance metrics of the classification algorithms. The selection of suitable features involves an intense human trial and error factor.

Recently, deep neural networks have received great interest in biomedical signal processing applications. Convolutional neural network (CNN) can be applied directly to raw ECG signals without any pre-filtering, or feature extraction. It deals with the problem of classification by extracting transient features through the ECG signal. In [5], block-based neural networks, optimized by the proposed evolutionary algorithm is employed for ECG heartbeat pattern classification. 1-D CNN algorithm is applied as a classification technique in [6]. In [7], it is recommended to add the softmax regression layer over the hidden representation layer as deep neural network, labeling the most relevant and uncertain ECG beats in the test record at each iteration and using this data to update the DNN weights, while a 9-layer deep CNN has been developed to classify 1-D ECG arrhythmias in [8]. This study was inspired from these different deep CNN structures used to solve various ECG signal classification and arrhythmia detection problems in the literature.

Machine learning algorithms are widely used in the systems developed to assist the experts. One of the important points in these algorithms is the use of samples of data to reduce the machine learning success. However, the number of samples especially in medical applications such as ECG arrhythmias may not be sufficient. When working on a classification algorithm with multiple classes, generally, if number of samples in each class is equal, the higher the training performance can be obtained. In real world applications, data of all kinds of diseases, especially in the health sector, may not be obtained in equal amounts. In common diseases, more samples can be collected, whereas in rare cases number of samples will be low. Researches are carried out for the classification performance in such imbalanced data sets. Data mining methods are being developed in imbalanced data sets. While developing these methods, the results of statistical metrics and performance measurements are examined. In order to increase the success of learning in imbalanced data sets, the main objective is to improve the recall of the data to the system without damaging the accuracy. However, recall and precise results can often be contradictory, because the number of false positives may increase while increasing the true positive for the minority classes [9].

In this study, it is aimed to increase the performance of the classification of the heart signals obtained from the MIT-BIH arrhythmia database by using CNN algorithm and data mining methods in imbalanced data sets.

II. METHODOLOGY

The extraction of appropriate features from an image or a signal is an important factor in the accuracy rates of the classification algorithms. In this study, one dimension raw ECG beats were obtained from ECG signals by using QRS extraction. The arrhythmias downloaded from the MIT-BIH database were segmented according to American National Standards Institute (ANSI) standard class labels. Data is prepared in the input format suitable for CNN which are explained in the following subsections.

A. Data

The source of the ECG data used for training and testing is obtained from the MIT-BIH Arrhythmia database from the Physionet website. [10]. The data classified according to Association for the Advancement of Medical Instrumentation (AAMI) standard which can be divided to the arrhythmias into 5 main branches which are called as non-ectopic (N), unknown (Q), ventricular ectopic (V), supraventricular ectopic (S) and fusion (F) [11]. These main branches contain different types of arrhythmias.

B. Preparing Data

The data was prepared for the CNN algorithm using Matlab compatible WFDB toolbox [12]. By utilizing the functions in this toolbox, QRS detection was applied the signals in records according to annotation codes which arranged in accordance with ANSI / AAMI-standard. Each QRS block was labeled with the type name. The QRS data of each type is divided into test, training and validation data.

C. CNN Model

CNN algorithms can be considered as multi-layer perceptron (MLP) with a special structure. The CNN algorithm has 3 basic layers: a convolutional linear layer, a pooling layer, and fully-connected layer. Unlike standard machine learning algorithms, CNN can use different feature extraction methods to obtain and classify discriminant features [13]. This means that raw data is given directly as an input of the network. This process is more effortless than the classical methods and the determination of the features with deep learning technique provides higher performance rates.

The proposed CNN architecture was inspired from Acharya and his research team's publication [8]. The CNN architecture includes of 9 layers: 3 convolutional layers, 3 max-pooling layers, and 3 fully-connected layers. Each convolutional layer provides the feature maps and followed by a batch normalization layer, then activated with a rectified linear unit, and down-sampled with the max-pooling layer. Final layer includes the fully-connected layers consist of respectively 30 and 30 and 5 output neurons. The softmax function and classification layer is used to separate outputs to each class which are N, V, Q, S and F.

Two different learning parameters were used in the applied CNN model and the results were compared. First, $3\text{x}10^{-3}$ initial learning rate applied and this value has not been changed during the training, momentum parameter was taken as 0.2 and maximum epoch set as 300. Ten-fold cross-validation was applied with these layer parameters.

Secondly, 0.01 initial learning rate applied and this value has been dropped with the factor of 0.2 during the training, momentum parameter was taken as 0.2 and maximum epoch set as 300. The numbers of majority class samples were subdivided to match the number of minority class samples. And the training was continued until all of the majority class samples were used in train.

D. Dealing with Imbalanced Data

The first approach is to increase the number of epochs. The raw ECG data was given directly to the network without applying any feature extraction method and filtering. 10-fold cross-validation was applied. Data randomly separated to 10 equal pieces. Each fold, one piece of the data was used for test and the rest was used for the train and validation. Increasing the number of training iteration, which is one of the factors of deep learning, increased the possibility of learning the data with few samples in classification algorithm.

In the second method, the train data ensemble which includes same number of samples from each class is constructed. Assume that a rare class A has r samples, and an abundant class B has s samples. $n = \lceil s/r \rceil$ train sessions are required to include all the samples. Thus, the set B is divided into n clusters as $B = [b_1\ b_2\ \cdots\ b_n\ ]$ where each subset b_i has r samples. The train sets are organized as show in Figure 1 for each n training sessions. In the first training, the initial weights are assigned as random. In the k^{th} training the weights of $k\text{-}1^{th}$ train are assigned as the initial weight. The training continues until all clusters of class B is included in the training.

Train 1	A	b_1
Train 2	A	b_2
Train n	A	b_n

Weights and biases are transferred

Figure 1. *Train sets for imbalanced data*

E. Evaluation Criteria

Sensitivity gives information about the percentage of heartbeats which have the relevant arrhythmia disorder arrhythmia disorder is detected correctly. This relationship can be expressed as

$$S_t = \frac{TP}{TP + FN} . 100[\%] \tag{1}$$

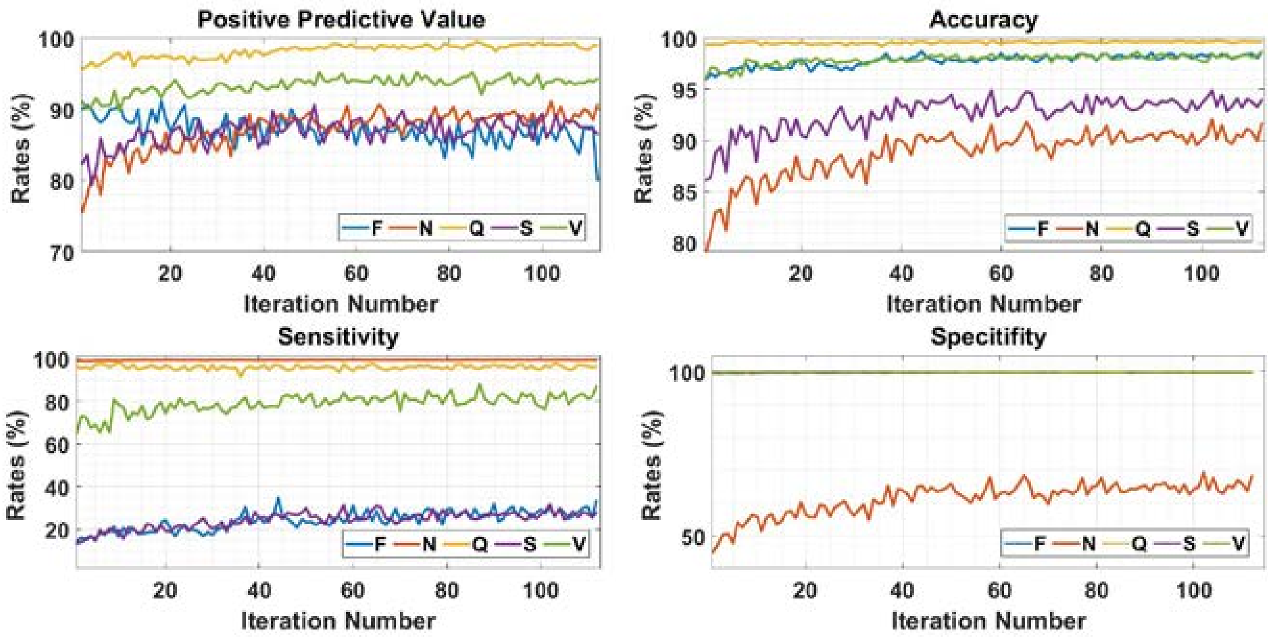

Figure 2. Performance metrics via iterations

Specificity gives the percentage of beats which do not have the relevant arrhythmia disorder is correctly identified. Specificity can be expressed as

$$S_P = \frac{TN}{TN + FP} . 100[\%] \quad (2)$$

Positive predictive value is a criterion of precision that gives the possibility of how the classification model can be correctly classify the beats presence of relevant arrhythmia. The positive predictive value (PPV) is expressed as

$$PPV = \frac{TP}{TP + FP} . 100[\%] \quad (3)$$

Equation (4) indicates the accuracy calculation of the classification algorithm. Accuracy rate A is calculated as

$$A = \frac{TP + TN}{TP + FP + TN + FN} . 100[\%] \quad (4)$$

III. RESULTS AND DISCUSSIONS

In this study CNN algorithm was applied to classify ECG arrhythmia data with imbalanced class sizes. The CNN architecture was created on Matlab and the algorithm runs on a PC workstation with Intel Xeon 3.50 GHz (E5-1650) processor and 64 GB of RAM.

The results are mostly compared with Acharya et al. which has similar structure [8]. It has been shown that the classification performance was increased without using synthetic data. In our first approach of incremented learning, CNN is trained for 300 epochs. The test results of statistical metric comparison of noisy ECG signals classification is given in Table 1. This parameter change makes positive effect on the PPV rates of the low sample classes.

Table 1. Results Comparison for incremented learning

Method	Overall Average Classification Results of Imbalance data of Noisy ECG				
	Type	*SEN %*	*SPEC %*	*PPV %*	*ACC %*
Acharya et al.[8]	N	88,35	95,90	99,04	89,65
With 300 epoch		**98,40**	**96,63**	**99,33**	**98,10**
Acharya et al.[8]	Q	95,45	99,24	90,87	98,96
With 300 epoch		**99,30**	**99,90**	**98,78**	**99,86**
Acharya et al.[8]	V	92,67	98,11	77,62	97,75
With 300 epoch		**95,38**	**99,60**	**94,27**	**99,32**
Acharya et al.[8]	S	85,26	95,11	31,25	94,86
With 300 epoch		**82,22**	**99.09**	**64.92**	**98,75**
Acharya et al.[8]	F	**88,15**	96,99	17,80	96,93
With 300 epoch		85,04	**99,77**	**68,38**	**99,68**

In the second approach new train clusters were created. Totally 90593-N beats, 8040-Q beats ,7235-V beats, 2781 S-beats and 802-F beats were used. These beats split into %80 training and %20 test data. The rare set F has 642 samples for training. Then each train ensemble is constructed with 642 samples from each class. A total of 113 training iterations were implemented with 300 epochs. That means, 1 time 72474-N samples, 11 times

6432-Q, 12 times 5788-V samples, 33 times 2225-S samples and 113 times 642-F samples were participated to the training.

Figure 2 illustrates the variation of performance with training iterations. In all graphs, the classes V and Q which have average data size, introduces stable characteristics. In the positive predictive value, the rare class F starts with higher values around 90% while N and S has lower starting values. With the addition of the new training samples from the other classes, the PPV value of F, N and S classes reaches to similar level. The effect of data size imbalance is seen when the sensitivity and specificity graphs are examined closely. While the number of true positives for the classes F and S increased, the class N, which had more samples, had an effect on the false positives of S and F classes. This reduced the sensitivity percentages of the classes with low sample size. Similarly, the abundant class N has the lowest specifity. The accuracy for each class increase with the iterations, again the abundant class N has the lowest accuracy in the classification.

Table 2. Comparison of SVEB Metrics

Method	SVEB			
	SEN %	*SPEC %*	*PPV %*	*ACC %*
Ince et al.[4]	**81.8**	98.5	63.4	96.1
Jiang et al.[5]	74.9	98.8	78.8	**97.5**
Kiranyaz et al.[6]	68.8	99.5	79.2	96.4
Proposed 2nd Method	26,85	**99,6**	**85,43**	93,72

Table 2 gives the SVEB (means S class versus N, V, and F) percentages comparison of researches which classifies MIT-BIH data with deep learning methods. Since Q-beats have been used in this study, the last line represents the percentage values of S class versus N, V, F and Q. It has been observed that the proposed method increases the specificity and positive predictive value, while the sensitivity is lower than the previous studies.

IV. CONCLUSIONS

Automatic or semi-automated ECG beat classification is important for the development of patient follow-up systems and for assisting specialists. Finding suitable features for the standard classification algorithms is important to get higher accuracies. CNN extracts features from the raw data by applying deep learning technique and trying to find best fit features. In order to increase the classification performance of the CNN architecture, it is important to have equal size classes in the training samples. In this study, data mining methods are applied for imbalanced data while training the CNN without creating any synthetic data. By increasing the number of epochs on a proposed CNN architecture was significantly increased the PPV rates of the classes with low samples. In addition, creating balanced train clusters increased the PPV ratios the 80% and above in non-dominant classes. The results showed that the number of training iteration, which is one of the factors of deep learning, increased the possibility of learning the data with few samples from classification algorithm.

REFERENCES

[1] Rosengarten, J., Maharatna, K., A time domain morphology and gradient based algorithm for ECG feature extraction. In: Proceedings of IEEE international conference on industrial technology (ICIT), pp. 117–122, 2012.

[2] Mironovova, Martina, and Jirí Bíla. "Fast fourier transform for feature extraction and neural network for classification of electrocardiogram signals." In Future Generation Communication Technology (FGCT), 2015 Fourth International Conference on, pp. 1-6. IEEE, 2015.

[3] Sarvan, Çağla, and Nalan Özkurt. "Multiwavelet feature sets for ECG beat classification." In Signal Processing and Communications Applications Conference (SIU), 2017 25th, pp. 1-4. IEEE, 2017.

[4] T. Ince et al., "A generic and robust system for automated patient-specific classification of electrocardiogram signals", IEEE Trans. Biomed. Eng., vol. 56, no. 5, pp. 1415-1426, May 2009.

[5] W. Jiang, S. G. Kong, "Block-based neural networks for personalized ECG signal classification", IEEE Trans. Neural Networks, vol. 18, no. 6, pp. 1750-1761, Nov. 2007.

[6] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks," IEEE Trans. Biomed. Eng., vol. 63, no. 3, pp. 664–675, 2016.

[7] Al Rahhal, Mohamad M., Yakoub Bazi, Haikel AlHichri, Naif Alajlan, Farid Melgani, and Ronald R. Yager. "Deep learning approach for active classification of electrocardiogram signals." Information Sciences 345 (2016): 340-354.

[8] Acharya, U. Rajendra, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, Muhammad Adam, Arkadiusz Gertych, and Ru San Tan. "A deep convolutional neural network model to classify heartbeats." Computers in biology and medicine 89 (2017): 389-396.

[9] Chawla, N. V. (2009). Data mining for imbalanced datasets: An overview. In Data mining and knowledge discovery handbook (pp. 875-886). Springer, Boston, MA.

[10] MIT-BIH Arrhythmia Database Directory, https://www.physionet.org/physiobank/database/html/mitdbdir/mitdbdir.htm (Access Date: 2016).

[11] Association for the Advancement of Medical Instrumentation. (1998). Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms. ANSI/AAMI EC38, 1998.

[12] WFDB Toolbox for MATLAB and Octave, https://physionet.org/physiotools/matlab/wfdb-app-matlab/, November 3, 2017 (Access Date: January 2019)

[13] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.