Independent University **Bangladesh (IUB)**

IUB Academic Repository

Electrical and Electronics Engineering Article

2020-06

Application of Machine Learning on ECG Signal Classification Using Morphological Features

Alim, Anika

IEEE

http://dir.iub.edu.bd:8080/xmlui/xmlui/handle/123456789/512

Downloaded from IUB Academic Repository

2020 IEEE Region 10 Symposium (TENSYMP), 5-7 June 2020, Dhaka, Bangladesh

Application of Machine Learning on ECG Signal Classification Using Morphological Features

Anika Alim
Dept. of Electrical and Electronic Engineering
Independent University, Bangladesh
Dhaka, Bangladesh
Email: anikaalim833@gmail.com

Md. Kafiul Islam
Dept. of Electrical and Electronic Engineering
Independent University, Bangladesh
Dhaka, Bangladesh
Email: kafiul_islam@iub.edu.bd

***Abstract*— An electrocardiogram (ECG) is a simple test that is used to check one's heart's electrical activity. Sensors attached to the skin are used to detect the electrical signal produced by one's heart each time it beats. Many people around the world suffer from cardiovascular diseases. So it is important to detect arrhythmia/abnormal and normal ECG signal more accurately. In this paper, ECG signal is classified by support vector machine (SVM) and neural network. The research is conducted on the normal and arrhythmia ECG datasets obtained from the *PhysioNet* website. The raw ECG data are preprocessed using different filters and then the features are extracted based on morphological values of the waveform. Twelve features are extracted and these are used to train classifiers to classify normal and abnormal ECG data. For SVM classifier, the accuracy is around 87% while for artificial neural network, MATLAB's pattern recognition app is used where the classification accuracy found is around 90% - 93%. The accuracy varies for different numbers of hidden neurons. Diverse ECG databases are used to prove the efficacy and robustness of the use of proposed morphological features and obtained results are compared with existing state-of-the-art research works. This work can be further developed in the future by incorporating deep learning for better performance and it can eventually help to detect cardiac diseases.**

Keywords—ECG, morphological features, SVM, ANN, classification.

I. INTRODUCTION

According to the world health organization (WHO) the number one cause of death around the world is cardiovascular diseases. Around 17.9 million people die every year from cardiovascular disease which is 31% of all deaths globally [1]. Of all cardiovascular diseases, most deaths are due to heart attacks and strokes. A large number of heart attacks and strokes can be prevented by early detection and proper treatment. ECG is one of the most commonly used tests to diagnose cardiac diseases. In recent years, many algorithms and methods have been developed for ECG classification. Machine learning has become very popular and widely used for classification for biosignals such as ECG. In this work, we classify ECG signal by support vector machine (SVM) and artificial neural network (ANN) algorithms with morphological feature extraction method. The main aim is to detect normal and abnormal ECG signal by using the machine learning algorithms along with the feature selection process accurately and in a less complicated way in MATLAB R2018a software.

II. LITERATURE REVIEW

A. Electrocardiogram (ECG)

Electrocardiogram (ECG) is also known as EKG. It is the way to visualize the electrical activity of the heart over a period of time using electrodes placed on the skin. These electrodes capture the tiny electrical changes that are a result of cardiac muscle depolarization followed by repolarization throughout every cardiac cycle (heartbeat) [2]. Dutch physician and physiologist Willem Einthoven invented the first practical electrocardiogram (ECG) [3].

The normal ECG is consist of 'P' wave, a 'QRS' complex, and a 'T' wave. The QRS complex is three separate waves- Q, R and S waves. P wave is generated during the atria depolarize before contraction. On the other hand, QRS complex is produced when the ventricles depolarize before contraction i.e. as the depolarization wave spreads through the ventricles. Hence, 'depolarization waves' refer to both P wave and QRS complex. Finally, T wave is generated during the ventricles recovering from the depolarization state and is called 'repolarization wave'. Thus, the electrocardiogram consists of both depolarization and repolarization waves [4].

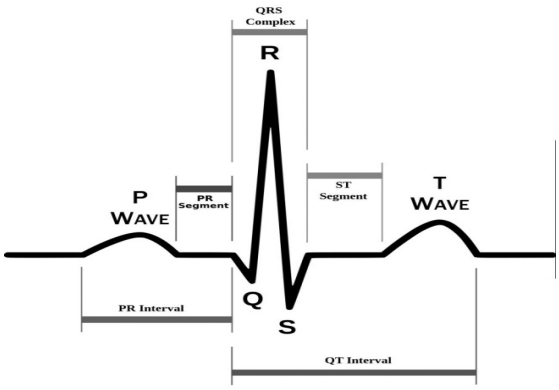


Fig. 1. A normal ECG waveform for one cardiac cycle.

B. Support Vector Machine (SVM)

Support Vector Machine (SVM) is one of the most popular supervised machine learning algorithms which is very efficient for binary and linear classification problems. It is commonly used for both classification and regression analysis. However, it is mostly utilized in classification problems [5]. SVM observes data and sorts them into two categories. It is trained with a series of labelled 2-class data building the model. The job of an SVM algorithm is to find out which class a new data point belongs to.

C. Artificial Neural Network (ANN)

ANN is originally inspired from biological neural networks to recognize any pattern. In practice, ANN has multi-layer fully-connected neural nets whose architecture usually contains an input layer at the left, multiple hidden layers in the middle, and an output layer at the right. Every neuron (i.e. node) in one layer is connected to every alternative neuron or node within the subsequent layer with different weights. The input layer receives various types of information

978-1-7281-7366-5/20/$31.00 ©2020 IEEE

from the outside world and produces some output. This is the labelled data that the network intents to process or learn about based on its weighted connections between different layers of neurons in its training phase. Once the weights are known, based on the inputs, the network calculates its output from the same mathematical equation it learns during the training [6].

III. Materials and Methods

A. Dataset Used

ECG dataset has been prepared with 50 normal and 50 arrhythmia databases. These data are collected from the PhysioNet website [7]. It offers free access to large collections of physiological and clinical data and related open-source software. The four databases are used for this project-

- Fantasia Database [8, 19]
- MIT-BIH Normal Sinus Rhythm Database [8]
- MIT-BIH Arrhythmia Database [8, 20]
- Sudden Cardiac Death Holter Database [8, 21]

B. Proposed Method

The downloaded row data are not suitable for further processing in the MATLAB R2018a. So the first step is to prepare dataset. Before the feature extraction we have to pre-process the signal to remove the unwanted noise. After feature selection a new dataset is created. This dataset is fed to SVM and also to neural network to train them. After the training we can classify normal and arrhythmia ECG signal.

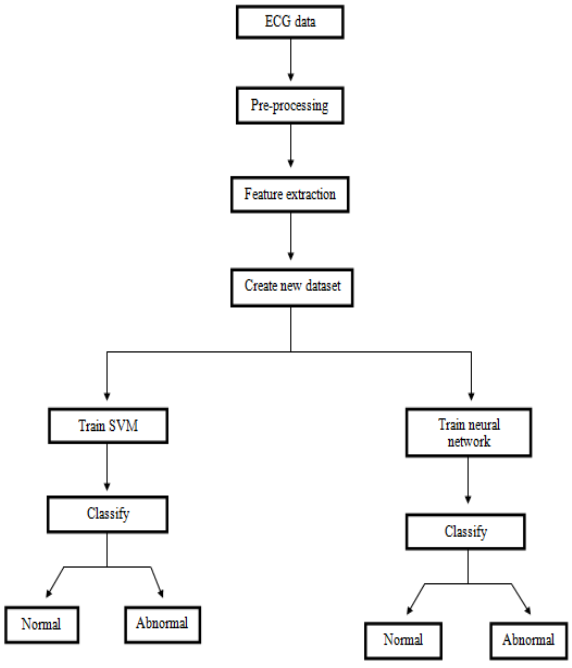


Fig. 2. Process flow chart of the method used in this work.

C. Pre-processing

Normally ECG signals are mixed with noise and artifacts such as baseline wander (due to muscle movement). For the highest accuracy these unwanted noises need to be removed. For denoising different types of filters are used. To remove the power line interface a 2nd order 50 Hz notch filter has been used. High frequency is removed using 128th order 80 Hz FIR low pass filter. As we are using four different types of database the LPF's order and cut off frequency is changed according to the database's cut off frequency.

D. Feature extraction

Feature extraction is very important for classification. For this project we have selected 12 features. It helps to identify normality and abnormality of an ECG signal. We are using morphological method for feature extraction. 12 features are extracted from the signal using algorithm found in [18]. These features are:

- **Maximum heart rate**: A normal resting heart rate of a healthy person is between 60 to 100 bpm.
- **Average heart rate**: It is between 60 to 100 bpm.
- **Minimum heart rate**: It is around 60 to 70 bpm
- **Total number of QRS**: The total number of QRS is not fixed. It changes according how long the ECG signal is.
- **Number of irregular beats**: In normal heart rate there might be few irregular beats for exercising or running. If the percentage of irregular beats is less than 10 then it is normal.
- **Percentage of irregular beats**: Percentage of irregular for normal heart rate is less than 10.
- **Number of episodes with consecutive beats**: It has to be less than 25 for normal heart rate.
- **Average PR interval**: Normal PR interval is between 120 to 200 ms.
- **Average QRS interval**: QRS duration is between 60 to 100 ms.
- **Average QTc interval**: QTc is referred to corrected QT interval. Average QTc< 430 ms is considered as normal.
- **Number of P absence**: When the number of absence P wave is multiply by 100 and divided by total number of QRS complex is less than 8 then the heart rate is consider normal.
- **Number of consecutive P absence**: If the number of consecutive P wave absence is multiply by 100 and divided by total number of QRS complex is less than 1 then the heart rate is normal.

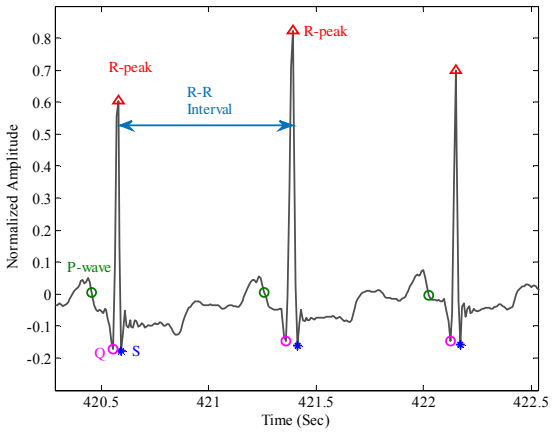


Fig. 3. QRS complex detection and extraction of relevant morphological features.

E. Preparing dataset

A dataset is created with 100 patient's ECG signals which are downloaded from the *PhysioNet* website. Among these 100 patients, 50 is normal ECG signal and 50 is abnormal ECG signal. Each contains the 12 features. So the new dataset contain 100 ECG signals of different patients with 12 features.

F. *Training SVM*

SVM is used as a classification method. After preparing dataset it is fed to SVM for training. After randomizing the data it is divided into train data and test data. 80% of the data is used as training data and 20% is used as test data. We use a linear kernel function to map the training data into kernel space. MATLAB R2018a is used to train SVM.

G. *Training ANN*

We use a two-layer feed-forward network, with sigmoid hidden and *softmax* output neurons. MATLAB's neural network pattern recognition app is utilized for it. Data is fed to the app as input data to present to the network and target data defining desired network output. Then it randomly divided up the 100 samples into 70% training set, 15% validation set and 15% testing set. Then we set different numbers of hidden neurons to get more accurate classification and train the network.

IV. RESULTS AND DISCUSSION

As the ECG signal is classify using two different classifiers, so the result is discuss in two different sections.

A. *SVM classifier result*

We have used kernel linear function for the classification. The data is divided in training set and test set in MATLAB R2018a software. The accuracy is around 87%. The accuracy changes every time we retrain the classifier.

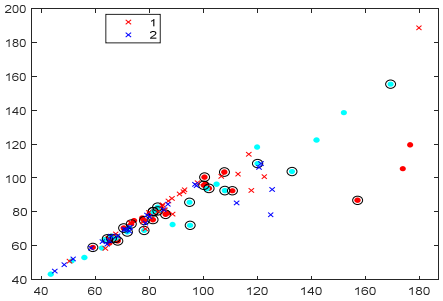


Fig. 4. Scatter diagram of training and test set during DVM

In this classification only 12 features are used to train the classifier. Though the existing studies on SVM are vast and have more accuracy than this project. But the only few of the studies have used these 12 features and four types of database. Table-I provides the comparison with other studies.

Most of the previous studies have used two feature extraction methods. This makes the detection more accurate. But we have used only one feature extraction method and selected only 12 features. For this reason, the process becomes easier although the accuracy is not high. The accuracy can be improved if we find the most dominant features and train the classifier with those features.

B. *ANN classifier result*

The overall accuracy of neural network is 93% for the number of hidden neurons 24. We can change the accuracy and performance by changing the number of hidden neurons. The whole dataset is divided into training, validation and test set. The training set is 70% of the whole dataset and the validation and test set is 15%.

TABLE I. COMPARISON WITH RELATED WORKS ON SVM

Related works	Database	Method	Performance
Ref [9]	MIT-BIH Arrhythmia	2 different feature extraction methods-The wavelet transform and autoregressive modeling (AR)	99.68% accuracy
Ref [10]	MIT-BIH arrhythmia	Discrete wavelet transform (DWT) and principle components analysis (PCA)	Performance is 99.6367% with LIBSVM.
Ref [11]	MIT-BIH arrhythmia	DWT based feature extraction; the R-peaks are detected to determine the HRV signal features.	96% accuracy
Ref [12]	MIT-BIH arrhythmia	Two kinds of features: 1) ECG morphology features and 2) ECG wavelet features with QRS width.	Accuracy: 97% for dataset 1 91% for dataset 2
Ref [13]	INCART 12-lead arrhythmia	Principal Component Analysis (PCA) feature extraction	The accuracy is 76.83% and 98.33% for MSVM and SIMCA classifier respectively
Ref [14]	MIT-BIH arrhythmia	Beat classification and episode detection and classification	95% and 94% accuracy
Ref [15]	MIT-BIH arrhythmia	Feature extraction includes-frequency information, RR intervals, QRS morphology and AC power of QRS detail coefficients	Accuracy is 97.2%
Ref [16]	MIT-BIH arrhythmia	Morphology feature extraction (Wavelet-SVM Method)	Accuracy is 97.59%
This work	Fantasia, MIT-BIH normal sinus rhythm, MIT-BIH arrhythmia, Sudden Cardiac Death Holter	Morphology Feature extraction method (12 features)	87% accuracy

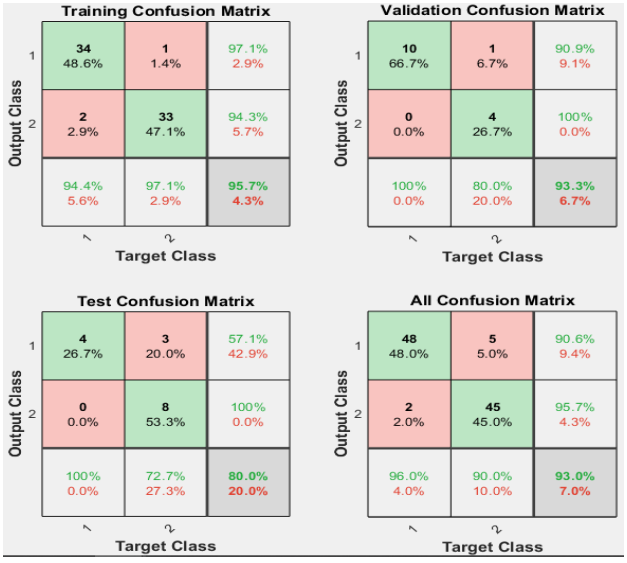


Fig. 5. Confusion matrix of neural network for 24 hidden neuron

From the confusion matrix we get the value of true positive (TP), true negative (TN), false positive (FP), false negative (FN) and total accuracy. Definition of these terms are as follows [17]:

- True positive (TP): Positive prediction means the subject has the disease and it is true prediction.
- True negative (TN): Negative prediction means the subject does not have the disease. This is also a true prediction.
- False positive (FP): The prediction is positive but the subject doesn't actually have the disease (aka "Type-I error").
- False negative (FN): The prediction is negative but the subject actually has the disease (aka "Type-II error").

The performance of neural network changes with different number of hidden neurons. The changes are given below-

TABLE II. CLASSIFICATION PERFORMANCE OF NEURAL NETWORK (PERCEPTRON) FOR DIFFERENT HIDDEN NEURONS

# Hidden Neurons	ANN Classifier's Performance			
	Training	**Validation**	**Test**	**All**
10	TP=34.3%, TN=42.4%, FP=18.6%, FN=5.7%, Accuracy=75.7%	TP=33.3%, TN=53.3%, FP=0%, FN=13.3%, Accuracy=86.7%	TP=33.3%, TN=26.7%, FP=20%, FN=20%, Accuracy=60%	TP=34%, TN=41%, FP=16%, FN=9%, Accuracy=75%
14	TP=42.9%, TN=44.3%, FP=7.1%, FN=5.7%, Accuracy=87.1%	TP=46.7%, TN=33.3%, FP=6.7%, FN=13.3%, Accuracy=80%	TP=40%, TN=46.7%, FP=6.7%, FN=6.7%, Accuracy=86.7%	TP=43%, TN=43%, FP=7%, FN=7%, Accuracy=75.7%
22	TP=40%, TN=50%, FP=7.1%, FN=2.9%, Accuracy=90%	TP=53.3%, TN=46.7%, FP=0%, FN=0%, Accuracy=100%	TP=60%, TN=33.3%, FP=0%, FN=6.7%, Accuracy=93.3%	TP=45%, TN=47%, FP=5%, FN=3%, Accuracy=92%
28	TP=12.9%, TN=45.7%, FP=40%, FN=1.4%, Accuracy=58.6%	TP=6.7%, TN=66.7%, FP=26.7%, FN=0%, Accuracy=73.3%	TP=20%, TN=46.7%, FP=33.3%, FN=0%, Accuracy=66.7%	TP=49%, TN=13%, FP=37%, FN=1%, Accuracy=62%

For 22 hidden neurons the accuracy is 92% and for 24 the accuracy is 93%. So for 21 to 25 hidden neurons the accuracy is 90% to 93%. Below 21 or above 25 hidden neurons the accuracy decline.

V. CONCLUSION

This paper proposed two different machine learning algorithms- SVM and ANN for detecting normal and arrhythmia ECG signals by using morphological features. This study is conducted on four different types of databases of 100 subjects downloaded from *PhysioNet*. The average accuracy found with SVM and ANN are around 87% and 94% (using 24 hidden neurons) respectively. The accuracy is good but can be further improved and the proposed system may be developed for practical use. From the features if we can find the most dominant features (i.e. feature selection) to train the classifier, it may improve the accuracy. We can try to classify using advanced neural network like deep neural network (DNN) and convolutional neural network (CNN). With CNN we can directly use any raw signal to classify which may allow to reduce the computational complexity.

ACKNOWLEDGMENT

We acknowledge the support from Dept. of EEE of Independent University, Bangladesh.

REFERENCES

[1] https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1 (accessed on 9 Febeuray, 2020)

[2] S. T. Alam, M. M. Hossain, M. D. Rahman, M. K. Islam," Towards Development of a Low Cost and Portable ECG Monitoring System for Rural/Remote Areas of Bangladesh ", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.10, No.5, pp. 24-32, 2018.DOI: 10.5815/ijigsp.2018.05.03

[3] https://en.wikipedia.org/wiki/Willem_Einthoven (accessed on 9 February, 2020)

[4] S Pithika, "Electrocardiogram (ECG): Definition and Characteristics | Veterinary Pharmacology," [online document], Available: http://www.biologydiscussion.com/veterinary-pharmacology/electrocardiogram-ecg-definition-and-characteristics-veterinary-pharmacology/74308 (accessed on Dececember 17, 2019)

[5] S. Ray, "Understanding Support Vector Machine algorithm from examples (along with code)" 6 October, 2015. [Online document], https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/ (accessed on 20 December, 2019)

[6] B. Marr, "What Are Artificial Neural Networks - A Simple Explanation For Absolutely Anyone," 24 September, 2018, Available: https://www.forbes.com/ (accessed on December 19, 2019)

[7] https://physionet.org/ (accessed on 20 December, 2019)

[8] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals (2003). Circulation. 101(23):e215-e220.

[9] Q. Zhao and L. Zhang, "ECG feature extraction and classification using wavelet transform and support vector machines," IEEE International Conference on Neural Networks and Brain (ICNN&B), vol. 2, pp. 1089–1092 (2015).

[10] D. Thanapatay, C. Suwansaroj and C. Thanawattano, "ECG beat classification method for ECG printout with Principle Components Analysis and Support Vector Machines", International Conference on Electronics and Information Engineering, vol. 1, pp. 72-75 (2010).

[11] C. Venkatesan, P. Karthigaikumar, A. Paul, S. Satheeskumaran, and R. Kumar, ''ECG signal preprocessing and SVM classifier-based abnormality detection in remote healthcare applications,'' IEEE Access, vol. 6, pp. 9767–9773, (2018).

[12] Y. Bazi, N. Alajlan, H. AlHichri and S. Malek, "Domain adaptation methods for ECG classification", in: IEEE International Conference on Computer Medical Application (ICCMA), January 2013, pp. 1–4.

[13] N. Jannah and S. Hadjiloucas, "Comparison Between ECG Beat Classifiers Using Multiclass SVM and SIMCA with Time Domain PCA Feature Reduction", UKSim-AMSS 19th International Conference on Modeling & Simulation, pp. 126-131, (2017).

[14] M.G. Tsipouras, D.I. Fotiadis and D. Sideris, "Arrhythmia classification system based on the RR-interval signal" Artificial Intelligence in Medicine, vol.33, 237—250, (2005).

[15] Z. Zidelmala, A. Amirou, D. O. Abdeslam and J. Merckle, "ECG beat classification using a cost sensitive classifier", vol. 111 (3), pp.570-577, (2013).

[16] M.K. Gautam, "Morphological Analysis of ECG Arrhythmias using SVM Methodology", International Journal of Computer Science and Information Security (IJCSIS), Vol. 15, No. 5, pp.421-427, (May 2017)

[17] https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/ (accessed on 27 December, 2019)

[18] https://github.com/redxlab/cardio24 (accessed on 15 March, 2020)

[19] Iyengar N, Peng C-K, Morin R, Goldberger AL, Lipsitz LA. Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics. Am J Physiol 1996;271:1078-1084.

[20] Moody GB, Mark RG. The impact of the MIT-BIH Arrhythmia Database. IEEE Eng in Med and Biol 20(3):45-50 (May-June 2001). (PMID: 11446209)

[21] Greenwald SD. Development and analysis of a ventricular fibrillation detector. M.S. thesis, MIT Dept. of Electrical Engineering and Computer Science, 1986.