Proceedings of the Fourth International Conference on Inventive Systems and Control (ICISC 2020)
IEEE Xplore Part Number: CFP20J06-ART; ISBN: 978-1-7281-2813-9

# EEG-ECG Signals Classification for Arrhythmia Detection using Decision Trees

Ghazanfar Latif, Faisal Yousif Al Anezi
Department of Computer Science
Prince Mohammad bin Fahd University
Al-Khobar, Saudi Arabia
glatif@pmu.edu.sa, falanezi@pmu.edu.sa

Mohammad Zikria, Jaafar Alghazo
Department of Computer Engineering
Prince Mohammad bin Fahd University
Al-Khobar, Saudi Arabia
mzikria@pmu.edu.sa, jghazo@pmu.edu.sa

*Abstract*— **The study analysis the electrocardiogram (ECG) and electroencephalogram (EEG) signals classification problem by using decision trees. The analysis and classification of heartbeats and brain traces associated with different types of Arrhythmia and Seizure is an active research topic in recent years. In this paper, we discuss different classification techniques for the analysis and classification of ECG and EEG signals. The training and testing is performed on the MIT-BIH arrhythmia and ECG database. For classification, the random forest tree and naïve Bayes algorithms are used. The results provide an improved performance over the current standards used for classification. The accuracy of the designed system is 97.45% which outperforms the various instance learning and supervised machine learning methods.**

***Keywords— ECG Classification, EEG Classification, Decision Trees, MIT BIH Arrhythmia Datasets***

## I. Introduction

Cardiovascular diseases have become a cause of major brain problems in the recent years indicated by the number of patients admitted in the hospitals. The American Heart Association (AHA) reports that every year in the United States approximately 295,000 emergency cases are caused by cardiac arrests [1]. This number is bound to increase unless some drastic measures are taken into consideration. Seizure is a brain disorder which is caused by sudden Abnormal firing of neurons [2]. At the top right chamber of the human heart, an electrical signal is generated from the Sino Atrial node which stimulates the heartbeat [1]. The heart may experience abnormal increase or decrease in its beat rate which is known as arrhythmia [3]. In order to detect this type of abnormality, an electrocardiogram (ECG) device is used that measures the variations in the electrical signals of the heart.

The aim of this research is to enhance Arrhythmia and Seizure analysis and detection ratios by using various decision tree classifier models. This analysis is devoted to decisions concerned with recognizing and classifying the patterns of the abnormal electrical signals of the heart and brain beats. This work combines different types of classifiers and compares their performance as individual models. The models used in this work include Multilayer Perceptron (MLP), Radial Basis Function (RBF), neural networks, logistical model trees, naïve Bayes trees and random forest trees. The simulation results are based on training and testing the extracted samples from the renowned arrhythmia database i.e. MIT-BIH and ECG Seizures database.

This organization of this paper is as follows: Section II provides a brief overview of recent work in the field of ECG beat detection and classification. Section III describes our infused classification model from different machine learning techniques. The source Data used in this research and the extracted classification features are explained in Sections IV and V, respectively. Section VI discusses the evaluation results of the different proposed classification models. Conclusions and future work is provided in Section VII.

## II. Related Work

The classification of ECG and EEG signals is a very important task in the field of biomedical engineering. It helps in providing early detection and monitoring of cardiac diseases which can play a vital role in patients life. There are different supervised and un-supervised Machine Learning (ML) techniques that are presented in recent years for the early detection of the disease [4-7]. Most of these methods use beats as input records to the system. Different recognition and classification models dedicated for ECG arrhythmia detection were proposed in the literature. An enhanced kernel based approach of SVM machine learning technique is used by the authors in [8] to classify the ECG signals and compare the results with the Linear Discriminant Analysis (LDA) based classification. The SVM based results outperforms the ones from LDA analysis. Another method based on MLP was proposed by Niwas et al. [9] in which an overall accuracy of 99.02% was achieved for nine different classes of arrhythmia. The high accuracy was achieved by training a large number of datasets ranging from 15 to 40 datasets per class. Each data set contained at least 1500 sample beats. Also, Zhu et al. [10] examined three models of binary MLP classification and achieved accuracy above 95% with a small number of training examples. In the study of Zeybekog et al. [11], five different types of ECG signals were classified using Multilayer perceptron (MLP) machine learning techniques. The system classified the input data into Normal ECG signal, Left Bundle Branch Block signal, Ventricular Tachycardia signal, Atrial Fibrillation signal and Right Bundle Branch Block signal. They achieved a general accuracy of 82% in their classification.

There are different machine learning techniques which are quite popular for ECG signals classification i.e. MLP [12], various other Artificial Neural Networks paradigms like fuzzy neural networks [13], radial basis networks [14], support vector machines [15], and self-organizing maps [16]. For radial basis function neural network, different studies achieved high classification results for multiclass recognition. In the study of Guangying et al. [17], they were able to reach a classification rate of 90% for the normal class and paced heart beats class. Moreover, a specificity of 97.42% and sensitivity of 88.37% were recognized for classifying six types of beats. Input data has been classified into various abnormal classes including Atrial Premature Beat (APB), Fusion (F), Premature Ventricular Contraction (PVC) and Right Bundle Branch Block Beat (RBBB). The dataset used in the process was the MIT-BIH database [18].

## III. Proposed Method

This work focuses on five classifier models including Logistic Model (LM), Multilayer Perceptron (MLP), Naïve Bayes (NB), Radial Basis Function (RBF) and Random Forest (RF) for prediction of different types of Arrhythmia and Seizures. The classifications of these models were then integrated into single models which are discussed in the following sections.

### A. *Multilayer Perceptron*

The general MLP training works by utilizing feed forward propagation and error feedback propagation for updating the weights of the model through gradient technique. Based on the results of Ebrahimzadeh et al. experiments, the Levenberg Marquardt (LM) technique has been used in our work where the first hidden layer comprises of 10 hidden units [3]. In the hidden layer a sigmoid activation function has been used, whereas the output layer incorporates a linear function.

### B. *Radial Basis Function Network*

In RBF neural network, the square of distances between the input vector x and mean vector cm for the radial basis functions are computed using $\|\mathrm{x}-\mathrm{c_m}\|^2$ and the output of the mth hidden unit is represented by $y_m$ as shown in Equation 1.

$$\mathrm{e}^{-[\frac{\|\mathrm{x}-\mathrm{c_m}\|^2}{2\mathrm{6}^2}]} \qquad (1)$$

Next, the output values of the hidden units are weighted and summed to produce the final results [11]. In (1) is the spread factor interpreted as the radius of the radial basis function. The spread parameter plays a major rule in classifying the data and a spread factor of 105 was selected experimentally to classify the ECG signals.

### C. *Logistic Model Decision Trees*

In this scheme, the arrhythmia and seizures are classified by utilizing the model trees containing linear regression function at the leaves [17]. LMT works based on the combination of Logistic regression and decision trees to create LM at every node of the trees that refines the constructed model at the higher levels of the tree.

### D. *Naïve Bayes Decision Trees*

Tree augmented NB models are created by addition of directional edges amongst the attributes. The performance of NB classifier has been observed to be better than traditional NB classifier and it also maintains a comparable performance to several popular rival techniques. The standard Bayes rule is defined in Equation 2 [18].

$$\arg_n \max\{\, P(C_n|w)\} = \frac{P(W \mid C_n) \times P\,(C_n)}{P(w)} \qquad (2)$$

Where $P\ (C_n)$ = the prior probability of category n, W = the new web page to be classified. $P(w \mid C_n)$ = the conditional probability of the test page, given category n.

Where $P\ (C_n)$ is the priori probability for n[th] category, w is the new web page for classification. $P(w \mid C_n)$ is the test page's conditional probability, given the n[th] category.

### E. *Random Forest Decision Trees*

RF tree can be referred to as an ensemble classifier, comprising of many decision trees. The mode of classes generated by individual trees is considered as the outcome of RF classifier. If the count of trees in the forest is increased the generalization error tends to converge to a limit. The degree of generalized error is dependent on strength/correlation of individual trees included in the forest.

## IV. ECG Features Extraction

Between 1975 and 1979, 48 annotated ECG records that belong to 47 subjects were analyzed by the BIH Hospital in Boston. 23 records in the built database were chosen randomly from a set of over 4000 24-hour ECG tapes. The other 25 records were chosen to include specific types of arrhythmias [2]. Each record represents approximately 30 minutes of the signal. The sampling frequency was set to 360 Hz. The database consists in total of roughly 109,000 beats that were manually annotated by at least two cardiologists.

In the work, segmentation for the QRS wave and waveform boundary intervals are determined using an algorithm whose implementation is available as part of the ecgpuwave Physio Toolkit software. The part of the function that segments the QRS wave is established on the algorithm of Pan and Tompkins with some modification to exploit the information of the slope. For the end points recognition of the P and T-waves, an algorithm described was used [16]. The evaluation of the system was performed based on [18] and was used as a component of the segmentation function.

In the preprocessing stage, the input dataset is examined to see if it contains any missing segmentation values or class labels. The checks are performed because R wave can make an appearance in the final segmentation which results in missing class labels. Beats corresponding to such missing information are excluded. As a result of preprocessing, from approximately 109,400 total beats, we used 108,232 beats, or 98.6% of the original data and remaining are discarded.

The system classifies the input beats into nine different types of arrhythmia. These arrhythmias classes are named Premature Ventricular Contraction (PVC) class, Atrial Premature Beat (APB) class, Right Bundle Branch Block (RBBB) class, Paced Beat (PB) class, Nodal Escape Beat (NEB) class, Fusion of ventricular and normal beat (FVNB) class, Fusion of paced and normal beat (FPNB) class, Ventricular Flutter Wave (VFW) class and Left Bundle Branch Block (LBBB) class of arrhythmia. Rest of the classes beats are placed in others class. The data split for different classes is given in Table 1. For our second test, we used 200 EEG traces provided by Epilepsy Center, University of Germany. The signals were sampled at 174 Hz. Out of these 200 traces, there are 100 seizure and 100 Non seizure (normal) traces.

TABLE 1. SUMMARY OF EXTRACTED ECG CLASSES DATA

| Type | No of datasets |
|---|---|
| Normal | 74384 |
| APB | 2356 |
| PVC | 6730 |
| PB | 6969 |
| FPNB | 974 |
| Others | 337 |
| NEB | 227 |
| LBBB | 8033 |
| FVNB | 790 |
| RBBB | 7205 |
| VFW | 226 |

## V. RESULTS AND DISCUSSIONS

Five different classifiers methods were applied to classify the input data into 11 classes that contains one normal class and 10 classes which fall under the abnormal category. Sensitivity and specificity were considered as performance measures in comparing t he classification models rather than accuracy since the data is not evenly distributed between classes. Uneven distribution of data will cause accuracy as a measure to be skewed and inaccurate, however, the sensitivity and specificity will give more accurate results. The summary of the results is represented in Table 1 which shows the data distribution of different classes. Table 2 to 6 represents results of different classification model. Each table shows the accuracy, sensitivity and specificity of each class.

As can be seen in Table 2, using RBF as a classifier results in an average of 91.9% sensitivity and 90.8% specificity, while the accuracy averages to 92%.

Table 3 shows the results of using MLP as a classifier. In this case, the average accuracy achieved is 93.74%, with an average sensitivity of 93.7% and average specificity of 92.1%.

TABLE 2. RBF NETWORK RESULTS FOR CLASSIFYING 11 CLASSES OF HEARTBEATS

| Class | Accuracy % | Sensitivity % | Specificity % |
|---|---|---|---|
| **Normal** | 91.68 | 96.00 | 86.90 |
| **APB** | 83.45 | 53.20 | 99.60 |
| **PVC** | 89.74 | 78.30 | 98.80 |
| **PB** | 98.08 | 96.40 | 99.70 |
| **FPNB** | 83.61 | 53.30 | 99.80 |
| **Others** | 100.00 | 0.00 | 100.00 |
| **NEB** | 100.00 | 0.00 | 100.00 |
| **LBBB** | 94.22 | 88.90 | 99.00 |
| **FVNB** | 83.64 | 53.50 | 99.80 |
| **RBBB** | 95.14 | 90.70 | 99.20 |
| **VFW** | 92.84 | 84.50 | 99.90 |
| **Com. Avg.** | **92.00** | **91.9** | **90.8** |

TABLE 3 MULTILAYER PERCEPTRON RESULTS FOR CLASSIFYING 11 CLASSES OF HEARTBEATS

| Class | Accuracy % | Sensitivity % | Specificity % |
|---|---|---|---|
| **Normal** | 93.80 | 98.40 | 88.70 |
| **APB** | 82.83 | 44.50 | 99.90 |
| **PVC** | 93.51 | 86.90 | 99.30 |
| **PB** | 98.07 | 96.70 | 99.40 |
| **FPNB** | 86.89 | 19.20 | 99.90 |
| **Others** | 90.99 | 11.30 | 100.00 |
| **NEB** | 100.00 | 0.00 | 100.00 |
| **LBBB** | 96.22 | 93.50 | 98.80 |
| **FVNB** | 83.48 | 51.80 | 99.90 |
| **RBBB** | 95.24 | 90.20 | 99.80 |
| **VFW** | 91.69 | 10.20 | 100.00 |
| **Com. Avg.** | **93.74** | **93.7** | **92.1** |

Table 4 shows the results of utilizing logistic model tree as a classifier. Results indicate an average accuracy of 95.54%, average sensitivity of 95.5%, and average specificity of 94.8%.

Table 5 shows the results of Naive Bayes tree as a classifier. Results indicate an average accuracy of 95.64%, average sensitivity of 95.7%, and average specificity of 95.3%.

TABLE 4. LOGISTIC MODEL TREES RESULTS FOR CLASSIFYING 11 CLASSES OF HEARTBEATS

| Class | Accuracy % | Sensitivity % | Specificity % |
|---|---|---|---|
| **Normal** | 95.43 | 98.10 | 92.60 |
| **APB** | 86.70 | 67.50 | 99.70 |
| **PVC** | 96.09 | 92.30 | 99.60 |
| **PB** | 98.57 | 97.30 | 99.80 |
| **FPNB** | 87.58 | 70.20 | 99.80 |
| **Others** | 83.14 | 48.90 | 99.90 |
| **NEB** | 82.96 | 36.10 | 99.90 |
| **LBBB** | 96.12 | 92.60 | 99.40 |
| **FVNB** | 86.66 | 66.90 | 99.90 |
| **RBBB** | 96.93 | 94.00 | 99.70 |
| **VFW** | 89.04 | 74.30 | 100.00 |
| **Com. Avg.** | **95.54** | **95.50** | **94.80** |

TABLE 5. NAIVE BAYES TREE RESULTS FOR CLASSIFYING 11 CLASSES OF HEARTBEATS

| Class | Accuracy % | Sensitivity % | Specificity % |
|---|---|---|---|
| **Normal** | 95.81 | 98.20 | 93.30 |
| **APB** | 88.35 | 72.80 | 99.70 |
| **PVC** | 95.71 | 91.60 | 99.50 |
| **PB** | 98.37 | 96.90 | 99.80 |
| **FPNB** | 89.58 | 76.20 | 99.80 |
| **Others** | 82.89 | 45.70 | 99.90 |
| **NEB** | 83.19 | 33.80 | 99.90 |
| **LBBB** | 96.27 | 92.80 | 99.50 |
| **FVNB** | 88.41 | 72.60 | 99.90 |
| **RBBB** | 96.61 | 93.30 | 99.70 |
| **VFW** | 86.43 | 65.80 | 100.00 |
| **Com. Avg.** | **95.67** | **95.70** | **95.30** |

Table 6 shows the results Random Forest Tree acting as the classifier. Results indicate an average accuracy of 97.45%, average sensitivity of 97.5%, and specificity of 95.9%.

Table 7 shows the results comparison of cumulative average of accuracy, sensitivity and specificity of the different classifiers used in this paper. Results clearly indicate that the Random Forest decision tree classifier outperforms not only the classifiers tested in this work, but also results of classifiers mentioned in previous literature. The Random Forest Tree achieve the best results i.e. accuracy, sensitivity and specificity of 97.45%, 97.50% and 95.90% respectively. Clearly, this implies that the classifier of choice would be the Random Forest Tree.

TABLE 6. RANDOM FOREST TREE RESULTS FOR CLASSIFYING 11 CLASSES OF HEARTBEATS

| Class | Accuracy % | Sensitivity % | Specificity % |
|---|---|---|---|
| **Normal** | 96.93 | 99.60 | 94.10 |
| **APB** | 90.48 | 78.50 | 99.90 |
| **PVC** | 97.36 | 94.80 | 99.80 |
| **PB** | 99.30 | 98.70 | 99.90 |
| **FPNB** | 90.69 | 78.90 | 100.00 |
| **Others** | 83.33 | 50.00 | 100.00 |
| **NEB** | 83.94 | 29.50 | 100.00 |
| **LBBB** | 97.32 | 94.70 | 99.80 |
| **FVNB** | 89.11 | 74.50 | 100.00 |
| **RBBB** | 98.04 | 96.10 | 99.90 |
| **VFW** | 88.57 | 72.90 | 100.00 |
| **Com. Avg.** | **97.45** | **97.50** | **95.90** |

A second test was performed on the EEG datasets for Seizure Detection. Table 8 shows the results of EEG for different classifiers which are also promoting just like ECG classification by using decision trees. Table 8 shows that the Logistic Model Tree achieved the best results for EEG classification with an average accuracy of 95.58%, average sensitivity of 95.6%, and average specificity of 95.6%.

As mentioned above, due the non-uniform distribution of data, the indicators of sensitivity and specificity would best reflect the results of the classifiers used in this paper. In all cases, the results remain the same with Random Forest indicating the best results for ECG classification and Logistic Model tree for EEG classification.

TABLE 7. SUMMARY OF PERFORMANCE RESULTS OF THE PROPOSED ECG CLASSIFICATION MODELS

| Classifier Type | Avg. Accuracy % | Avg. Sensitivity % | Avg. Specificity % |
|---|---|---|---|
| RBF Network | 92.00 | 91.90 | 90.80 |
| Multilayer Perceptron | 93.74 | 93.70 | 92.10 |
| Logistic Model Trees | 95.54 | 95.50 | 94.80 |
| NB Tree | 95.69 | 95.70 | 95.20 |
| Random Forest Tree | **97.45** | **97.50** | **95.90** |

TABLE 8. SUMMARY OF PERFORMANCE RESULTS OF THE PROPOSED EEG CLASSIFICATION MODELS

| Classifier Type | Avg. Accuracy % | Avg. Sensitivity % | Avg. Specificity % |
|---|---|---|---|
| RBF Network | 92.64 | 92.60 | 92.60 |
| Multilayer Perceptron | 94.11 | 94.10 | 94.10 |

| | | | |
|---|---|---|---|
| Logistic Model Trees | 95.58 | 95.60 | 95.60 |

## CONCLUSION

In this work, a classification process is achieved that uses different ML algorithms like MLP, RBF NN, Logistic Model Tree, NB Tree and Random Forest Trees. The MIT BIH arrhythmia ECG dataset is separated into 11 classes. A binary classification of EEG dataset by using these classifiers for Seizure Detection is also provided. The results indicate that the Random Forest Tree classifier attained the best performance in detecting arrhythmias. Furthermore, these results are comparable to the most recent available works and they outperform the recent work in most cases. For future work, we aim to implement more sophisticated integrating and ensemble methods that can optimize this research's results. Also, we look forward to classifying further classes of arrhythmia and Seizure. Moreover, we also aim at applying our suggested models on different well-representative databases.

## REFERENCES

[1] M. Valderrama, C. Alvarado, S. Nikolopoulos, J. Martinerie, C. Adam, V. Navarro, M. Le Van Quyen, "Identifying an increased risk of epileptic seizures using a multi-feature EEG–ECG classification", *Biomedical Signal Processing and Control 7 (2012) 237– 244.*

[2] Tani s Mar, Sebastian Zaunseder, Juan Pablo Martinez, Mariano Llamedo, and Rudiger Poll, "Optimization of ECG Classification by Means of Feature Selection", *Ieee transactions on Biomedical Engineering, Vol. 58, no. 8, August 2011, 2168-2177.*

[3] Ataollah, E., & Khazaee, A. (2010). Detection of premature ventricular contractions using MLP neural networks: A comparative study. *Measurement, 103-112.*

[4] Elhaj, Fatin A., et al. "Arrhythmia recognition and classification using combined linear and nonlinear features of ECG signals." *Computer methods and programs in biomedicine* 127 (2016): 52-63.

[5] shensheng Xu, Sean, Man-Wai Mak, and Chi-Chung Cheung. "Deep neural networks versus support vector machines for ECG arrhythmia classification." *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2017.

[6] Luz, Eduardo José da S., et al. "ECG-based heartbeat classification for arrhythmia detection: A survey." *Computer methods and programs in biomedicine* 127 (2016): 144-164.

[7] Houssein, Essam H., Moataz Kilany, and Aboul Ella Hassanien. "ECG signals classification: a review." *International Journal of Intelligent Engineering Informatics* 5.4 (2017): 376-396.

[8] Varatharajan, R., Gunasekaran Manogaran, and M. K. Priyan. "A big data classification approach using LDA with an enhanced SVM method for ECG signals in cloud computing." *Multimedia Tools and Applications* 77.8 (2018): 10195-10215.

[9] Issac Niwas, S.; Shantha Selva Kumari, R.; Sadasivam, V.; , "Artificial neural network based automatic cardiac abnormalities classification," Computational Intelligence and Multimedia Applications, 2005. Sixth International Conference on , vol., no., pp. 41- 46, 16-18 Aug. 2005.

[10] Farah Nur Atiqah Francis Abdullah, Fazly Salleh Abas, Rosli Besar, "ECG Classification using Wavelet Transform and Discriminant Analysis", *2012 International Conference on Biomedical Engineering (ICoBE),27-28 February 2012,Penang 191-196.*

[11] Zeybekoglu, ;.S O z an, M., "Classification of ECG Arrythmia beats ",with Artificial Neural Networks*Biomedical Engineering Meeting (BIYOMUT), 2010 15th National , vol., no., pp.1-4, 21-24 April 2010.*

[12] Y. H. Hu, S. Palreddy, and W. Tompkins, "A patient-adaptable ECG beat classifier using a mixture of experts approach," IEEE Trans. Biomed. Eng., vol. 44, no. 9, pp. 891–900, Sep. 1997.

[13] S. Osowski and T. H. Linh, "ECG beat recognition using fuzzy hybrid neural network," IEEE Trans. Biomed. Eng., vol. 48, no. 11, pp. 1265–1271, Nov. 2001.

[14] N. Maglaveras, T. Stamkopoulos, K. Diamantaras, C. Pappas, and M. Strintzis, "ECG pattern recognition and classification using non-linear transformations and neural networks: A review," Int. J. Med. Inf., vol. 52, no. 1–3, pp. 191–208, 1998.

[15] K. Park, B. Cho, D. Lee, S. Song, J. Lee, Y. Chee, I. Kim, and S. Kim, "Hierarchical support vector machine based heartbeat classification using higher order statistics and hermite basis function," in Proc. Comput. Cardiol., Sep., 2008, pp. 229–232.

[16] Jun Dong, Jia-fei Tong and Xia Liu, "The Abnormal vs. Normal ECG Classification Based on Key Features and Statistical Learning", *HAIS 2010, Part I, LNAI 6076, pp. 136–143, 2010.*

[17] Yang Guangying; Chen Yue; , "The Study of Electrocardiograph Based on Radial Basis Function Neural NetworkThird International Symposium on ," Intelligent Information Technology and Security Informatics (IITSI), pp.143-145, 2-4 April 2010.

[18] Dogan, B.; Korurek, M.; , "Performance evaluation of Radial Basis Function Neural Network on ECG beat classification," *Biomedical Engineering Meeting. BIYOMUT 2009. 14th National , pp.1-4, 20-22 May 2009*.