Ensemble Neural Network Algorithm for Detecting Cardiac Arrhythmia

S. Aruna and L.V. Nandakishore

Abstract Cardiac arrhythmias are electrical malfunctions in rhythmic beating of the heart. Sometimes, they cause life-threatening conditions. Hence, they need to be diagnosed quickly and accurately to save life and prevent further complications and effective management of the disease. In this paper, we propose an ensemble neural network algorithm to detect arrhythmia. Bagging approach with multilayer perceptron and radial basis neural networks is used to classify the standard 12-lead Electrocardiogram (ECG) recordings in the cardiac arrhythmia database available in UCI Machine Learning Repository. The classification performance of the diagnostic model was analyzed using the following performance metrics, namely precision, recall, F-measure, accuracy, mean absolute error, root mean square error, and area under the receiver-operating curve. The classifier accuracy obtained for the ensemble neural network (ENN) model is 93.9 and 94.9 % for ENN-RBFN and ENN-MLP, respectively.

Keywords Bagging · Cardiac arrhythmia · Correlated feature selection · Multilayer perceptron · Radial basis function neural networks

1 Introduction

Cardiac arrhythmias are abnormal rhythmic activity of heart. Arrhythmias are categorized into two types, those starting in the upper two chambers (atria or auricles) and those starting in the lower two chambers (ventricles). Arrhythmias

S. Aruna (✉)
Department of Computer Science, A.M. Jain College, Meenambakkam,
600114 Chennai, India
e-mail: arunalellapalli@yahoo.com

L.V. Nandakishore
Department of Mathematics, M.G.R Educational and Research Institute University,
Maduravoyal, 600095 Chennai, India
e-mail: lvnandaishore@gmail.com

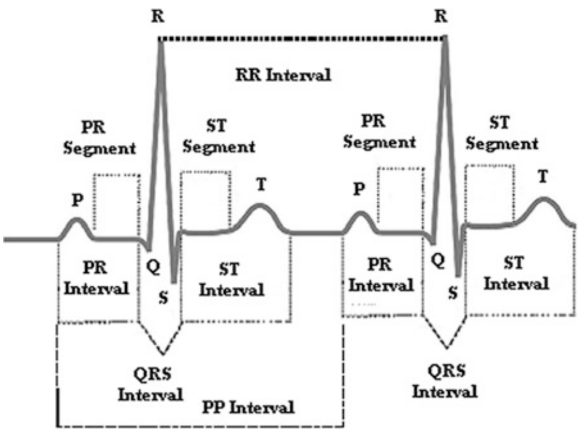
L.P. Suresh et al. (eds.), *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, Advances in Intelligent Systems and Computing 324,
DOI 10.1007/978-81-322-2126-5_4

starting at ventricles are more serious than those starting at auricles. The normal resting heart rate is about 60–100 beats per minute (bpm). According to the speed of the heart rate, arrhythmias can be further categorized into bradycardia, tachycardia, premature contraction, and fibrillation. Arrhythmias represent a serious threat to the patient recovering from acute myocardial infarction, especially ventricular arrhythmias such as ventricular tachycardia (VT) and ventricular fibrillation (VF) [1]. In the USA, more than 850,000 people are hospitalized every year for arrhythmia [2]. Ventricular tachycardia kills an estimated 120,000 people in the UK each year [3]. The detection of arrhythmia is an important task in clinical reasons which can initiate life-saving operations [4]. Some patients do not have any symptoms of arrhythmia. They are diagnosed during routine examination. Treatment varies depending on the type of arrhythmia. Several methods for automated arrhythmia detection have been developed in the past few decades to simplify the monitoring task [5].

Electrocardiogram (ECG) records the electronic activities of the heart and has been widely adopted for diagnosing cardiac arrhythmia [6]. The state of cardiac health is generally reflected in the shape of the ECG waveform and heart rate and contains important pointers to the nature of the disease attacking the heart [7]. Figure 1 shows the normal ECG signal. The normal ECG signal is described by *P*–QRS–*T* waves. *P* wave records the atrial depolarization. QRS complex records the ventricular depolarization. *T* wave records the repolarization of the ventricles. *PP* interval records the atrial rate. *RR* interval records the ventricular rate. *PR* interval measures the AV node function. *ST* interval is measured from the point at which QRS complex finishes with the end of *T* wave. *PR* segment connects *P* and QRS complex. *ST* segment represents the period of depolarization of the ventricles.

In this paper, we propose an ensemble neural network (ENN) algorithm based on bagging to diagnose cardiac arrhythmia from 12-lead ECG recordings. The rest of the paper is organized as follows. Section 2 gives details about the dataset used for the experiment and ensemble neural network algorithm. Section 3 gives the results obtained. Concluding remarks are given in Sect. 4 to address further research.

Fig. 1 Normal ECG signal



2 Materials and Methods

2.1 Dataset Description

The cardiac arrhythmia database used in this study was obtained from the UCI Machine Learning Repository [8]. The dataset has 452 instances of 16 classes. Class 01 is normal, classes 02–15 represent different types of cardiac arrhythmias, and unclassified data come under class 16. Each instance has 279 attributes, of which first four attributes, namely age, sex, height, and weight, give the general description of the patient. Remaining attributes are extracted from the standard 12-lead ECG recordings. There are 206 linear-valued attributes, and the rest is nominal-valued attributes.

2.2 Ensemble Neural Network Algorithm

The ENN algorithm is based on the bagging approach of the ensemble classification method. Bagging is a statistical resample and combine technique [9] based on bootstrapping and aggregating techniques. Bootstrap resampling technique generates multiple versions of the predicting model. Then, the aggregating technique combines those together [10]. Bagging reduces the variance for the classifier. Artificial neural networks (ANN), namely multilayer perceptron (MLP) and radial basis function neural networks (RBFN), were used as base classifiers for the diagnostic model. ANN is mathematical models inspired by biological neural networks where nodes represent neurons and arcs represent axons, dendrites, and synapses. MLP is a feedforward neural network with three layers, the input layer, one or more hidden layers, and output layer. The training and testing vectors presented to the input layer are processed by hidden and output layers. The computational capabilities of MLP are presented by Lippman [11]. RBF networks have a static Gaussian function as the nonlinearity for the hidden layer processing elements and the Gaussian function, responds only to a small region of the input space where the Gaussian is centered [12]. The key to a successful implementation of these networks is to find suitable countries for the Gaussian functions [13]. Figure 2 shows the ENN diagnostic model.

The sequence of steps in the ENN algorithm is as follows:

```
Step 1: Calculation of mean and mode for all the attributes
in the training set.
Step 2: Replacement of missing values of the attributes
using the values obtained in Step 1.
Step 3: Calculation of correlation coefficients for all the
attributes.
Step 4: Removal of attributes with low-class correlation
coefficients.
```

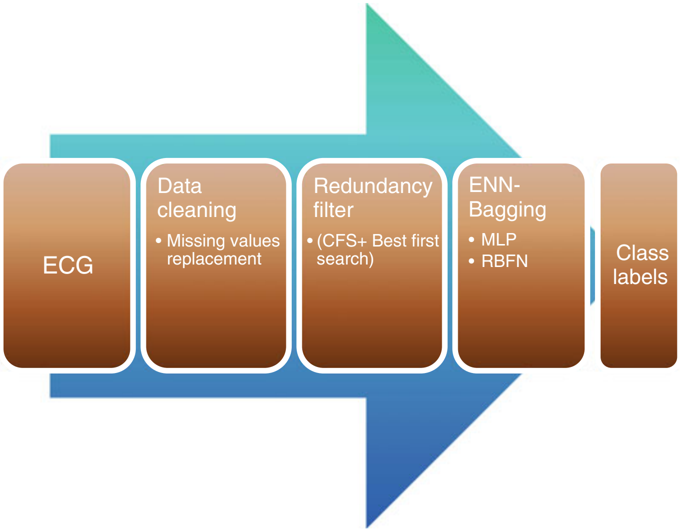


Fig. 2 ENN diagnostic model

Step 5: Selection of best feature subsets of the remaining attributes using best first search criterion.
Step 6: Building predictor model from the feature subset obtained in Step 5 using bootstrap resampling with aggregation for MLP and RBFN classifiers.
Step 7: Prediction of class labels using the model built in Step 6.

The ENN diagnostic model has two phases, data preprocessing phase and classification phase. In the data preprocessing phase, two filters are used. The arrhythmia database has noisy, redundant, and about 0.33 % of missing values which may cause error in classification. Hence, attribute-based filter is first applied to replace the missing values with the modes and means of the training data. Then, to remove the redundant attributes from the dataset correlation feature selection (CFS), redundancy filter is applied. Correlation coefficients for all the attributes were computed using Eq. 1 where CL is the correlation between the summed feature subsets and the class variable, N is the total number of subset attributes, A_{C} is the average of the correlations between the class variable and the subset of attributes, and A_{I} is the average intercorrelation between a subset of attributes.

$$\mathrm{CL} = \frac{N\overline{A_{\mathrm{C}}}}{\sqrt{N + N(N-1)\overline{A_{\mathrm{I}}}}} \tag{1}$$

Feature subsets having high class correlation and low intercorrelation are selected using best first search criterion after removing features with low CL values. Finally, the resultant feature subset is classified using the bagging-based ENN approach with MLP and RBFN as base classifiers.

3 Results

WEKA [14], Java-based data mining tool, is used for conducting the experiments with tenfold cross-validation. Tenfold cross-validation has been proven to be statistically good enough in evaluating the performance of the classifier [15].

3.1 Performance Metrics

The performance criterion for the diagnostic model is analyzed by computing precision, recall, *F*-measure, accuracy, mean absolute error (MAE), root mean square error (RMSE), and area under the ROC (AUC) from the confusion matrix. The precision is the computational measure of predictive accuracy of a particular class. Recall is the measure of positive samples predicted as positive. Accuracy is the ratio of the predictions that are correct. *F*-Measure. MAE is the statistical measure of how far is the estimated value from the actual value. RMSE measures the difference between the actual values and the values predicted by the model. Precision, recall, *F*-measure, accuracy, MAE, and RMSE are calculated using Eqs. 2–7.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{3}$$

$$F\text{ - Measure} = 2 * \frac{\text{precision} - \text{recall}}{\text{precision} + \text{recall}}. \tag{4}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}. \tag{5}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} |P_i - C_i|. \tag{6}$$

$$\mathrm{RMSE} = \sqrt{\left\langle (M_i - C_i)^2 \right\rangle}. \tag{7}$$

TP is the true positives, FP is the false positives, TN is the true negatives, FN is the false negatives, N is the number of instances, P_i is the prediction value of the instance i, C_i is the class value of the instance i, and M is the measured value of the instance i. AUC analyzes the variance independent of the decision's sensitivity. The AUC is obtained by the nonparametric method based on the Wilcoxon's trapezoidal rule to approximate the area [4].

3.2 Diagnosis of Cardiac Arrhythmia

Cardiac arrhythmia database is classified by both linear and ensemble classifiers for neural networks. From 450 attributes after replacing the missing values, CFS selected 14 attributes among 279 attributes. The total number of subsets examined is 3,289. Merit of the best subset found is 0.592, and then, bagging-based ENN-MLP and ENN-RBFN are used for classification of the dataset with 14 attributes. Bag size percent used is 100. Five hundred epochs are used for ENN-MLP. For ENN-RBFN, the ridge value for regression set is 1.0E−8, minimum standard deviation is 0.1, and the number of k-means cluster is 2. Table 1 shows the comparison results for arrhythmia classification by MLP, RBFN, ENN-MLP, and ENN-RBFN classifiers. Figure 3 shows the ROC for ENN-MLP and ENN-RBFN classifiers. A receiver-operating characteristic (ROC) curve is constructed by plotting false-positive rate versus the true-positive rate for varying cutoff values.

ROC analysis originated in electrical engineering in the early 1950s where the technique was developed to assess the performance of signal detection devices (receivers) and later spread into other fields, finding useful applications in both psychology and medical diagnosis [16]. From the results, apparently ENN-MLP achieved better classification performance than all the other classifiers. The results infer that ensemble approach improved the classification performance of both MLP and RBFN classifiers.

Table 1 Cardiac arrhythmia classification results

Performance metrics	Linear classifiers		Ensemble classifiers	
	RBFN	MLP	ENN-RBFN	ENN-MLP
Precision	0.89	0.71	0.94	0.95
Recall	0.88	0.62	0.94	0.95
F-measure	0.88	0.66	0.94	0.95
Accuracy (%)	88.14	62.03	**93.9**	**94.9**
MAE	0.15	0.45	0.11	0.06
RMSE	0.30	0.47	0.23	0.22
AUC	0.82	0.49	0.92	0.94

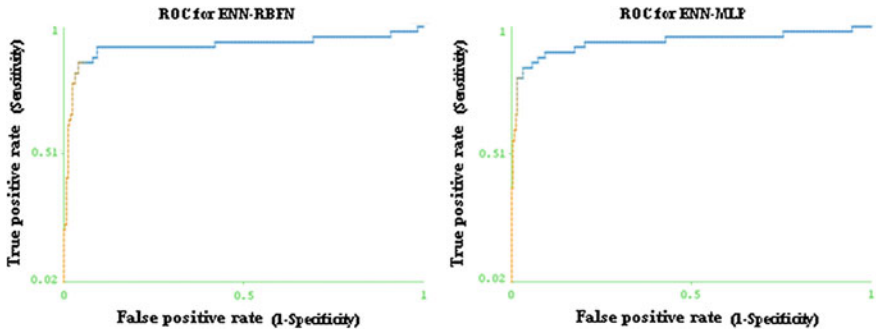


Fig. 3 ROC for ENN-RBFN and ENN-MLP for ten iterations

3.3 Related Work

In [17], the authors used neural network with weighted fuzzy membership function to classify cardiac arrhythmia with an accuracy of 81.32 %. In [18], Elsayad used learning vector quantization neural networks and obtained the classification accuracy of 76.92 %. Zuo et al. [4] used kernel difference KNN to detect arrhythmia. They achieved an accuracy of 70.66 %. In [20], authors used weighted fuzzy artificial immune recognition system for medical diagnosis. They obtained an accuracy of 80.71 % for ECG arrhythmia database. Uyar et al. [22] classified arrhythmia using a serial fusion of support vector machines and logistic regression with an accuracy of 76.1 %. In [23], authors used a novel pruning approach using expert knowledge for data running for diagnosing the arrhythmia with an accuracy of 68.47 %. Ozcan [24] used fuzzy support vector machines for ECG arrhythmia classification. For adaptive neuro fuzzy inference system (ANFIS) and fuzzy

Table 2 Comparison of classification accuracies of our method and other methods in the literature

Model		Accuracy (%)
NEWFM [17]		81.32
HLVQ [18]		76.92
KDFW-KNN [19]		70.66
Fuzzy weighted AIRS [20]		80.71
SVM with Gaussian kernel [22]		76.10
A novel pruning approach [23]		68.47
ANFIS		79.43
FSVM-DTCM [24]		83.33
MNN model		82.22
GFFNN model		82.35
MLP model [25]		86.67
Our method	ENN-RBFN	93.90
	ENN-MLP	94.91

support vector machines distance to class mean method they got accuracy of 79.43 and 83.33 %, respectively. In [25], authors proposed an effective ANN-based approach for cardiac arrhythmia classification. They got classification accuracy of 82.22, 82.35, and 86.67 % for modular neural network model, generalized feed-forward neural network model, and multilayer perceptron model, respectively (Table 2).

4 Conclusion

Cardiac arrhythmias are the irregular rhythm of the heart show a serious medical problem sometimes threatening the life. Detection of arrhythmia is an important task in effective management of the disease. Automated system for arrhythmia detection helps the physicians in simplifying the task. In this paper, we propose an ensemble neural network algorithm based on the bagging approach for detecting the presence or absence of arrhythmia. Multilayer perceptron and radial basis function neural networks are used as base classifiers for the proposed diagnostic model. The performance of the model was evaluated using the following metrics, namely precision, recall, F-measure, accuracy, mean absolute error, root mean square error, and area under ROC. The experiments were conducted using the WEKA data mining tool with tenfold cross-validation. Cardiac arrhythmia database from the UCI Machine Learning Repository is used for the study. Ensemble methods achieved a classification accuracy of 94.9 and 93.9 % for ENN-MLP and ENN-RBFN, respectively. In the present work, the ensemble classifier is used for detecting the presence or absence of arrhythmia. The future work will be concentrated in classifying different types of arrhythmias. Different types of ensemble approaches with neural networks will be compared with the present model for a broader experimental evaluation and further enhancement of the algorithm.

References

1. K.H. Ryu, A data mining approach and framework of intelligent diagnosis system for coronary artery disease prediction, IEICE technical report (2008), pp. 33–34
2. H.G. Lee, K.H. Ryu, A data mining approach for coronary heart disease prediction using HRV features and carotid arterial wall thickness. Bio. Eng. Inform. 200–206 (2008)
3. R.D. Raut, S.V. Dudul, Arrhythmias classification with MLP neural network and statistical analysis. First IEEE International Conference on Emerging Trends in Engineering and Technology (2008), pp. 553–558
4. W.M. Zuo, W.G. Lu, K.Q. Wang, H. Zhang, in *Computers in Cardiology* Diagnosis of cardiac arrhythmia using kernel difference weighted KNN classifier (2008), pp. 253–256
5. D.M. Dubin, Rapid Interpretation of EKG's. (2001)
6. L. Breiman, Bagging predictors. Mach. Learn. **24**, 123–140 (1996)

7. J. Carney, P. Cunningham, The neural BAG algorithm: optimizing generalization performance in Bagged Neural Networks, in *Proceedings of the 7th European Symposium on Artificial Neural Networks* (1999), pp. 35–40
8. R. Lippman, An introduction to computing with neural nets. IEEE Trans. ASSP Mag. **4**, 4–22 (1987)
9. M.D. Buhmann, *Radial Basis Functions: Theory and Implementations Cambridge Monographs on Applied and Computational Mathematics* (Cambridge University Press, Cambridge, 2003)
10. S.V. Chakravarthy, J. Ghosh, Scale based clustering using radial basis function networks, in *Proceeding of IEEE International Conference on Neural Networks* (1994), pp. 897–902
11. P. Baldi, S. Brunak, Y. Chauvin, Assessing the accuracy of prediction algorithms for classification: and overview. Bioinformatics **5**, 412–424 (2000)
12. J.A. Hanley, B. McNeil, The meaning and use of the area under a receiver operating characteristic (ROC) Curve. Radiology **143**, 29–36 (1982)
13. T. Fawcett, *ROC Graphs: Note and Practical Considerations for Data Mining Researchers*, HP Labs Technical Report, (2003)
14. S.H. Lee, J.K. Uhm, J.S. Lim, Extracting input features and fuzzy rules for detecting ecg arrhythmia based on NEWFM. International Conference on Intelligent and Advanced Systems
15. A.M. Elsayad, Classification of ECG arrhythmia using learning vector quantization neural networks, in *Proceedings of Computer Engineering Systems* (2009), pp. 139–144
16. K. Polat, S. Şahan, S. Güneş, A new method to medical diagnosis: Artificial immune recognition system (AIRS) with fuzzy weighted pre-processing and application to ECG arrhythmia. Expert Syst. Appl. **31**, 264–269 (2006)
17. A. Uyar, F. Gurgen, Arrhythmia classification using serial fusion of support vector machines and logistic regression, in *Proceedings of Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications* (*IDAACS 2007*) (2007), pp. 560–565
18. A.M. Mahmood, M. Rao Kuppa, A novel pruning approach using expert knowledge for data specific pruning. Eng. Comput. **28**, 21–30 (2012)
19. N.O. Ozlem, F. Gurgen, Fuzzy support vector machines for ECG arrhythmia detection, in *Proceedings of Pattern Recognition* (*ICPR 2010*) (2010), pp. 2973–2976
20. S.M. Jadhav, S.L. Nalbalwar, A.A. Ghatol, Artificial neural network models based cardiac arrhythmia disease diagnosis from ECG signal data. Int. J. Comput. Appl. **44**, 8–13 (2012)