Methodology Article

RR Interval-based Atrial Fibrillation Detection using Traditional and Ensemble Machine Learning Algorithms

Abstract

Atrial fibrillation (AF) is a life threatening disease and can cause stroke, heart failure, and sometimes death. To reduce the rate of mortality and morbidity due to increased prevalence of AF, early detection of the same becomes a prior concern. Traditional machine learning (TML) algorithms and ensemble machine learning (EML) algorithms are proposed to detect AF in this article. The performances of both these methods are compared in this study. Methodology involves computation of RR interval features extracted from electrocardiogram and its classification into: normal, AF, and other rhythms. TML techniques such as Classification and Regression Tree, K Nearest Neighbor, C4.5, Iterative Dichotomiser 3, Support Vector Machine and EML classifier such as Random Forest (RF), and Rotation Forest are used for classification. The proposed method is evaluated using PhysioNet challenge 2017. During the tenfold cross validation, it is observed that RF classifier provided good classification accuracy of 99.10% with area under the curve of 0.998. Apart from contributing a new methodology, the proposed study also experimentally proves higher performance with ensemble learning method, RF. The methodology has many applications in health care management systems including defibrillators, cardiac pacemakers, etc.

Keywords: *Atrial fibrillation, area under the curve, C4.5, classification and regression tree, Discrete wavelet transform, Electrocardiogram, Iterative Dichotomiser 3, K-NN, Random Forest, rotation forest, Support Vector Machine*

S. K. Shrikanth Rao[1], Roshan Joy Martis[2]

[1]Department of Electronics and Communication Engineering, Vivekananda College of Engineering and Technology, Puttur, Karnataka, India, [2]Department of Computer Science Engineering, Global Academy of Technology, Bengaluru, Karnataka, India

Submitted: 10-Jan-2022 **Revised:** 23-Apr-2022 **Accepted:** 16-May-2022 **Published:** 12-Jul-2023

Introduction

Atrial fibrillation (AF) is one of the most commonly observed types of cardiovascular disease which is clinically important. Worldwide, it occurs with an incidence and prevalence rate of 0.4% and 2.75% per year, respectively.[1,2] According to the recent report of American Heart Association, more than 33 million individuals are affected by AF worldwide.[3,4] Every year 5 million new cases of AF are being reported globally.[5] The statistics of AF indicate that the incidence of AF is progressively increasing with years. Furthermore, it is noted that the incidence of AF is higher in the elderly population. The prominent causes of AF include obesity, physical inactivity, increased intake of fatty food, smoking habit, consumption of alcohol, and family inheritance. Clinically, AF can easily be diagnosed by the electrocardiogram (ECG) pattern. Usual findings of an individual's ECG with AF include irregularity in the R peak to R peak time interval followed by the absence of P wave.

Many authors have proposed a number of computational techniques and algorithms applying Machine Learning (ML) and Deep Learning (DL) approaches in diagnosing AF using digital ECG signal. Graphical abstract of the proposed approach is shown in Figure 1. Table 1 provides the literature review of some prominent work related to ECG classification.

Luo *et al*.[6] proposed polar coordinate transformation method on Poincare plot and recorded specificity of 99.14%. Radhakrishnan *et al*.[7] proposed chirplet transform method using deep convolutional Bidirectional Long Short-Term Memory (BiLSTM) network and achieved accuracy of 99.18%. Shi *et al*.[8] suggested convolutional neural network (CNN)-based

How to cite this article: Rao SK, Martis RJ. RR interval-based atrial fibrillation detection using traditional and ensemble machine learning algorithms. J Med Sign Sens 2023;13:224-32.

Address for correspondence:
Prof. S. K. Shrikanth Rao, Department of Electronics and Communication Engineering, Vivekananda College of Engineering and Technology, Puttur D.K -574201, Karnataka, India.
E-mail: skanthrao@gmail.com

Access this article online
Website: www.jmssjournal.net
DOI: 10.4103/jmss.jmss_4_22
Quick Response Code:

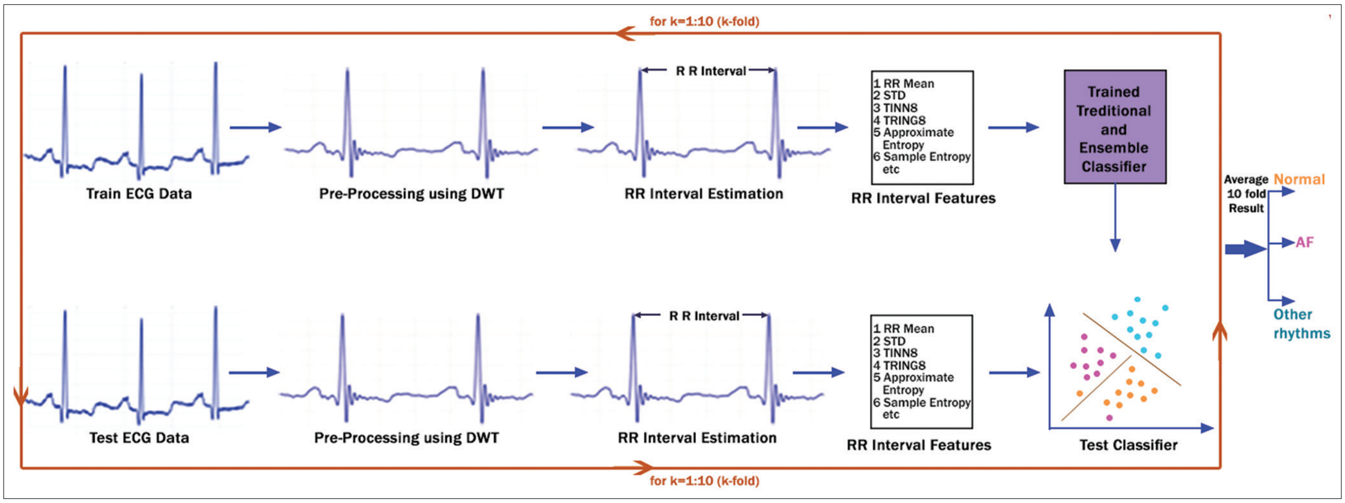

Figure 1: Graphical abstract of the proposed approach

Table 1: Literature review

Literature	Techniques	Datasize	Validation	Classifier	Overall accuracy
Luo *et al.*[6]	Poincare plot	229 ECG signal from different database	1 fold	Threshold	Sensitivity of 97.91% and specificity of 99.14%
Radhakrishnan *et al.*[7]	Chirplet transform	34206 ECG signals from 3 database	10 fold	CNN-BiLSTM	99.18%
Shi *et al.*[8]	Canonical correlation	8528 ECG signals	1 fold	CNN	91.7%
Nurmaini *et al.*[9]	DWT	5504 Training signals 824 Testing signals 612 Validation signals	10 fold	1D-CNN	99.98%
Lown *et al.*[10]	Lorenz plot	19749 ECG signals from different database 415 Validation signals	1fold	SVM	Sensitivity 100% specificity 97.6%
Liang *et al.*[11]	Deep features	7877 ECG signals	10 fold and random oversampling	CNN - BiLSTM	85%
Ghosh *et al.*[12]	Fractional norm feature	190 ECG signals 30 testing signals	10 fold	Hierarchical extreme learning	99.40%
Wu *et al.*[13]	Deep features	5546 ECG signals	10 fold	RF	F1 score of 96%
Wang *et al.*[14]	Wavelet packet transform	141556 ECG segments	10 fold	Artificial neural network	98.8%
Jin *et al.*[15]	Residual block	92789 ECG signals from different data base	1 fold	CNN	98.84%
Yue and Jinjing[16]	Empirical mode decomposition	8528 ECG signals	5 fold	Gradient boosting classifier	86%
Horoba *et al.*[17]	RR interval	25 ECG signals of 10 hour	10 fold	SVM	94%
Kong *et al.*[18]	RR interval	1960 ECG signals	10 fold	SVM	98.16%
Guo *et al.*[19]	Spectrogram	8528ECG signals	1 fold	CNN	78%
Aligholipour *et al.*[20]	Morphological features	100 ECG signals	4 fold	Neural network	93.05%
Firoozabadi *et al.*[21]	Self organizing map	8809 ECG signals from different database	5 fold	Decision tree	F1 score of 96%
Maji *et al.*[22]	Empirical mode decomposition	165 ECG signals	5 fold		Sensitivity of 96.14 % and specificity of 93.51%.

ECG – Electrocardiogram; DWT – Discrete wavelet transform; CNN – Convolutional neural network; SVM – Support vector machine; RF – Random forest

method to detect AF by extracting deep features and RR interval features and recorded accuracy of 91.7%. Nurmaini *et al.*[9] proposed discrete wavelet transform (DWT) along with 13 layers of 1 dimensional CNN to discriminate ECG signal and recoded F1 score of 99.98%. Lown *et al.*[10] developed AF detection algorithm by using Lorenz plot of

RR intervals. They have also used wavelet transformation to compress the image of ECG signal and is given as input to Support Vector Machine (SVM) classifier and recorded sensitivity and specificity of 100% and 97.6%. Liang *et al*.[11] suggested the combination of CNN and BiLSTM classifier extracting deep features to classify heartbeat events and observed accuracy of 85%. Ghosh *et al*.[12] suggested hierarchical extreme ML approach for the diagnosis of AF using multirate cosine bank filter combined with fractional norm feature and observed accuracy of 99.40%. In a work by Shankar *et al*.,[23] various dimensionality reduction techniques, features used and different ML methods are systematically reviewed. Kleyko *et al*.[24] conducted a study on computational complexity of automatic detection of AF and classification using different databases. Wu *et al*.[13] proposed deep features-based approach to diagnose AF using Random Forest (RF) classifier and recorded F1 score of 96%. Wang *et al*.[14] suggested wavelet packet transform method for efficient feature extraction and detection of AF. The author used artificial neural network for the classification purpose and recorded accuracy of 98.8%. Jin *et al*.[15] proposed CNN method combined with residual block method to detect AF and recorded accuracy of 98.84%. Yue *et al*.[16] designed ensemble empirical mode decomposition filter and extreme gradient boosting classifier to identify AF signal and observed accuracy of 86%. Horobo *et al*.[17] proposed an approach based on RR interval features using Lagrangian SVM classifier to detect AF and reported accuracy of 94%. Kong *et al*.[18] proposed RR interval-based method for the diagnosis of AF using relevance vector machine algorithm and observed accuracy of 98.16%. Guo *et al*.[19] suggested a algorithm that incorporates spectrogram and CNN for the diagnosis of AF with accuracy of 78%. Aligholipour *et al*.[20] developed neural network-based algorithm using nonlinear method-based features to identify AF and recorded accuracy of 93.05%. Hagiwara *et al*.[25] reviewed different computer-aided diagnosis methods consisting of ML algorithms and DL algorithms to detect AF developed by various researchers. Firoozabadi *et al*.[21] suggested self-organizing map technique for atrial and ventricular activity features to diagnose AF using decision tree to record F1 score of 96%. Maji *et al*.[22] proposed a methodology to detect AF using the empirical mode decomposition method with sensitivity of 96% and specificity of 93%.

With close observation of existing literature on AF detection, made by authors using Traditional ML (TML) methods or DL methods, yet an exhaustive experimentation and corresponding broader comparative analysis with respect to effective performance does not exist. Different authors have designed experiments with different dataset, varied data size, different split of training, and test data during cross validation, but the comparison of F1score of these experiments is not logical.

The contribution of the present study is the usage of ensemble learning method and verification of experimental proof that ensemble method performs better with high accuracy. The present study involves the extraction of RR series features from ECG signals from the Physionet Challenge 2017 database and training of TML classifiers such as classification and regression tree (CART), K-Nearest Neighbor (KNN), C4.5, Iterative Dichotomiser (ID3), and SVM and ensemble ML (EML) classifiers such as RF, and Rotation Forest and reporting of the obtained performance in the classification of ECG signal into normal, AF, and other rhythms. The proposed approach is illustrated in section II, the results obtained and their significance are discussed in section III, and the paper concludes with section IV.

Methodology

The proposed approach is shown in Figure 2. It involves the removal of baseline wander and other noise components by using DWT, R peak detection using Pan Tompkins algorithm, RR series feature extraction and features selection and finally classification into normal, AF, and other rhythms. Each of the used method in the proposed approach is explained in this section.

Data set used

In this study, open source dataset, Physionet Challenge 2017, is used. The dataset used in this study comprises of 5154 signals with normal rhythm (N), 771 signals with AF rhythm, and 2557 signals with other rhythms. Each of the ECG signals has varied length between 9 s and slightly more than 1 min.[26]

Preprocessing

ECG signals include distinct noises such as baseline wander, muscle and movement artifacts, interference

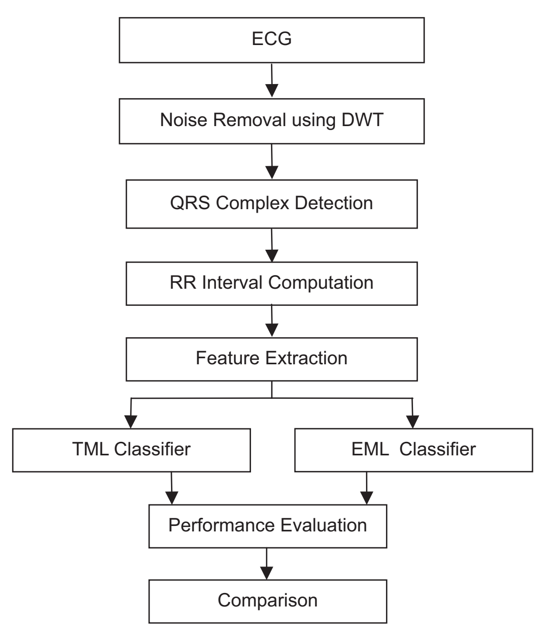

Figure 2: Proposed approach

of power line, electrode pop noise, etc., DWT is used to remove the noises present in the ECG signal. Daubechies6 (db6) is used as basis function in the time frequency decomposition of ECG signal using DWT.[27] The ECG signal sampled at 300 Hz is decomposed into eight levels. The frequency band from 75 Hz to 150 Hz does not include the required components of the ECG signal. This sub-band represents the Ist level detail. Hence, this sub-band is not considered during the reconstruction step. The frequency component between 0 and 0.5 Hz represents baseline wander and is the eighth level approximation sub-band which is also not required. The necessary bands include IInd, IIIrd, IVth, Vth, VIth, and VIIth level's detail sub-bands. Hence, first level detail coefficients and VIIIth level approximation coefficients are replaced with zeros during the reconstruction process of inverse wavelet transformation to obtain the denoised ECG signal. The denoised ECG signal is further subjected to QRS complex detection with the help of Pan Tompkins algorithm[27,28] to detect RR intervals.

Feature extraction

Various features are extracted which are enumerated[25,29-32] in Table 2.

Feature selection

The selection of appropriate features plays an important role in enhancing the accuracy of classification and also reducing the computational time. In this study, we have used the Analysis of Variance (ANOVA) test to select features. The value of *F* and p obtained from ANOVA test is enumerated in Table 2. The relative importance of each feature is expressed in terms of rank.

Classification

In order to discriminate the ECG signal into 3 rhythms viz: normal, AF and other rhythms, TML classifiers such as CART, KNN, C4.5, ID3, and SVM and EML classifiers such as RF, and Rotation Forest are used.

Classification and regression tree

CART is a nonparametric supervised learning method used for classification. CART makes use of *If* and *Else* rules to arrive at a particular condition. It is treelike structure and has three nodes, namely root node, child node, and leaf node. Information gain is calculated at each node. This information gain is used for the further division of the tree. This division continues till leaf node, where the information gain becomes zero.[27]

K-Nearest Neighbor

K-NN is a nonparametric and supervised ML algorithm. It works on the concept of similarity measures. It associates the vector of given feature with the category that is more similar to the available categories. The similarity can be calculated using distance measurements. In the present study, we have used Euclidean distance which is given by.[28]

$$d\left(x_i, x_j\right) = \left(\sum_{i=1}^{n}(x_i - x_j)^2\right)^{1/2} \tag{1}$$

Support vector machine

In the present study, the SVM classifier is used to classify heart rate variability derived features into normal, AF, and other rhythms. SVM classifier works on the principle of constructing a hyperplane which linearly separates two classes.[33,34] A hyperplane can be constructed using

$$g\left(x\right) = w^T x + w_0 = 0 \tag{2}$$

In our study, we have used Radial Basis Kernel function which is given by

$$\Phi(x_i, x_j) = e^{-r\|x_i - x_j\|2} \tag{3}$$

Where (x_i, y_i) represent training set, $x_i \in$ Rd, i = 1,2,….l and $y_i \in (-1,1)$

Rotation forest

Rotation forest is a decision tree-based ensemble classifier. The rotation forest increases the accuracy of the individual tree, yet maintains the diversity within the ensemble. The RF algorithm is sensitive to the rotation of feature axes.[35] The final average performance of rotation forest is the average performance of all individual trees. The rotation forest can achieve better performance with less number of trees. It is more immune to noise present in the data. In the present study, the rotation forest is used to classify HRV derived features into normal, AF, and other rhythms. As the rotation forest is robust to noise and as it provides a simple tree structure with minimum number of parameters, it is expected to provide relatively higher performance.[36]

C4.5

C4.5 is a decision tree-based algorithm and is an extension of Quinlan's ID3 algorithm used for classification of data. C4.5 constructs decision tree from the set of training data using the concept of information entropy. C4.5 selects the attribute of the data at each tree node that most efficiently splits its sample set into subsets enriched in one or the other class.[37] The splitting criteria are based on the difference in entropy of data. To make the decision, the attribute with the highest difference in entropy is selected. The C4.5 algorithm then splits by selected attribute to produce remaining on the sub-list.

Iterative dichotomiser 3

ID3 algorithm is used in ML and natural language processing. It is used to generate the decision tree from the given dataset. ID3 algorithm begins with the original set S as the root node. At every iteration of the algorithm, it iterates through every unused attribute of the data set

Table 2: Time domain features extracted from electrocardiogram signal

Feature	Computational equation	*F*	*P*	Rank
RRμ	Mean of RR intervals RRμ= $\frac{1}{N}\sum_{i=1}^{N} RR_i$ Where RR_i is the RR interval at i^{th} instant and N is the length of RR interval	8.02	0.0003	9
CVRR	Coefficient variance of RR intervals CVRR= $\frac{1}{N}\sum_{i=1}^{N}(RR_i - RR_\mu)^2$	1.29	0.2753	15
SDRR	Standard deviation of RR intervals SDRR= $\sqrt{CVRR}$	3.79	0.227	13
SDSD	Standard deviation of RR interval differences SDSD=STD (RR_i-RR_{i+1}) Where STD is standard deviation	5.38	0.0046	10
RMSSD	RMS of successive differences RMSSD= $\sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}((RR)_{i+1} - (RR)_i)^2}$	2.93	0.0533	14
RR50	Number of pairs of adjacent RR intervals differing by 50 ms RR50=(RR_i-RR_{i+1})>0.05	3.92	0.0198	11
pRR50	Proportion successive RR interval >50 ms pRR50=100*RR50/Length of RR intervals	9.12	0.0001	7
Triang8	It is the total number of RR intervals divided by the height of histogram in 8 m bins	13.77	1.07e-04	5
TINN8	Multilinear function is q is defined such that q (t)=0 for t <A and t >B and q (X)=Y TINN8=B-A where 8 stands for 8 ms bins	17.72	2.09e-08	3
pRR20	Proportion successive RR interval >20 ms pRR20=100*(RR_i-RR_{i+1}) >0.02/(Length of RR intervals)	21.21	6.47e-10	1
pRR30	Proportion successive RR interval>30 ms pRR30=100*(RR_i-RR_{i+1}) >0.03/(Length of RR intervals)	15.08	2.88e-07	4
pRR6.25	Proportion successive difference >1/16 ms pRR6.25=100*(RR_i-RR_{i+1}) >0.00625/(Length of RR intervals)	17.95	1.66e-08	2
RSA 5RR	Difference betweenmean of 5 largest and5 smallest RR intervals	9.9	5.09e-05	6
SampEn	Sample Entropy SampEn (m, r, N)= $-In\left[\frac{D}{E}\right]$ where D=Number of guided vector pairs having distance function d[X_{m+1}(i), X_{m+1}(j)]<r E=Number of guided vector pairs having distance function d[X_m (i), X_m (j)]<r	2.8	0.00061	12
ApEn	Approximate entropy ApEn= $\left\|\theta^m(r) - \theta^{m+1}(r)\right\|$ $\theta^m(r) = \left(\frac{1}{N-m+1}\right)\sum_{i=1}^{N-m+1} logB_r^m(i)$ where B_r^m (i)=[number of x (j) such that d (x(i)), x (j)<r/(N-m+1)	8.16	0.0003	8

S.[38] After iteration, it calculates entropy of the attribute. Algorithm then selects the attribute which has smallest entropy value. The set S is then split by selected attribute to produce sublist of the data.

Results and Discussion

In this study, a three class pattern classification problem of detection of normal, AF, and other rhythms using the ECG derived RR interval features is proposed. The methodology is implemented in MATLAB using Physionet Challenge 2017 dataset. Baseline wander and high frequency noise present in the ECG are removed using DWT. Pan-Tompkins algorithm is used to detect QRS complex from denoised ECG. Figure 3 (a) shows raw ECG consisting of baseline wander. Figure 3 (b) shows denoised ECG using DWT where the baseline wander is removed. Figure 3 (c) provides the visualization of detected QRS complex which is marked in red asterisk from the denoised ECG. Various features, extracted from the RR interval time series, which are listed in Table 2. It can be seen that each feature has a different *F* and p indicating different discrimination potential and statistical significance, respectively, during ANOVA test. In total, the extracted 15 features are subjected to classification using TML algorithms as well as EML algorithms.

The average 10-fold cross validation results and area under the curve (AUC) for CART, KNN, RF, Rotation Forest, C4.5, ID3, and SVM classifier are shown in Table 3. It can be observed that RF classifier achieved an overall accuracy of 99.10%.

From Table 3, it could be concluded that the RF classifier provides higher performance metrics (NCSA of 98.64%, AFCSA of 99.50%, OCSA of 97.98%, and OA of 99.10%) in comparison with other classifiers. During the training and the testing of the classifier, the 10 fold cross validation is used. It can be seen that RF classifier provides the highest overall accuracy with a minimum variation in respect to different folds. From the classified signals receiver operating characteristic curve is plotted to calculate AUC. It is observed that AUC of RF classifier is 0.998 which is more than the other classifiers.

The scatter plot is presented in Figure 4 for the first 2 features of 100 signals. The various distinct regions resulted due to classifier training can be seen in the scatter plot.

In the present study, the RF classifier provided higher performance in comparison with other EML and TML classifiers. It is interesting to compare the performance obtained in the present study with the existing state-of-the-art methods available in the literature. Table 4 highlights some of the work carried out by authors on detecting AF using Physionet challenge 2017 dataset. Rao *et al*.[27] proposed effective noise removal method along with dimensionality reduction techniques using CNN to detect AF and recorded accuracy of 91.71%. Quang *et al*.[33] suggested a method based on statistical features of segments derived from CNN and recorded F1 score of 84.19%. Plesinger *et al*.[39] proposed a method to classify ECG signal using PQRS morphology features using Bagged tree ensemble and shallow NN and achieved F1 score of 91%. Liu *et al*.[40] suggested a method using SVM classifier for discrimination of ECG signal using *P* wave, RR interval, spectrum, entropy features, and recorded average F1 score of 90.82%. Chen *et al*.[41] proposed morphological and heart rate variable features method to detect AF. Piecewise linear splines method was used for the selection of features and recorded F1 score of 81%. Mei *et al*.[42] proposed heart rate variability and frequency features for the diagnosis of AF using SVM classifier and achieved accuracy of 96%. Shao *et al*.[43] proposed delta RR interval, morphology, similarity index feature to detect AF using AdaBoost classifier and recorded F1 score of 82%. Sanchez *et al*.[44] suggested gramian angular summation field method to detect AF using CNN and achieved accuracy of 97.6%. Najmeh *et al*.[45] suggested neural architecture search method to

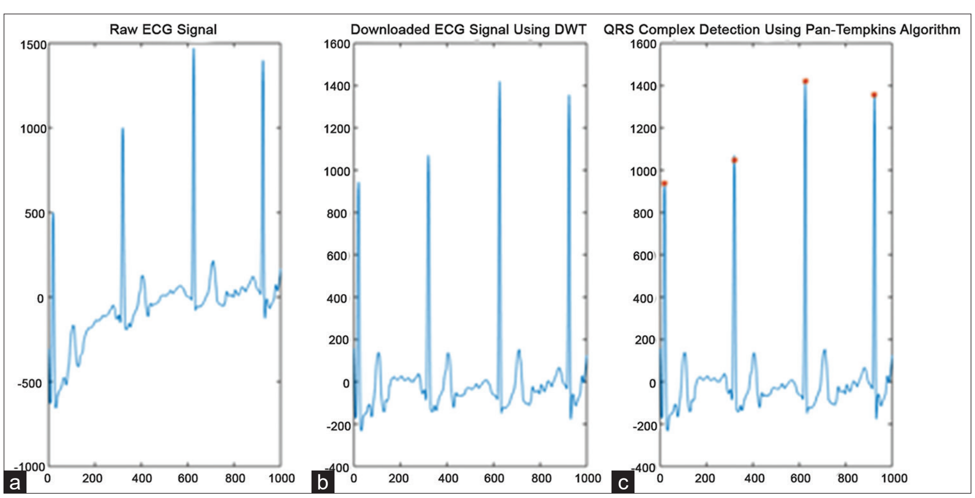

Figure 3: (a): Raw ECG. (b) Denoised ECG using DWT. (c): QRS complex detection. ECG: Electrocardiogram, DWT: Discrete wavelet transform

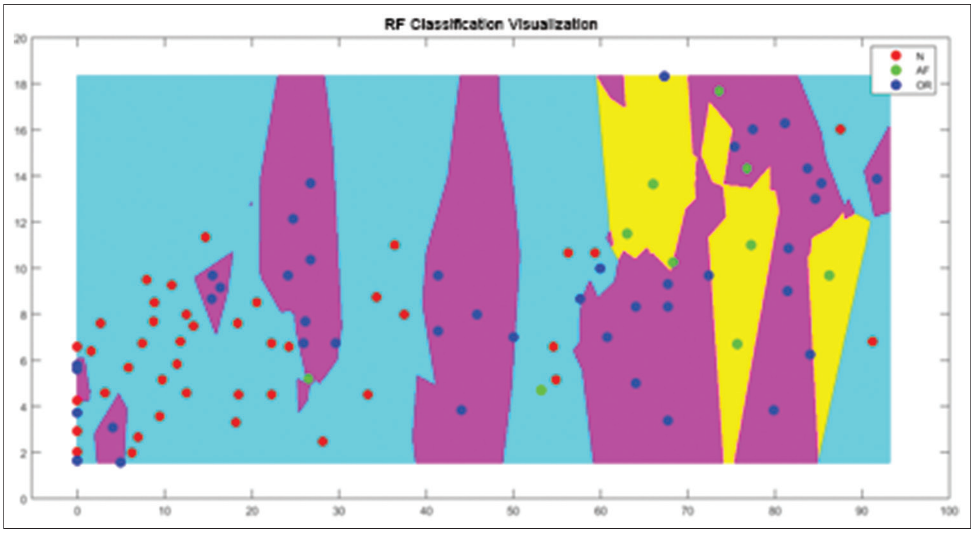

Figure 4: Scatter plot of first two features with class boundary during RF classification. RF: Random forest

Table 3: Average classification performance of classification and regression tree, K-nearest neighbor, random forest, Rotation Forest, C4.5, iterative dichotomiser, and support vector machine classifier

Classifier	NCSA±SD	AFCSA±SD	OCSA±SD	OA±SD	AUC±SD
CART	98.64±0.44	97.86±0.71	97.94±1.24	98.34±0.43	0.972±0.0066
KNN	98.24±0.89	92.67±3.66	93.30±2.99	96.16±1.75	0.968±0.0060
RF	99.65±0.18	99.50±0.18	97.98±0.47	99.10±0.25	0.998±0.0006
Rotation Forest	99.27±0.24	97.81±0.78	97.29±0.69	98.50±0.38	0.976±0.0059
C4.5	99.19±0.18	98.35±0.60	97.20±0.28	98.47±0.23	0.979±0.0037
ID3	98.91±0.09	98.31±1.03	96.98±1.28	98.24±0.49	0.989±0.0042
SVM	96.18±0.32	97.26±0.38	95.56±0.64	96.16±0.54	0.956±0.0032

NCSA – Normal class specific accuracy; AFCSA – AF class specific accuracy; OCSA – Other rhythm class specific accuracy; OA – Overall accuracy; AUC – Area under the curve; CART – Classification and regression tree; KNN – K-nearest neighbor; RF – Random forest; ID3 – Iterative dichotomiser; SVM – Support vector machine; SD – Standard deviation

Table 4: Overview of studies on classification of electrocardiogram signal using physionet challenge 2017 database

Literature	Techniques	Classifier	Overall accuracy
Shrikanth *et al*.[27]	PCA and ICA	CNN	91.71%
Nguyen *et al*.[33]	Stacking method	SVM and CNN	F1 score of 84.19%
Plesinger *et al*.[39]	PQRS morphology features	Bagged tree ensemble and shallow neural network	F1 score of 91%
Liu *et al*.[40]	P wave, RR interval, spectrum, entropy features	SVM	F1 score of 90.82%
Chen *et al*.[41]	Morphological and heart rate variability features	XGBoost	F1 score of 81%
Mei *et al*.[42]	Heart rate variability and frequency features	SVM	96%
Shao *et al*.[43]	Delta RR interval, morphology, similarity index features	AdaBoost classifier	F1 score of 81%
Sánchez *et al*.[44]	Gramian angular summation fields	CNN	97.6%
Fayyazifar *et al*.[45]	Neural architecture search	CNN	F1 score of 84.15%
Rao *et al*.[46]	Power spectrum	CNN	94.67%
Current Study	RR series feature	RF	99.10%

CNN – Convolutional neural network; SVM – Support vector machine; RF – Random forest; PCA – Principal component analysis; ICA – Independent component analysis

diagnose AF using CNN classifier and observed F1 score of 84.15%. Rao *et al*.[46] proposed power spectrum-based method for the classification of ECG signal using CNN and reported 94.67% of accuracy.

The current study focuses on extracting RR interval based features and their classification using TML and EML classifiers. The 15 features are extracted from RR interval series are subjected to classification using CART, KNN, RF, Rotation Forest, C4.5, ID3, and SVM classifier. During the study, it is found that RF performed better compared to other classifiers achieving an overall accuracy of 99.10%.

It is noted that there are many studies in the literature which classify the ECG into different classes (as seen in Table 4). Innovation in this proposed study is the usage of ensemble learning method and its superiority in terms of high classification performance in accurately categorizing the ECG data into normal, AF and other rhythms. Apart from the methodological contribution, the method also provides improved performance, which is the key contribution of this study. Referring Table 4, all other authors in previous studies obtained relatively lower performance compared to the proposed study.

The proposed study has the application in monitoring of patients in hospitals, critical care units, Holter monitoring of patients during their normal activity, pacemakers, and defibrillators. The developed system can be used in mass screening of population, remote patient monitoring, and in many other applications. Automated computer diagnosis of AF can bring a change in the conventional practice by immensely helping the physicians in accurately diagnosing and mobilizing patients.

Conclusion

In this work, we propose RR interval features-based methodology to detect AF. Physionet challenge 2017 database is used in this study. DWT method is used for the denoising of ECG signal. TML algorithms and EML algorithms are used to evaluate the performance. It is observed that RF classifier gave a better discrimination with NCSA of 99.65%, AFCSA of 99.50%, OCSA of 97.98%, and OA of 99.10%.

There is an increase in the number of patients with heart arrhythmia, worldwide. Manually diagnosing each arrhythmia and its classification into different categories is extremely tedious and time consuming for any physician. An accurate method of classification of ECG signal into various rhythms is very important. Each abnormality requires different method of treatment and any misclassification may cause serious complications. In this direction, obtaining high performance is highly desirable. The proposed methodology provided an improved accuracy on the benchmark dataset. The developed methodology can be used in the health care management system to screen ECG signal into various rhythms.

Financial support and sponsorship

None.

Conflicts of interest

There are no conflicts of interest.

References

1. Johansson C, Dahlqvist E, Andersson J, Jansson JH, Johansson L. Incidence, type of atrial fibrillation and risk factors for stroke: A population-based cohort study. Clin Epidemiol 2017;9:53-62.
2. Lippi G, Sanchis-Gomar F, Cervellin G. Global epidemiology of atrial fibrillation: An increasing epidemic and public health challenge. Int J Stroke 2021;16:217-21.
3. Chung MK, Eckhardt LL, Chen LY, Ahmed HM, Gopinathannair R, Joglar JA, *et al*. Lifestyle and risk factor modification for reduction of atrial fibrillation: A scientific statement from the American Heart Association. Circulation 2020;141:e750-72.
4. Tsao CW, Aday AW, Almarzooq ZI, Alonso A, Beaton AZ, Bittencourt MS, *et al*. Heart disease and stroke statistics – 2022 update: A report from the American Heart Association. Circulation 2022;145:e153-639.
5. Tang DH, Gilligan AM, Romero K. Economic burden and disparities in health care resource use among adult patients with cardiac arrhythmia. Appl Health Econ Health Policy 2014;12:59-71.
6. Luo C, Li Q, Rao H, Huang X, Jiang H, Rao N. An improved Poincaré plot-based method to detect atrial fibrillation from short single-lead ECG. Biomed Signal Proc Control 2021;64:102264.
7. Radhakrishnan T, Karhade J, Ghosh SK, Muduli PR, Tripathy RK, Acharya UR. AFCNNet: Automated detection of AF using chirplet transform and deep convolutional bidirectional long short term memory network with ECG signals. Comput Biol Med 2021;137:104783.
8. Shi Jingjing, Chao Chen, Hui Liu, Yinglong Wang, Minglei Shu, and Qing Zhu. Automated atrial fibrillation detection based on feature fusion using discriminant canonical correlation analysis. Computational and Mathematical Methods in Medicine 2021;2021.6691177. https://doi.org/10.1155/2021/6691177.
9. Nurmaini S, Tondas AE, Darmawahyuni A, Rachmatullah MN, Partan RU, Firdaus F, *et al*. Robust detection of atrial fibrillation from short-term electrocardiogram using convolutional neural networks. Future Gener Comput Syst 2020;113:304-17.
10. Lown M, Brown M, Brown C, Yue AM, Shah BN, Corbett SJ, *et al*. Machine learning detection of Atrial Fibrillation using wearable technology. PLoS One 2020;15:e0227401.
11. Liang Y, Yin S, Tang Q, Zheng Z, Elgendi M, Chen Z. Deep learning algorithm classifies heartbeat events based on electrocardiogram signals. Front Physiol 2020;11:569050.
12. Ghosh SK, Tripathy RK, Paternina MR, Arrieta JJ, Zamora-Mendez A, Naik GR. Detection of atrial fibrillation from single lead ECG signal using multirate cosine filter bank and deep neural network. J Med Syst 2020;44:1-15.
13. Wu X, Zheng Y, Chu CH, He Z. Extracting deep features from short ECG signals for early atrial fibrillation detection. Artif Intellig Med 2020;109:101896.
14. Wang J, Wang P, Wang S. Automated detection of atrial fibrillation in ECG signals based on wavelet packet transform and correlation function of random process. Biomed Signal Proc Control 2020;55:101662.
15. Jin Y, Qin C, Liu J, Lin K, Shi H, Huang Y, Liu C. A novel domain adaptive residual network for automatic atrial fibrillation detection. Knowl Based Syst 2020;203:106122.
16. Yue Z, Jinjing Z. Atrial fibrillation detection based on EEMD and XGBoost. J Phys Conf Ser 2019;1229:012074.
17. Horoba K, Czabanski R, Wrobel J, Matonia A, Martinek R, Kupka T, *et al*. Recognition of atrial fibrilation episodes in heart rate variability signals using a machine learning approach. In: 2019 MIXDES-26th International Conference" Mixed Design of Integrated Circuits and Systems. Rzeszow, Poland: IEEE; 2019. p. 419-24. Available from: https://doi. 10.23919/MIXDES0.2019.8787048. [Last accessed on 2019 Aug 05].
18. Kong D, Zhu J, Wu S, Duan C, Lu L, Chen D. A novel IRBF-RVM model for diagnosis of atrial fibrillation. Comput Methods Prog Biomed 2019;177:183-92.
19. Guo JM, Yang CC, Wang ZH, Hsia CH, Chang LY. Atrial fibrillation detection in spectrogram based on convolution neural

networks. In: 2019 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS). Taipei, Taiwan: IEEE; 2019. p. 1-2. Available from: https://doi. 10.1109/ISPACS48206.2019.8986347. [Last accessed on 2020 Feb 10].

20. Aligholipour O, Kuntalp M, Sadaghiyanfam S. Silent paroxysmal atrial fibrillation detection by neural networks based on ECG records. In: 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT). Istanbul, Turkey: IEEE; 2019. p. 1-4. Avaialble from: https://doi. 10.1109/EBBT0.2019.8741771. [Last accessed on 2019 Jun 20].
21. Firoozabadi R, Gregg RE, Babaeizadeh S. P-wave analysis in atrial fibrillation detection using a neural network clustering algorithm. In: 2018 Computing in Cardiology Conference (CinC). Vol. 45. Maastricht, Netherlands: IEEE; 2018. p. 1-4.
22. Maji U, Mitra M, Pal S. Automatic detection of atrial fibrillation using empirical mode decomposition and statistical approach. Proc Technol 2013;10:45-52.
23. Shankar MG, Babu CG. An exploration of ECG signal feature selection and classification using machine learning techniques. Int J Innov Technol Explor Eng Regul 2020;9:797-804.
24. Kleyko D, Osipov E, Wiklund U. A comprehensive study of complexity and performance of automatic detection of atrial fibrillation: Classification of long ECG recordings based on the physionet computing in cardiology challenge 2017. Biomed Phys Eng Exp 2020;6:025010.
25. Hagiwara Y, Fujita H, Oh SL, Tan JH, Tan RS, Ciaccio EJ, *et al*. Computer-aided diagnosis of atrial fibrillation based on ECG signals: A review. Inf Sci 2018;467:99-114.
26. Clifford GD, Liu C, Moody B, Lehman LH, Silva I, Li Q, *et al.* AF classification from a short single lead ECG recording: The physionet/computing in cardiology challenge 2017. Comput Cardiol (2010) 2017;44:1-4.
27. Shrikanth RS, Krithika K, Akhila M, Martis RY. Deep learning based atrial fibrillation detection using effective denoising methods and dimensionality reduction techniques. In: 2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC). Bangalore, India: IEEE; 2021. p. 01-7. Available form: https:// doi. 10.1109/R10-HTC53172.2021.9641550. [Last accessed on 2021 Dec 22].
28. Martis RJ, Acharya UR, Adeli H. Current methods in electrocardiogram characterization. Comput Biol Med 2014;48:133-49.
29. Rizwan M, Whitaker BM, Anderson DV. AF detection from ECG recordings using feature selection, sparse coding, and ensemble learning. Physiol Meas 2018;39:124007.
30. Mali B, Zulj S, Magjarevic R, Miklavcic D, Jarm T. Matlab-based tool for ECG and HRV analysis. Biomed Signal Proc Control 2014;10:108-16.
31. Smith AL, Owen H, Reynolds KJ. Heart rate variability indices for very short-term (30 beat) analysis. Part 1: Survey and toolbox. J Clin Monit Comput 2013;27:569-76.
32. Smith AL, Owen H, Reynolds KJ. Heart rate variability indices for very short-term (30 beat) analysis. Part 2: Validation. J Clin Monit Comput 2013;27:577-85.
33. Nguyen QH, Nguyen BP, Nguyen TB, Do TT, Mbinta JF, Simpson CR. Stacking segment-based CNN with SVM for recognition of atrial fibrillation from single-lead ECG recordings. Biomed Signal Proc Control 2021;68:102672.
34. Qayyum A, Meriaudeau F, Chan GC. Classification of atrial fibrillation with pre-trained convolutional neural network models. In: 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES). Sarawak, Malaysia: IEEE; 2018. p. 594-9. Available from: https://doi. 10.1109/IECBES0.2018.8626624. [Last accessed on 2019 Jan 28].
35. Ozcift A, Gulten A. Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms. Comput Methods Programs Biomed 2011;104:443-51.
36. Rodríguez JJ, Kuncheva LI, Alonso CJ. Rotation forest: A new classifier ensemble method. IEEE Trans Pattern Anal Mach Intell 2006;28:1619-30.
37. Mohanty M, Sahoo S, Biswal P, Sabut S. Efficient classification of ventricular arrhythmias using feature selection and C4. 5 classifier. Biomed Signal Proc Control 2018;44:200-8.
38. Yang S, Guo JZ, Jin JW. An improved Id3 algorithm for medical data classification. Comput Electr Eng 2018;65:474-87.
39. Plesinger F, Nejedly P, Viscor I, Halamek J, Jurak P. Parallel use of a convolutional neural network and bagged tree ensemble for the classification of Holter ECG. Physiol Meas 2018;39:094002.
40. Liu N, Sun M, Wang L, Zhou W, Dang H, Zhou X. A support vector machine approach for AF classification from a short single-lead ECG recording. Physiol Meas 2018;39:064004.
41. Chen Y, Wang X, Jung Y, Abedi V, Zand R, Bikak M, Adibuzzaman M. Classification of short single-lead electrocardiograms (ECGs) for atrial fibrillation detection using piecewise linear spline and XGBoost. Physiol Meas 2018;39:104006.
42. Mei Z, Gu X, Chen H, Chen W. Automatic atrial fibrillation detection based on heart rate variability and spectral features. IEEE Access 2018;6:53566-75.
43. Shao M, Bin G, Wu S, Bin G, Huang J, Zhou Z. Detection of atrial fibrillation from ECG recordings using decision tree ensemble with multi-level features. Physiol Meas 2018;39:094008.
44. Sánchez FA, Cervera JA. ECG classification using artificial neural networks. J Phys Conf Ser 2019;1221:012062.
45. Fayyazifar N. An accurate CNN architecture for atrial fibrillation detection using neural architecture search. In: 2020 28th European Signal Processing Conference (EUSIPCO). Amsterdam, Netherlands: IEEE; 2021. p. 1135-9. Avaialble from: https://doi. 0.23919/Eusipco47968.2020.9287496. [Last accessed on 2020 Dec 2020].
46. Rao SK, Kolekar MH, Martis RJ. Frequency domain features based atrial fibrillation detection using machine learning and deep learning approach. In: 2021 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT). Bangalore, India: IEEE; 2021. p. 1-6. Available form: https://doi: 10.1109/CONECCT52877.2021.9622533. [Last accessed on 2021 Dec 07].