ORIGINAL ARTICLE

Deep Learning Based Atrial Fibrillation Detection Using Combination of Dimensionality Reduction Techniques and RR Interval Features

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

[1] *Department of Electronics & Communication Engineering, Vivekananda College of Engineering & Technology, Puttur, Dakshina Karnataka, India*

[2] *Department of Computer Science & Engineering, Global Academy of Technology, Bangalore, Karnataka State, India*

*Corresponding Author: Shrikanth Rao S.K
Email: skanthrao@gmail.com

Received: 15 September 2022 / Accepted: 19 June 2023

Abstract

Purpose: Accurate detection of Atrial Fibrillation (AF) has great significance in the field of medical science which can reduce the rate of mortality and morbidity. The present study focuses on Electrocardiography (ECG) signal classification using dimensionality reduction techniques combined with R wave to R wave interval (RR interval) features.

Materials and Methods: In the first approach, Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA), and Probabilistic Principal Component Analysis (PPCA) are performed independently on denoised ECG signal using Discrete Wavelet Transform (DWT) for the classification of ECG signal. In the second approach, the dimensionality reduction techniques combined with RR interval features are used for the classification of ECG signal.

Results: Machine Learning (ML) algorithms such as Decision Tree (DT), Support Vector Machine (SVM), and Deep Learning (DL) algorithms such as Long Short Term Memory (LSTM) and Bi-Directional LSTM (BiLSTM) are used for classification purposes.

Conclusion: The proposed methodology provided an overall accuracy of 93.65% with PCA and LSTM classifier and an overall accuracy of 99.45% with PCA combined with RR interval features and LSTM classifier. The developed technology has potential applications in many practical solutions.

Keywords: Atrial Fibrillation; Electrocardiography; Discrete Wavelet Transform; Long Short Term Memory; Support Vector Machine; Decision Tree.

DOI: https://doi.org/10.18502/fbt.v11i4.16504

1. Introduction

Cardiovascular Diseases (CVDs) are considered one of the major health concerns that pose an immense economic burden globally. Arrhythmia is a condition where heartbeat is irregular. Atrial Fibrillation (AF) is a type of arrhythmia that is closely related to CVD. The prevalence of AF is expected to be 12.1 million by 2030 in the US [1]. According to American Heart Association statistics, the number of citizens with newly diagnosed AF is projected to be 2.6 million by 2030 in the US [1]. AF affects more than 40 million individuals globally [2]. The increasing rate of AF leads to a financial burden for developing countries. From a health care point of view, it is necessary to develop a cost-effective method for the detection of AF using Electrocardiography (ECG) signals.

ECG measures the depolarization and repolarization activity of the heart which is represented by the P-QRS-T wave. During AF there are minute changes in the depolarization and repolarization patterns. This results in the absence of P waves and variation of RR intervals in the ECG. Discrimination of ECG signal for different rhythms is computationally implemented using classification techniques. Several Machine Learning (ML) and Deep Learning (DL) methods have been proposed in the literature for the discrimination of ECG signals. Shi *et al.* [3] suggested Support Vector Machine (SVM) based method to detect AF by observing variations in face skin color and recorded an accuracy of 92.56%. Petmezas *et al.* [4] suggested a focal loss method to address the data imbalance problem and used a Convolutional Neural Network (CNN) - Long Short Term Memory (LSTM) classifier and recorded a sensitivity of 97.87 % and specificity of 99.29%. Radhakrishnan *et al.* [5] proposed a chirplet transform method using a deep convolutional Bi-Directional LSTM (BiLSTM) network and achieved an accuracy of 99.18%. Liang *et al.* [6] suggested a combination of CNN-BiLSTM classifier extracting deep features to classify heartbeat events and observed an accuracy of 85%. Ghosh *et al.* [7] suggested a DL approach to detect AF using a multirate cosine bank filter combined with a fractional norm feature and observed an accuracy of 99.40%. Ahmed *et al.* [8] developed a low-cost AF detection machine using variation in RR intervals. The developed device was able to measure variation in heartbeat and notify when AF detected. In a work by Shankar *et al.* [9], various dimensionality reduction techniques, features used and different ML methods are systematically reviewed. Kleyko *et al.* [10] conducted a study on the computational complexity of automatic detection of AF and classification using different databases. Wu *et al.* [11] proposed a deep features-based method to detect AF using a random forest classifier and recorded an F1 score of 96%. Wang *et al.* [12] suggested a wavelet packet transform method for efficient feature extraction and detection of AF. The author used an artificial neural network for the classification purpose and recorded an accuracy of 98.8%. Li *et al.* [13] suggested a method to identify AF signals using heart rate, depth feature, and principal components. Using SE-ResNet architecture they achieved an accuracy of 99.68%. Hagiwara *et al.* [14] reviewed different computer-aided diagnosis methods consisting of ML algorithms and DL algorithms to detect AF developed by various researchers. Faust *et al.* [15] proposed an LSTM network-based method to detect AF and recorded an accuracy of 99.77%. Maji *et al.* [16] proposed a methodology to detect AF using an empirical mode decomposition method with a sensitivity of 96% and specificity of 93%. Dash *et al.* [17] suggested a robust algorithm to detect AF using heart rate variability features and observed sensitivity and specificity of 90.2% and 91.2 %, respectively.

From the existing literature on AF detection, various authors have proposed feature-based methods for pattern classification of ECG using either traditional machine learning algorithms or deep learning algorithms where features are automatically mapped into respective output classes. A few kinds of literature are available where RR interval features along with dimensionally reduced features are used for classification, yet systematic and exhaustive experimental evidence to prove which combination of methods provides a superior performance is lacking in this domain.

In the present study, we propose simple AF detection methods using ECG as follows:

1. Dimensionality reduction methods such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA), and Probabilistic Principal Component Analysis (PPCA) along with the absence of P wave features are subjected to traditional ML and DL algorithms and the method that provided superior performance is identified.

2. A combination of dimension reduction techniques, RR interval-based features, and absence of P wave feature are classified using ML and DL methods and performance is compared.

Two ML algorithms namely DL, and SVM; and two DL algorithms namely LSTM and BiLSTM are used for the discrimination of ECG signals into normal, AF, and other rhythms, and class-specific accuracy is calculated. The proposed methodology is discussed in section 2, and in section 3, results are discussed; the paper is concluded in section 4.

2. Materials and Methods

Figure 1 shows the proposed methodology for pattern classification of ECG signal. The methodology consists of pre-processing using Discrete Wavelet Transform (DWT), and Q wave, R wave and S wave (QRS) complex detection using the Pan-Tompkins algorithm. From the detected QRS complex, 15 time-series features are extracted and ECG signals are segmented such that each segment contains 200 sample points. The 200 sample points constitute 99 points before the R peak, and 100 points after the R peak along with the R peak. These 200 points are used for dimensionality reduction using PCA, LDA, ICA, and PPCA. Also, during AF, a P wave is absent and a fibrillatory wave (f wave) is present which can be manifested in the energy of the ECG signal. In total 15 RR time series features, 10 dimensionally reduced features, and one energy feature constituting a total of 26 features which are given to two ML classifiers independently viz: Decision Tree (DT) and SVM and two DL algorithms separately viz: LSTM and BiLSTM. The performance due to all classifiers is recorded and a comparative study is carried out.

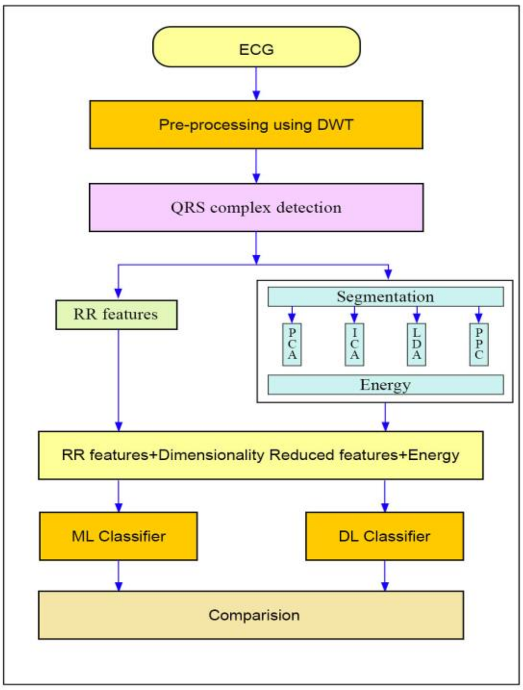

Figure 1. Proposed Methodology

2.1. Dataset Used

In this study open-source dataset, Physionet Challenge 2017 [18] is used. The dataset used in this study comprises 5154 signals with normal rhythm (N), 771 signals with AF rhythm, and 2557 signals with other rhythms. Each of the ECG signals has varied lengths between 9 seconds and more than slightly one minute.

2.2. Pre-processing

The noise components present in the ECG signal (baseline wander and high-frequency noise components) are removed using DWT [2, 19]. In this research, DWT with db6 basis function is used to decompose the ECG in the time-frequency domain [20]. The ECG signals are sampled at 300Hz, and are decomposed into 8 levels using DWT. The frequency band of 75-150Hz contains high-frequency noise components. Hence, these coefficients are replaced with zeros during the reconstruction. Similarly, the frequency components in the band of 0-0.5Hz consisting of baseline wander are replaced with zeros during reconstruction. The denoised ECG signal is further subjected for the detection of QRS complex with the help of the Pan Tompkins algorithm. Using the detected QRS complexes ECG signals are segmented into 200 sample windows for a given ECG beat.

2.3. RR Series Feature Extraction

From the detected QRS complexes the RR interval time series is derived from which RR series features are extracted [16, 21-26] which are shown in Table 1.

2.3.1. PCA

Given the training vector, X={$x_1, x_2, \ldots, x_n$} belonging to C classes, the principal components analysis algorithm requires the definition of covariance matrix [19, 26] as Equation 1:

$$K = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(x_i - \bar{x})^T \quad (1)$$

where $\overline{x}$ = (1 / n) $\sum_{i=1}^{n} x_i$. In the next step, the eigenvectors and eigenvalues of the covariance matrix K are calculated. Let P = $\{\eta_1, \eta_2, ..., \eta_r\}$ be the r < Q Eigenvectors corresponding to the r largest Eigenvalues of K. Eigenfeature vectors E_i are calculated by projecting the Eigenvectors **onto** the Eigenspace (Equation 2):

$$E_i = P^T x_i \quad (2)$$

In total first 10 PCA components are considered for classification purposes.

2.3.2. LDA

LDA maximizes the discrimination among different classes in the data. It projects the data into a new axis in a way to maximize the separation of classes [27, 28].

Accordingly, LDA is applied to the Eigenfeature vectors X_i as follows. Between-class (S_B) and within-

Table 1. RR series features extracted from ECG signal

Feature	Computational Equation
RRµ	Mean of RR intervals: **RRµ** = $\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{RR_i}$ Where **RR$_i$** is the RR interval at **ith** instant and N is the length of RR interval
CVRR	Coefficient of variant of RR intervals: **COVRR** = $\frac{1}{N}\sum_{i=1}^{N} (\boldsymbol{RR_i} - \mathbf{RR\mu})^2$
SDRR	Standard deviation of RR intervals: **SDRR** = $\sqrt{\boldsymbol{CVRR}}$
SDSD	Standard deviation of RR interval differences: **SDSD = STD(RR$_i$ - RR$_{i+1}$)** Where **STD** is standard deviation
RMSSD	RMS of successive differences: **RMSSD** = $\sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}((\boldsymbol{RR})_{i+1} - (\boldsymbol{RR})_i)^2}$
RR50	Number of pairs of adjacent RR intervals differing by 50 ms: **RR50 = (RR$_i$ - RR$_{i+1}$) > 0.05**
pRR50	Proportion of successive RR interval > 50 ms: **pRR50 = 100 * RR50 / (Length of RR intervals)**
Triang8	It is the total number of RR intervals divided by the height of histogram in 8 ms bins [36, 37]
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 [24, 25]** where 8 stands for 8ms bins
pRR20	Proportion of successive RR interval > 20 ms **pRR20 = 100 * (RR$_i$ - RR$_{i+1}$) > 0.02 / (Length of RR intervals)**
pRR30	Proportion of successive RR interval > 30 ms **pRR30 = 100 * (RR$_i$ - RR$_{i+1}$) > 0.03 / (Length of RR intervals)**
pRR6.25	Proportion of successive difference > 1/16ms **pRR6.25 = 100 * (RR$_i$ - RR$_{i+1}$) > 0.00625 / (Length of RR intervals)**
RSA 5RR	Difference between the mean of 5 largest and 5 smallest RR intervals [25]
SampEn	Sample Entropy, **SampEn(m,r,N) = -In[$\frac{D}{E}$]** 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**
ApEn	Approximate entropy: **ApEn** = $\lvert\boldsymbol{\theta^m(r)} - \boldsymbol{\theta^{m+1}(r)}\rvert$ $\boldsymbol{\theta^m(r)} = \left(\frac{\mathbf{1}}{\boldsymbol{N-m+1}}\right)\sum_{i=1}^{N-m+1} \boldsymbol{logB_r^m(i)}$ Where, **B$_r^m$(i)** = Number of x(j) such that d(x(i)), **x(j) < r / (N – m + 1)**
Energy	Energy: **E** = $\sum_{i=1}^{99} \boldsymbol{x_i}^2$ where $\boldsymbol{x_i}$ is the amplitude of ECG signal at **ith** point.

class (S_W) scatter matrices are calculated respectively as follows [29] (Equation 3, 4):

$$S_B = \sum_{k=1}^{C} q^k (\bar{X}^k - \bar{X})(\bar{X}^k - \bar{X})^T \quad (3)$$

$$S_W = \sum_{k=1}^{C} \sum_{i=1}^{q^i} (X_i - \bar{X}^k)(X_i - \bar{X}^k)^T \quad (4)$$

where $\overline{X} = (1/Q)\sum_{i=1}^{Q} X_i$ is the ensemble mean, and $\bar{X}^k = \left(\frac{1}{q^k}\right)\sum_{k=1}^{q^k} X_i^k$ is the mean of Class k which has q^k samples. An optimal sub-space E_o is determined using the generalized eigenvectors of $\boldsymbol{S_B}, \boldsymbol{S_W}$. Finally, the feature vectors with the optimal discrimination are calculated as Equation 5:

$$Z_i = E_o^T X_i \quad (5)$$

In total first 10 LDA components are used for classification purpose.

2.3.3. ICA

PCA produces mutually uncorrelated features. But in some applications of lower dimensional subspace, PCA can provide minimum class separation. This problem can be overcome by ICA. The ICA mixture model is computed using [28] (Equation 6):

$$x_j = W_j s_j = \sum_{i=1}^{n} w_{ji} s_{ji} \quad (6)$$

where W_j is the total number of patterns present in the data, w_{ji} is the mixing coefficient, S_{ji} is an independent signal which is given by Equation 7:

$$S = Ax \quad (7)$$

A and W are inverse to each other that need to be computed. In our case, we have used Fast ICA algorithm to compute A and W. Here in this work, in total first 10 ICA components are considered for classification purposes.

2.3.4. PPCA

PPCA is a formulation of PCA as a latent variable model. Each data point is assumed to be generated as a linear function of Gaussian latent variables and noise. PPCA involves performing inference in a model involving multivariate Gaussians and performing maximum likelihood. In total first 10 PPCA components are considered for classification purposes.

2.4. Classification

In order to detect AF, two ML algorithms, namely DT and SVM, 2 DL algorithms, namely LSTM and Bi LSTM networks are used.

2.4.1. DT

DT is a non-parametric supervised learning method used for classification. DT makes use of If -and -else rules to arrive at a particular condition. It has a tree like 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 further dividing of the tree. This dividing continues till the leaf node where information gain becomes zero.

2.4.2. SVM

SVM is a supervised machine learning algorithm that classifies the data into one of the two classes. Here C different hyperplanes are constructed for C different classes where each hyperplane separates the class in question from all the other classes. The output of SVM is given by Equation 8:

$$y_j(\mathrm{X}) = \mathrm{sign}\left(\sum_{i=1}^{n_s} \hat{\lambda}_i \, d_i K(X, X_i) + b\right) \quad (8)$$

Where $y_j(\mathrm{X})$, $(1 \le j \le C)$

The given pattern X is assigned to class ϱ_j such that, (Equation 9)

$$\mathrm{J} = \mathit{arg}\max_{\mathrm{j}} \mathrm{y_j}(X) \quad (9)$$

2.4.3. LSTM

The architecture of LSTM is shown in Figure 2. LSTM is used to overcome the problem of vanishing gradient in Recurrent Neural Network (RNN). LSTM structure is

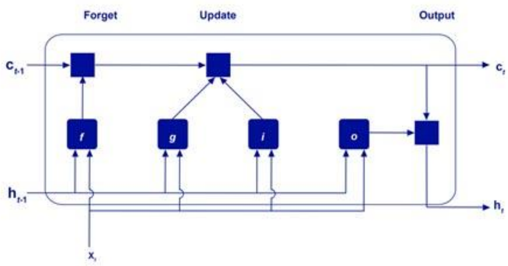

Figure 2. LSTM architecture

a modification of the hidden layer in RNN. LSTM is capable of remembering RNN weights and their input over a long period of time.

2.4.4. BiLSTM

The architecture of the BiLSTM network is shown in Figure 3. BiLSTM is a combination of 2 LSTMs connected in a forward and backward direction by transferring the information in a forward as well as in a backward direction. This makes the architecture to remember future and past information. Figure 4 shows the steps involved in the DL classifier, LSTM, and BiLSTM, respectively. The number of hidden units, choice of the optimizer, and the number of mini-batches are selected using a trial and error-based method. Table 2 highlights the hyperparameters used for the classification purpose.

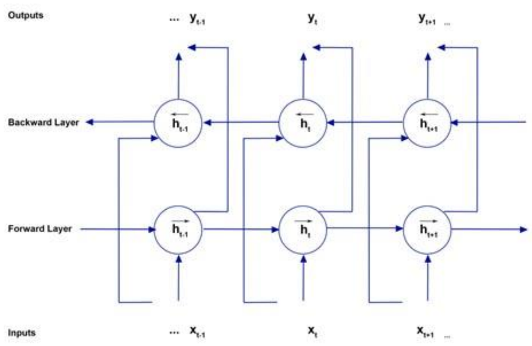

Figure 3. BiLSTM architecture

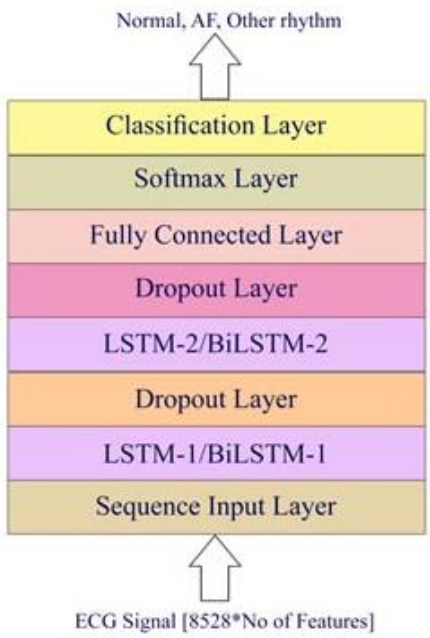

Figure 4. Proposed DL method

Table 2. LSTM/BiLSTM network hyperparameters

Layer	Parameter Name	Parameter Value
LSTM/BiLSTM	Sequence input with 1 dimension Hidden units	100
Other parameters Optimizer	Adam Epochs Dropout rate Mini-batches Classification output	 300 20% 26 crossentropyex

2.4.5. Statistical Metrics

Consider a confusion matrix as shown in Figure 5 (Equation 10-13).

	Normal	AF	Other	
Normal	C_{11}	C_{12}	C_{13}	ΣN
AF	C_{21}	C_{22}	C_{23}	ΣAF
Other	C_{31}	C_{32}	C_{33}	ΣO
	Σn	Σaf	Σo	

Figure 5. Confusion matrix

$$\text{Normal Class Specific Accuracy (AFSA)} = \frac{C_{11}}{\text{Total Normal signals under test}} \tag{10}$$

$$\text{AF Class Specific Accuracy (AFCSA)} = \frac{C_{22}}{\text{Total AF signals under test}} \tag{11}$$

$$\text{Other rhythm Class Specific Accuracy (OCSA)} = \frac{C_{33}}{\text{Total other rhythm signals under test}} \tag{12}$$

$$\text{Overall Accuracy (OA)} = \frac{C_{11} + C_{22} + C_{33}}{\text{Total number of signals under test}} \tag{13}$$

3. Results and Discussion

The proposed algorithm for the classification of ECG signals into normal, AF and other rhythm signals are implemented in MATLAB R2020a. DWT method is used for the removal of high-frequency noise and correction of baseline wander. QRS complex is detected using the Pan-Tompkins algorithm. The raw ECG signal, denoised ECG signal, and detected QRS complex for normal, AF and other rhythm signals are shown in Figure 6, Figure 7 and Figure 8, respectively.

PCA, ICA, LDA, and PPCA methods are used for dimensionality reduction. RR interval features such as RRµ, CVRR, SDRR, SDSD, RMSSD, RR50, pRR50, Triang8, TINN8, pRR20, pRR30, pRR6.25 %, RSA 5RR, SampEn, ApEn are extracted from detected QRS complex. Also, during AF, a P wave is absent which can be manifested in the energy of the ECG signal. Therefore, we have considered energy present in the ECG segment till the R peak is considered as a feature. In total ten dimensionality reduced features, fifteen RR interval features (as discussed in section 2.3), and one energy feature are subjected to classification in order to predict the respective class of the given signal and hence the classification accuracy is calculated. In this study, two ML algorithms, Classification and Regression Trees (CART) and SVM; and two DL algorithms, LSTM and BiLSTM are implemented to predict AF independently. The class-specific accuracy for normal, AF, and other rhythms are calculated. Here 10-fold cross-validation is used and average performance over the 10 folds is obtained. The confusion matrix for the LSTM classifier for PCA and RR interval features combined during the 8th fold of testing is shown in Figure 9. The average accuracy using dimensionality reduction techniques and energy

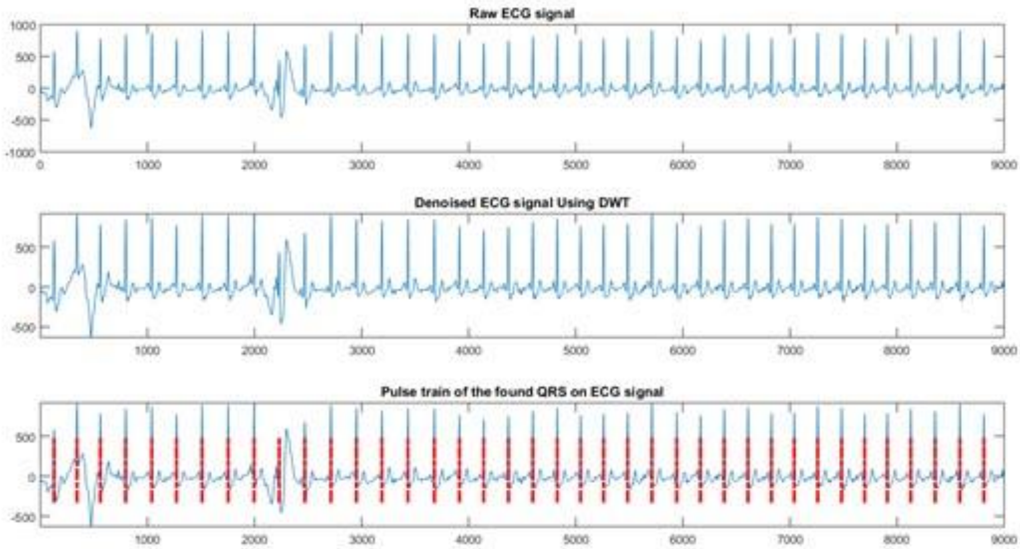

Figure 6. Normal ECG signal

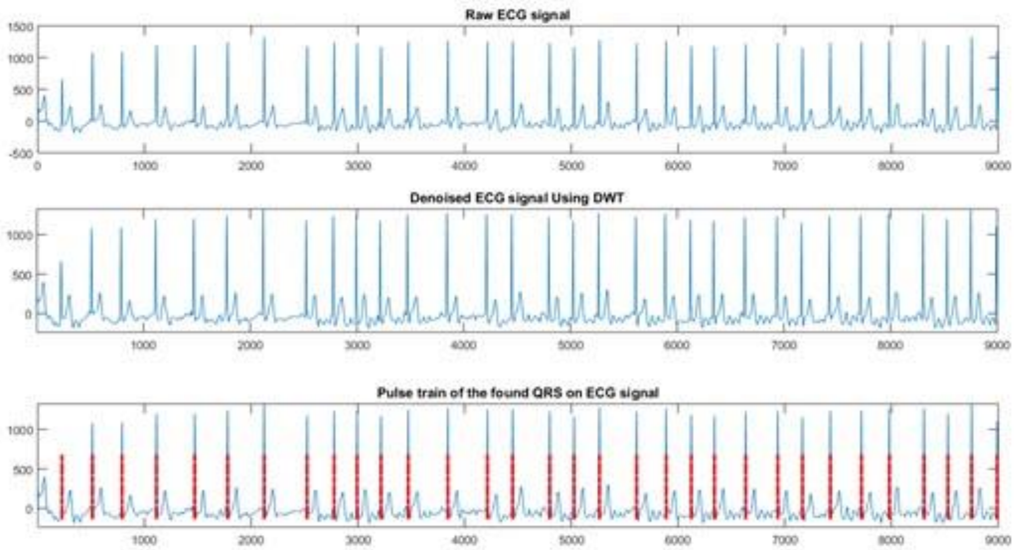

Figure 7. AF signal

features is shown in Table 3. It can be seen that the PCA method combined with the LSTM classifier outperforms in comparison with other classifiers. Figure 10 provides overall accuracy for different classifiers. It can be observed that the LSTM consistently outperforms compared to other classifiers.

The average accuracy using dimensionality reduction techniques, RR interval features, and energy features is shown in Table 4. It is observed that the PCA method combined with the RR interval features along with the LSTM classifier outperformed in comparison with other classifiers.

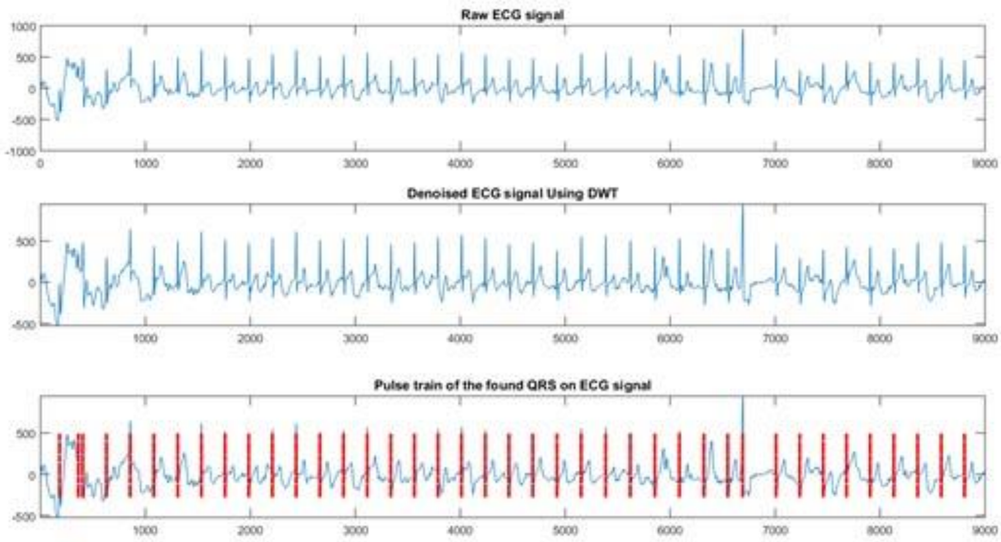

Figure 8. Other rhythm signal

Table 3. Average classification performance of DT, SVM, LSTM, and BiLSTM classifier using dimensionality reduction method

Classifier	NCSA ± std	AFCSA ± std	OCSA ± std	OA ± std
PCA + E + LSTM	**94.67** ± 1.269	**98.14** ± 2.101	**90.52** ± 1.616	**93.65** ± 1.268
PCA + E + BILSTM	93.63 ± 2.019	95.03 ± 4.843	90.19 ± 2.462	92.73 ± 1.925
ICA + E + LSTM	87.08 ± 1.445	79.75 ± 3.277	88.84 ± 2.004	86.68 ± 1.591
ICA + E + BiLSTM	87.86 ± 1.228	80.88 ± 5.550	87.88 ± 2.651	84.10 ± 1.194
PPCA + E + LSTM	96.12 ± 1.622	84.68 ± 6.370	89.64 ± 1.434	93.19 ± 1.208
PPCA + E + BiLSTM	95.53 ± 2.058	83.51 ± 5.113	88.70 ± 1.913	92.39 ± 1.489
LDA + E + LSTM	87.51 ± 1.775	44.14 ± 7.520	84.98 ± 2.875	82.68 ± 1.189
LDA + E + BiLSTM	87.16 ± 1.041	42.83 ± 2.912	85.35 ± 2.304	82.49 ± 1.002
PCA + E + DT	88.69 ± 0.6262	90.18 ± 1.975	80.67 ± 1.387	86.37 ± 0.8402
PCA + E + SVM	81.61 ± 1.867	94.80 ± 1.133	86.26 ± 1.722	87.62 ± 1.404
ICA + E + DT	87.49 ± 0.9954	83.44 ± 1.803	90.24 ± 1.109	87.97 ± 0.5666
ICA + E + SVM	88.57 ± 2.916	95.74 ± 1.172	87.90 ± 1.216	90.73 ± 1.469
PPCA + E + DT	93.81 ± 0.4910	86.43 ± 2.121	87.78 ± 2.393	91.30 ± 0.4663
PPCA + E + SVM	91.10 ± 0.4927	87.05 ± 0.5035	92.94 ± 0.9767	91.25 ± 0.3805
LDA + E + DT	82.36 ± 1.786	49.62 ± 4.832	84.39 ± 1.151	80.16 ± 1.160
LDA + E + SVM	85.33 ± 0.9674	91.62 ± 0.6810	88.17 ± 0.8308	88.38 ± 0.6219

Table 4. Average classification performance of DT, SVM, LSTM, and BiLSTM classifier using dimensionality reduction method and RR interval features

Classifier	NCSA ± std	AFCSA ± std	OCSA ± std	OA ± std
PCA + E + RR + LSTM	**99.54** ± 1.104	**97.51** ± 1.468	**99.50** ± 1.636	**99.45** ± 0.6335
PCA + E + RR + BILSTM	97.64 ± 1.017	95.10 ± 3.284	94.88 ± 1.219	96.43 ± 0.7591
ICA + E + RR + LSTM	89.39 ± 1.696	77.92 ± 5.621	81.98 ± 3.472	86.12 ± 1.778
ICA + E + RR + BiLSTM	90.82 ± 1.704	74.96 ± 4.144	81.47 ± 3.302	86.57 ± 1.858
PPCA + E + RR + LSTM	97.82 ± 1.222	86.46 ± 5.169	89.02 ± 2.973	94.10 ± 0.9710
PPCA + E + RR + BiLSTM	96.99 ± 1.343	84.23 ± 5.428	89.76 ± 1.553	93.64 ± 0.7059
LDA + E + RR + LSTM	91.75 ± 1.982	69.16 ± 8.825	84.21 ± 1.997	87.37 ± 1.219
LDA + E + RR + BiLSTM	90.77 ± 1.869	71.36 ± 6.017	76.69 ± 4.088	84.79 ± 1.488
PCA + E + RR + DT	98.12 ± 0.2224	98.00 ± 1.639	95.98 ± 0.3445	97.42 ± 0.1898
PCA + E + RR + SVM	96.85 ± 0.4198	97.67 ± 0.4529	95.07 ± 0.5245	96.53 ± 0.3147
ICA + E + RR + DT	99.45 ± 0.3869	96.34 ± 1.915	98.13 ± 0.9759	98.77 ± 0.3788
ICA + E + RR + SVM	96.40 ± 0.6582	97.66 ± 0.2868	95.18 ± 0.2540	96.45 ± 0.3586
PPCA + E + RR + DT	99.48 ± 0.2257	97.25 ± 1.698	98.39 ± 0.7488	98.93 ± 0.2815
PPCA + E + RR + SVM	97.09 ± 0.5920	97.89 ± 0.4286	95.56 ± 0.7737	96.85 ± 0.5494
LDA + E + RR + DT	99.23 ± 0.2734	96.50 ± 1.140	97.91 ± 0.5501	98.50 ± 0.2673
LDA + E + RR + SVM	99.43 ± 0.4473	91.07 ± 0.7932	90.89 ± 1.028	94.04 ± 0.5301

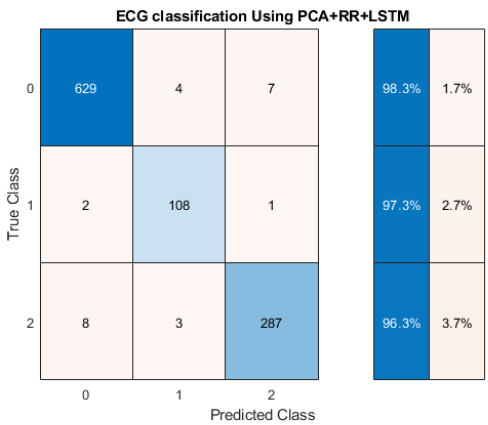

Figure 9. Confusion matrix

Figure 11 shows variation in overall accuracy plotted against classification using DT, SVM, LSTM, and BiLSTM classifiers. It can be seen that the PCA combined with the RR interval features and the LSTM classifier provided the highest median accuracy and lowest variability during the ten-fold cross-validation.

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 5 highlights some of the work carried out by authors on detecting AF using the Physionet Challenge 2017. Rao *et al.* [2] suggested a power spectrum-based method for the classification of ECG signals using CNN and observed an accuracy of 94.67%. The present study focuses on ECG signal classification using dimensionality reduction techniques combined with energy and RR interval features. Chang *et al.* [30] proposed a method to classify ECG signals using temporal and spectral features using an LSTM classifier and achieved an accuracy of 98.3%. Athif *et al.* [31] proposed a statical and morphological feature-based method to detect AF using an SVM classifier and recorded an accuracy of 96.1%. Sanchez *et al.* [32] suggested a gramian angular summation field method to detect AF using CNN and achieved an accuracy of 97.6%. Shi *et al.* [33] suggested a multiple-feature fusion method to diagnose AF using 1D CNN and achieved an accuracy of 91.7%. Zhao *et al.* [34] proposed a 24-layer CNN network to diagnose AF and recorded an accuracy of 87.1%. Mousavi *et al.* [35] suggested a hierarchical attention model using BiLSTM classifier to discriminate ECG signals into different rhythms and recorded an accuracy of 96.98%. Alsaleem *et al.* [36] suggested spectrogram-based features using a two- layer BiLSTM classifier to diagnose AF and observed an accuracy of 91.4%. Liaqat *et al.* [37]

Table 5. Overview of studies on the classification of ECG signal using Physionet Challenge 2017 dataset

Literature	Methods	Classifier	OA
Rao *et al.* [2] 2021	Power Spectrum	CNN	94.67%
Chang *et al.* [30] 2018	Spectral and temporal features	LSTM	98.3%
Athif *et al.* [31] 2018	Statical and morphological features	SVM	96.1%
Sanchez *et al.* [32] 2019	Gramian angular summation fields	CNN	97.6%
Shi *et al.* [33] 2021	Deep features and AF features	1D CNN	91.7%
Zhao *et al.* [34] 2020	Deep features.	24-layerCNN	87.1%
Mousavi *et al.* [35] 2020	Hierarchical attention model	BiLSTM	96.98%
Alsaleem *et al.* [36] 2020	Spectrogram based features	2-layer BiLSTM	91.4%
Liaqat *et al.* [37] 2020	RR interval, single and full-wave features	LSTM	87.5%
Zhang *et al.* [38] 2021	FFT	LSTM-CNN	95.28%
Disha *et al.* [39] 2022	Deep features	VGG16	97.60%
Geweid *et al.* [40] 2022	Hybrid approach	Dual SVM	99.27%
Rao *et al.* [41] 2024	RR intervals	CNN-LSTM	98.25%
Current Study	Dimensionally reduced, Energy and RR series feature	2-layer LSTM	99.45%

suggested an LSTM architecture to diagnose AF using RR features, single and full-wave method and achieved an accuracy of 87.5%. Zhang *et al.* [38] suggested a method based on Fast Fourier Transform (FFT) using an LSTM-CNN classifier and recorded an accuracy of 95.28%. Disha *et al.* [39] suggested a VGG16 based method for AF extracting deep features and observed an accuracy of 97.6%. Geweid *et al.* [40] designed a hybrid approach to detect AF using dual SVM and observed an accuracy of 99.27%. Rao *et al.* [41] implemented CNN_LSTM based architecture using RR intervals of ECG and observed OA of 98.25%. In the first approach, only dimensionality reduction techniques such as PCA, LDA, ICA, and PPCA are performed on denoised ECG signals using the DWT method and classifiers. In the second approach, features derived from dimensionality reduction techniques combined with RR interval features are used for the classification of ECG signals. ML models such as DT, SVM, and DL models such as LSTM and BiLSTM are used for classification purposes.

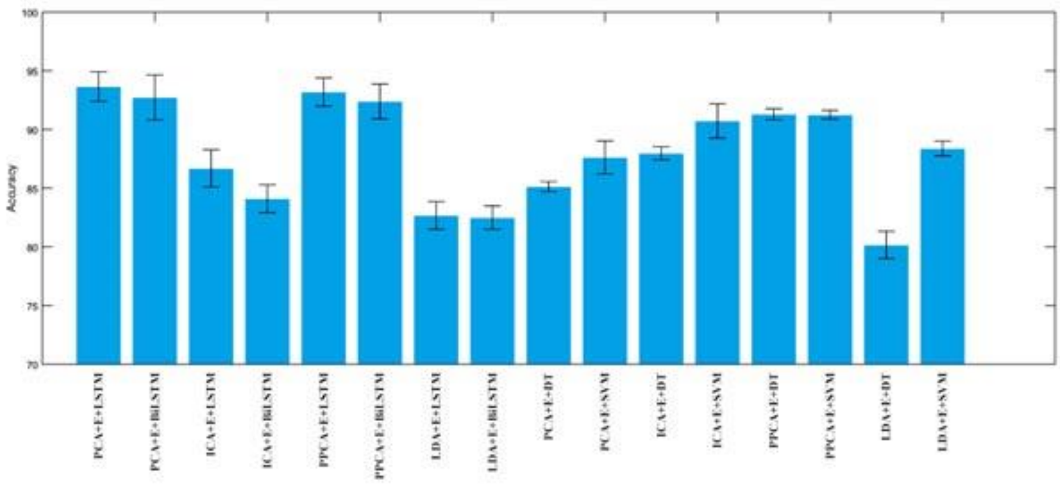

Figure 10. Variation of overall accuracy with dimension reduction technique

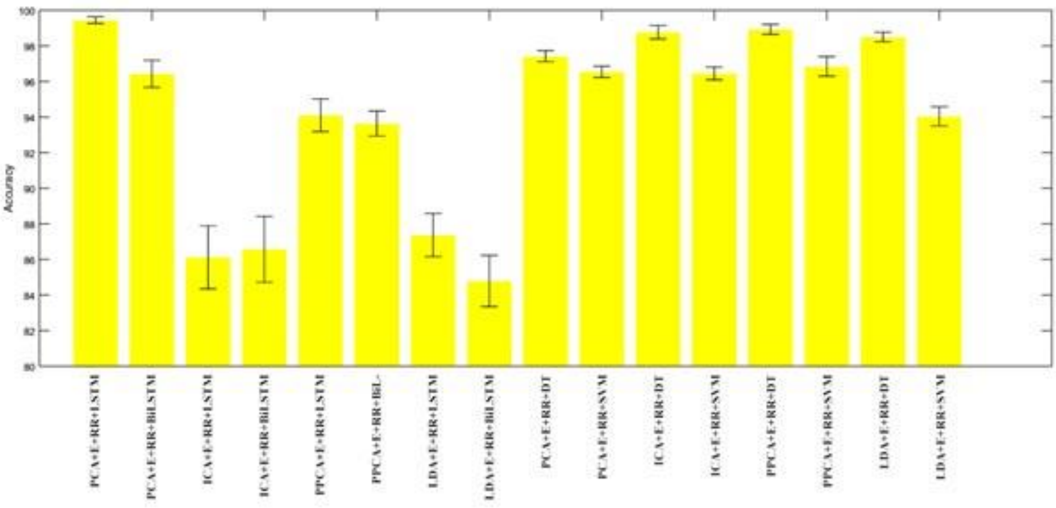

Figure 11. Variation of overall accuracy combined with dimension reduction technique and RR interval features

4. Conclusion

In this study, we have proposed a framework to detect AF using dimensionally reduction techniques and RR intervals. In the first approach, AF is detected using dimensionality reduction techniques. In the second approach, the features extracted from dimensionality reduction techniques are combined with RR interval features to detect AF. In total ten dimensionality reduced features, fifteen RR interval features and one energy feature are used in this study. Four dimensionality reduction techniques namely, PCA, ICA, LDA, and PPCA are used in this study. Two ML algorithms, namely DT and SVM, and two DL algorithms, namely LSTM and BiLSTM are used to predict AF. The accuracy is found to be consistent without much variability in all the folds during 10-fold cross-validation.

An accurate method to detect AF using features extracted from dimensionality reduction techniques and features extracted from RR intervals is very much required for long-term monitoring. The long-term monitoring system consists of more features that carry more information about AF detection. The computer-assisted cardiac diagnosis tool not only reduces the workload of physicians but also reduces the variability due to intra-observer validation.

References

1- Virani, Salim S., Alvaro Alonso, Hugo J. Aparicio, Emelia J. Benjamin, Marcio S. Bittencourt, Clifton W. Callaway, April P. Carson et al. *"Heart disease and stroke statistics—2021 update: a report from the American Heart Association*." Circulation 143, no. 8: e254-e743, (2021); https://doi.org/10.1161/CIR.0000000000000950.

2- Rao, SK Shrikanth, MaheshKumar H. Kolekar, and Roshan Joy Martis. "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*), pp. 1-6. IEEE, (2021); https://doi: 10.1109/CONECCT52877.2021.9622533.

3- 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); https://doi.org/10.1155/2021/6691177.

4- Petmezas, Georgios, Kostas Haris, LeandrosStefanopoulos, Vassilis Kilintzis, Andreas Tzavelis, John A. Rogers, Aggelos K. Katsaggelos, and NicosMaglaveras. "Automated atrial fibrillation detection using a hybrid CNN-LSTM network on imbalanced ECG datasets." *Biomedical Signal Processing and Control* 63: 102194, (2021); https://doi.org/10.1016/j.bspc.2020.102194.

5- Radhakrishnan, Tejas, Jay Karhade, S. K. Ghosh, P. R. Muduli, R. K. Tripathy, and U. Rajendra Acharya. "AFCNNet: Automated detection of AF using chirplet transform and deep convolutional bidirectional long short term memory network with ECG signals." *Computers in*

Biology and Medicine 137: 104783, (2021); https://doi.org/10.1016/j.compbiomed.2021.104783.

6- Liang, Yongbo, Shimin Yin, Qunfeng Tang, Zhenyu Zheng, Mohamed Elgendi, and Zhencheng Chen. "Deep learning algorithm classifies heartbeat events based on electrocardiogram signals." *Frontiers in Physiology* 11 (2020); https://doi: 10.3389/fphys.2020.569050.

7- Ghosh, Samit Kumar, Rajesh K. Tripathy, Mario RA Paternina, Juan J. Arrieta, Alejandro Zamora-Mendez, and Ganesh R. Naik. "Detection of atrial fibrillation from single lead ECG signal using multirate cosine filter bank and deep neural network." *Journal of medical systems* 44: 1-15, (2020); https://doi:10.1007/s10916-020-01565-y.

8- Ahmed, Nuzhat, and Yong Zhu. "Early detection of atrial fibrillation based on ECG signals." Bioengineering 7, no. 1: 16, (2020); https://doi.org/10.3390/bioengineering7010016

9- Shankar, M. Gowri, and C. Ganesh Babu. "An exploration of ECG signal feature selection and classification using machine learning techniques." *Int. J. Innov. Technol. Explor. Eng.* Regul 9: 797-804, (2020); https://doi:10.35940/ijitee.C8728.019320.

10- Kleyko, Denis, EvgenyOsipov, and Urban Wiklund. "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." *Biomedical Physics & Engineering Express* 6, no. 2: 025010, (2020); https://doi.org/10.1088/2057-1976/ab6e1e.

11- Wu, Xiaodan, Yumeng Zheng, Chao-Hsien Chu, and Zhen He. "Extracting deep features from short ECG signals for early atrial fibrillation detection." *Artificial Intelligence in Medicine* 109: 101896, (2020); https://doi.org/10.1016/j.artmed.2020.101896.

12- Wang, Jibin, Ping Wang, and Suping Wang. "Automated detection of atrial fibrillation in ECG signals based on wavelet packet transform and correlation function of random process." *Biomedical Signal Processing and Control* 55: 101662, (2020); https://doi.org/10.1016/j.bspc.2019.101662.

13- Li, Mingchun, Gary He, and Baofeng Zhu. "Atrial Fibrillation Detection Based on the Combination of Depth and Statistical Features of ECG." In Proceedings of the 2019 3rd International Conference on Graphics and Signal Processing, pp. 105-112, (2019); https://doi.org/10.1145/3338472.3338485.

14- Hagiwara, Yuki, Hamido Fujita, Shu Lih Oh, Jen Hong Tan, Ru San Tan, Edward J. Ciaccio, and U. Rajendra Acharya. "Computer-aided diagnosis of atrial fibrillation based on ECG signals: A review." *Information Sciences* 467: 99-114, (2018); https://doi.org/10.1016/j.ins.2018.07.063.

15- Faust, Oliver, Alex Shenfield, Murtadha Kareem, Tan Ru San, Hamido Fujita, and U. Rajendra Acharya. "Automated detection of atrial fibrillation using long short-term memory network with RR interval signals." *Computers in biology and medicine* 102: 327-335, (2018); https://doi.org/10.1016/j.compbiomed.2018.07.001.

16- Maji, U., M. Mitra, and S. Pal. "Automatic detection of atrial fibrillation using empirical mode decomposition and statistical approach." *Procedia Technology* 10: 45-52, (2013); https://doi.org/10.1016/j.protcy.2013.12.335.

17- Dash, S., K. H. Chon, S. Lu, and E. A. Raeder. "Automatic real time detection of atrial fibrillation." *Annals of biomedical engineering* 37, no. 9: 1701-1709, (2009); https://doi:10.1007/s10439-009-9740-z.

18- Clifford, Gari D., Chengyu Liu, Benjamin Moody, H. Lehman Li-wei, Ikaro Silva, Qiao Li, A. E. Johnson, and Roger G. Mark. "AF classification from a short single lead ECG recording: The PhysioNet/computing in cardiology challenge 2017." *In 2017 Computing in Cardiology (CinC*), pp. 1-4. IEEE, (2017); https://doi.10.22489/CinC.2017.065-469.

19- Martis, Roshan Joy, U. Rajendra Acharya, and HojjatAdeli. "Current methods in electrocardiogram characterization." *Computers in biology and medicine* 48: 133-149, (2014); https://doi.org/10.1016/j.compbiomed.2014.02.012.

20- Rao,SKShrikanth, Krithika, K., M. Akhila, and Roshan Joy Martis. "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), pp. 01-07, (2021). IEEE; https:// doi.10.1109/R10-HTC53172.2021.9641550

21- Rizwan, Muhammed, Bradley M. Whitaker, and David V. Anderson. "AF detection from ECG recordings using feature selection, sparse coding, and ensemble learning." *Physiological measurement* 39, no. 12: 124007, (2018); https://doi.org/10.1088/1361-6579/aaf35b.

22- Shaffer, Fred, and J. P. Ginsberg. "An overview of heart rate variability metrics and norms." *Frontiers in public health* 5: 258, (2017); https://doi.org/10.3389/fpubh.2017.00258.

23- Mali, Barbara, Sara Zulj, Ratko Magjarevic, DamijanMiklavcic, and TomazJarm. "Matlab-based tool for ECG and HRV analysis." *Biomedical Signal Processing and Control* 10: 108-116, (2014); https://doi.org/10.1016/j.bspc.2014.01.011.

24- Smith, Anne-Louise, Harry Owen, and Karen J. Reynolds. "Heart rate variability indices for very short-term (30 beat) analysis. Part 1: survey and toolbox." *Journal of clinical monitoring and computing* 27, no. 5: 569-576, (2013); https://doi.org/10.1007/s10877-013-9471-4.

25- Smith, Anne-Louise, Harry Owen, and Karen J. Reynolds. "Heart rate variability indices for very short-term (30 beat) analysis. Part 2: validation." *Journal of clinical monitoring and computing* 27, no.5: 577-585, (2013); https://doi: 10.1007/s10877-013-9473-2.

26- Martis, Roshan Joy, U. Rajendra Acharya, K. M. Mandana, Ajoy Kumar Ray, and Chandan Chakraborty. "Application of principal component analysis to ECG signals for automated diagnosis of cardiac health." *Expert Systems with Applications* 39, no. 14: 11792-11800, (2012); https://doi.org/10.1016/j.eswa.2012.04.072.

27- Ke, Calvin, Rajeev Gupta, Denis Xavier, Dorairaj Prabhakaran, Prashant Mathur, Yogeshwar V. Kalkonde, Patrycja Kolpak et al. "Divergent trends in ischaemic heart disease and stroke mortality in India from 2000 to 2015: a nationally representative mortality study." *The Lancet Global Health* 6, no. 8: e914-e923, (2018); https://doi.org/10.1016/S2214-109X(18)30242-0.

28- Martis, Roshan Joy, U. Rajendra Acharya, and Lim Choo Min. "ECG beat classification using PCA, LDA, ICA and discrete wavelet transform." *Biomedical Signal Processing and Control* 8, no. 5: 437-448, (2013); https://doi.org/10.1016/j.bspc.2013.01.005.

29- Sadaghiyanfam, Safa, and Mehmet Kuntalp. "Analysing the Performance of LDA (Linear Discriminant Analysis) Feature for Diagnosing PAF (Paroxysmal Atrial Fibrillation) Patients." *In 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science* (EBBT), pp. 1-4. IEEE, (2019); https://doi:10.1109/EBBT.2019.8741871.

30- Chang, Yen-Chun, Sau-Hsuan Wu, Li-Ming Tseng, Hsi-Lu Chao, and Chun-Hsien Ko. "AF detection by exploiting the spectral and temporal characteristics of ECG signals with the LSTM model." *In 2018 Computing in cardiology conference* (CinC), vol. 45, pp. 1-4. IEEE, (2018); https:// doi: 10.22489/CinC.2018.266.

31- Athif, Mohamed, Pamodh Chanuka Yasawardene, and Chathuri Daluwatte. "Detecting atrial fibrillation from short single lead ECGs using statistical and morphological features." *Physiological measurement* 39, no. 6: 064002, (2018); https:// doi 10.1088/1361-6579/aac552.

32- Sánchez, FA Rivera, and JA González Cervera. "ECG classification using artificial neural networks." *In Journal of Physics: Conference Series*, vol. 1221, no. 1, p. 012062. IOP Publishing, (2019); https://doi.org/10.1088/1742-6596/1221/1/012062.

33- Shi, Jingang, Iman Alikhani, Xiaobai Li, Zitong Yu, TapioSeppänen, and Guoying Zhao. "Atrial fibrillation detection from face videos by fusing subtle variations." *IEEE Transactions on Circuits and Systems for Video Technology* 30, no. 8: 2781-2795, (2019); https://doi:10.1109/TCSVT.2019.2926632.

34- Zhao, Yunxiang, Jinyong Cheng, Ping Zhang, and Xueping Peng. "ECG classification using deep CNN improved by wavelet transform." *Computers, Materials & Continua* 64, no. 3: 1615-1628, (2020); https://doi:10.32604/cmc.2020.09938.

35- Mousavi, Sajad, Fatemeh Afghah, and U. Rajendra Acharya. "HAN-ECG: An interpretable atrial fibrillation detection model using hierarchical attention networks." *Computers in Biology and Medicine* 127: 104057, (2020); https://doi.org/10.1016/j.compbiomed.2020.104057.

36- Alsaleem, Mona, and Md Saiful Islam. "POSTER: Atrial Fibrillation Detection Using a Double-Layer Bi-Directional LSTM Neural Networks." *In 2020 First International Conference of Smart Systems and Emerging Technologies (SMARTTECH*), pp. 266-267. IEEE, (2020); https://doi:10.1109/SMART-TECH49988.2020.00071.

37- Liaqat, Sidrah, Kia Dashtipour, Adnan Zahid, Khaled Assaleh, Kamran Arshad, and Naeem Ramzan. "Detection of atrial fibrillation using a machine learning approach." *Information* 11, no. 12: 549, (2020); https://doi.org/10.3390/info11120549.

38- Zhang, Xiangyu, Jianqing Li, Zhipeng Cai, Li Zhang, Zhenghua Chen, and Chengyu Liu. "Over-fitting suppression training strategies for deep learning-based atrial fibrillation detection." *Medical & Biological Engineering & Computing* 59, no. 1: 165-173, (2021); https://doi.org/10.1007/s11517-020-02292-9.

39- Disha, S., B. Deekshitha, A. Anwitha, U. M. Kavyashree, Shrikanth Rao SK, and Roshan Joy Martis. "Deep Learning assisted tool for Atrial Fibrillation detection using RR Intervals." *In 2022 International Conference on Connected Systems & Intelligence* (CSI), pp. 1-4. IEEE, (2022); https:// doi: 10.1109/CSI54720.2022.9924134.

40- Geweid, Gamal GN, and Jiande DZ Chen. "Automatic classification of atrial fibrillation from short single-lead ECG recordings using a Hybrid Approach of Dual Support Vector Machine." *Expert Systems with Applications* 198: 116848, (2022); https://doi.org/10.1016/j.eswa.2022.116848.

41- SK, Shrikanth Rao, MaheshKumar H. Kolekar, and Roshan Joy Martis. "A deep learning approach for detecting atrial fibrillation using RR intervals of ECG." Frontiers in Biomedical Technologies 11, no. 2 (2024): 255-264; https://doi.org/10.18502/fbt.v11i2.15343