International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10s
DOI: https://doi.org/10.17762/ijritcc.v11i10s.7599
***Article Received**: 29 May 2023 **Revised**: 25 July 2023 **Accepted**: 10 August 2023*

Classification of Atrial Fibrillation using Random Forest Algorithm

Suguna G C[1], Sunita Shirahatti[2], Sowmya R Bangari[3], *Veerabhadrappa S T[4]
[1]Dept. ECE, JSS Academy of Technical Education, Bengaluru, India
e-mail: sugunagc@jssateb.ac.in
[2]Dept. ECE, JSS Academy of Technical Education, Bengaluru, India
e-mail: sunithalshirahatti@jssateb.ac.in
[3]Dept. ECE, JSS Academy of Technical Education, Bengaluru, India
e-mail: sowmyarbangari @jssateb.ac.in
[4]Dept. ECE, JSS Academy of Technical Education, Bengaluru, India
*Corresponding Author: e-mail: veerabhadrappast@jssateb.ac.in

Abstract—The electrocardiogram is indicates the electrical activity of the heart and it can be used to detect cardiac arrhythmias. In the present work, we exhibited a methodology to classify Atrial Fibrillation (AF), Normal rhythm, and Other abnormal ECG rhythms using a machine learning algorithm by analyzing single-lead ECG signals of short duration. First, the events of ECG signals will be detected, after that morphological features and HRV features are extracted. Finally, these features are applied to the Random Forest classifier to perform classification. The Physionet challenge 2017 dataset with more than 8500 ECG recordings is used to train our model. The proposed methodology yields an F1 score of 0.86, 0.97, and 0.83 respectively in classifying AF, normal, other rhythms, and an accuracy of 0.91 after performing a 5-fold cross-validation.

Keywords- Classification; ECG morphological ; HRV features; Random Forest.

I. INTRODUCTION

The mortality due to heart failure is associated with population aging. The automatic detection and analysis of cardiac arrhythmias are playing an important role in hospitals. The most prevalent continuous heart rhythm problem, Atrial fibrillation (AF), results in severe cardiac arrest. It is distinguished by disorganized electrical activity in the atria and rapidly circulating waves of abnormal electrical signals that constantly stimulate the atrium rather than the sinus node. The AF is commonly observed in senior citizen and on many occasions, its diagnosis is highly complex due to paroxistic behavior and the absence of symptoms in some cases. Atrial fibrillation is an abnormal and frequently very rapid coronary heart rhythm (arrhythmia) that could lead to blood clots inside the heart. Normal heart rhythm is often known as regular sinus rhythm due to the fact the SA (sinus) node fires at the rate of 60 – 120 beats per minute.

II. RELATED WORK

The ECG signal is preprocessed to remove the baseline wandering, power line, and high-frequency noise using various filtering operations, de-noising before extraction of features [1-3]. Many research groups demonstrated the detection of arrhythmias is based on the irregularities in the cardiac cycle and as well as morphological parameters [4,5]. To support physicians, the automatic detection of cardiac arrhythmias is essential. Several researchers are attempted to detect and classify arrhythmias using machine learning and deep learning algorithms. In most of the studies, the morphological features ECG and HRV were extracted, FFNN, SVM, and ensemble classifiers to detect AF [6-10]. This study aims to explore the feasibility of AF classification based on heart rate variability (HRV) and beat morphology analysis using a random forest classifier.

Recent research has shown that computerized algorithms for classifying arrhythmias are capable of detecting cardiac arrhythmias with a fairly high accuracy rate. The research group describes optimum-path forest (OPF) classifier-based automatic arrhythmia identification from ECG and demonstrated that OPF and SVM-RBF classifiers produce better results [11]. The Random Forests (RF) classifier, which examines wavelet properties from the ECG heartbeat signal, was proposed for the classification of arrhythmia [12]. The five most prevalent arrhythmia classes may be classified with a high degree of accuracy using an RF classifier. The researchers have shown five groups of arrhythmias using a linear discriminate-based classification model.

In several investigations, neural network-based cardiac arrhythmia classifiers were also suggested. A block-based neural network classifier for identifying arrhythmias was introduced and used the PSO technique to train their neural network, and the shown higher accuracy [13]. The MLP was employed used to classify arrhythmias with wavelet and Fourier characteristics features from the ECG [14]. In, a classifier based on feed-forward neural networks was implemented to categorize arrhythmia ECG signals where wavelet and PCA characteristics were used as the input for a neural network classifier. They used the MD-PSO approach to optimize the neural network. Bayesian neural network, SVM and logistic regression back propagation algorithm , a voting feature interval-based classifier, KNN with kernel difference weight, Markov Blanket algorithm and Fusion techniques provide a trade-off between error and rejection rates, were implemented to detect arrhythmias[15-16]. To tackle this problem of several classes, data resampling is used, allowing the classifier to correctly categorize each class [17].

III. ORIGINALITY

In this study, the Physionet 2017 datasets were used to classify four different classes. The dataset includes 5050, 738, 2456, and 248 normal rhythms, AF, Other rhythms, and noisy ECG signals respectively. The proposed method for ECG classification involves four phases, which include preprocessing, feature extraction, training the machine learning model, and testing the model. Methodology of ECG classification algorithm is depicted in Figure 1 with an accuracy of 91%.

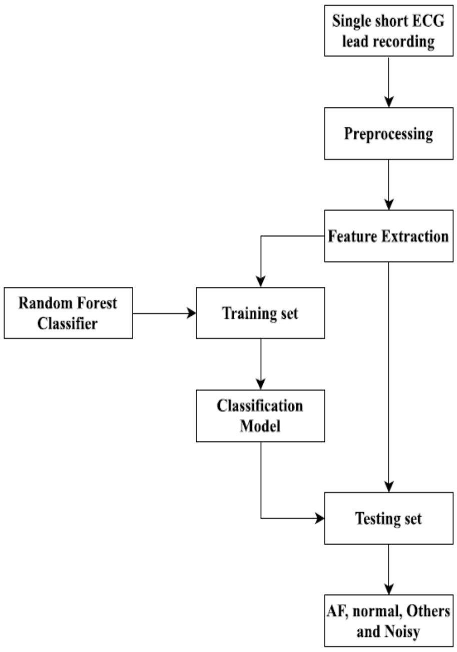


Figure 1. Methodology of ECG classification system

IV. SYSTEM DESIGN

System design of ECG classification system involves Preprocessing with band pass filter , Morphological Features P,Q,R,S T of ECG based on wavelet techniques, linear and nonlinear HRV analysis of ECG signal and random forest classifier, classify the AF,normal,other and noisy signal.

A. Preprocessing

The goal of preprocessing phase is to reduce various noises such as baseline wander, high-frequency noises due to body movement, power line interference, voltage fluctuation of the sensor and many other reasons which can heavily corrupt signal [18].

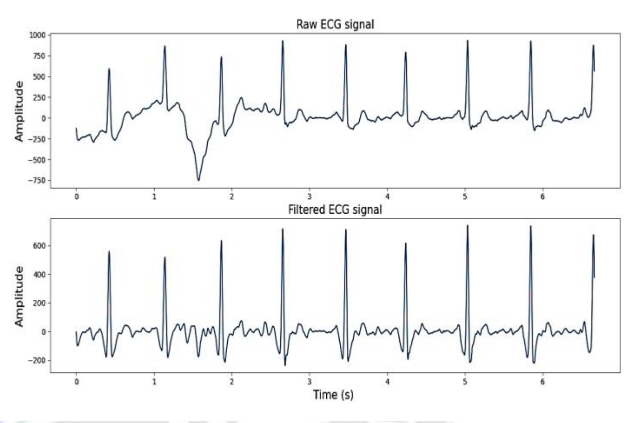


Figure 2. Filtered ECG curve

Designed of a digital FIR band pass filter with order 91 obtained from from equation (1), and relative side lobe attenuation was -42.6 dB for the given specification, having cutoff frequencies of 3 Hz and 45 Hz and normalize the frequency to Nyquist frequency (Fs/2) for removal noise in ECG. The original and filtered ECG signal is shown in Figure 2 and features were extracted for the classification.

$$\text{Filter Order} = 0.3 * \text{F}_s \qquad (1)$$

B. Morphological Features

The amplitudes and the corresponding location of P, Q, R, S and T were estimated from filtered ECG signals. R peak detected based on the combination of finite machine state and second derivative using Pan Tompkins algorithm [18-19]. This algorithm is also included a first-order derivative to remove low-frequency noise and baseline wandering. The higher values of the first derivative of ECG correspond to QRS-complex and a zero value at the inflection point of R-peak. To isolate the R-peak from ECG based on the second derivative the local maxima or minima were determined and further enhance the nonlinearity of the signal by squaring the second derivative which removes the effect of changes in the signal polarity. At last moving average window of N-samples is obtained by the integration window. Sampling frequency Fs depends on the width N of the integration window. In this study, empirically we found N=7 and which is optimal for a sampling frequency of 300 Hz. In the last stage, the adaptive amplitude and duration thresholding technique is applied to determine the R peak of the ECG signal.

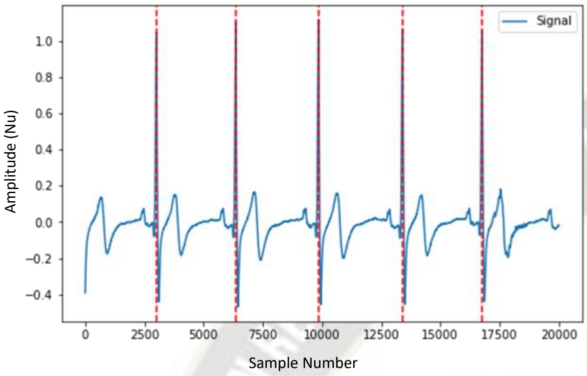


Figure 3. Visuals of R-peak. Red dashed line represents the detected R-peaks

In addition to the R peak, the duration of P, ST-segment and width of QRS are required to classify AF arrhythmia. A wavelet-based ECG delineator is used to detect P, Q, R, S and T peaks. Wavelet transform is used to decompose signal as a combination of a set of basis functions, which is obtained through scaling and translation of a mother wavelet Ψ(t). Wavelet transform of signal x(t) is defined as

$$W_{a,b}(x) = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} x(t)\psi\left(\frac{t-b}{a}\right)dt, \quad a > 0 \qquad (2)$$

where a and b are time scaling and shifting operators respectively. In this study, a quadratic spline mother wavelet is used. Li et al, proposed a multiscale approach using wavelet transform to estimate QRS points. In the later stage based on the QRS delineation, the onset points and peak points of the QRS complex, P wave, and T wave are detected as shown in Figure 3 and Figure 4 .

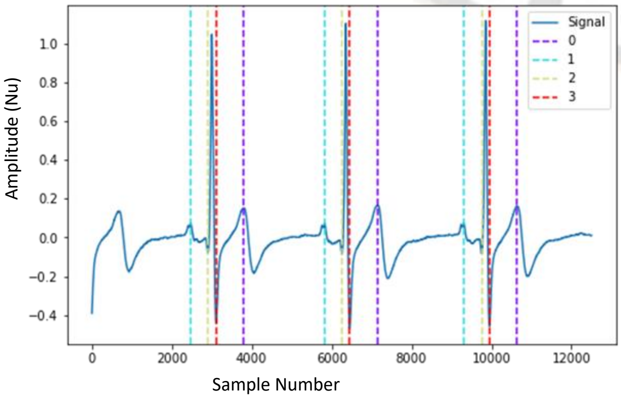


Figure 4.Visuals of PQST-peaks. 0,1,2,3 represents P,Q,S,T peaks respectively

C. *HRV Features*

The linear and nonlinear HRV parameters were extracted from RR intervals. The time-domain parameters such as normal heart rate (HR), standard deviation (SDNN), root mean square (RMSSD), the difference between consecutive heart rates (pNN50, pNN20) and SDSD features were extracted from RR intervals [20]. These features have been used for the classification of four classes of arrhythmia. HRV features defined as below

(a) Standard deviation (SDNN)

The standard deviation measures the spread of the data, where σ and μ are the standard deviation and arithmetic mean of {X1, X2, X3, X4, X5, ……. Xn} then

$$\sigma = \sqrt{\frac{1}{n}\sum_{m=1}^{m=n}(x_m - \mu)^2} \qquad (3)$$

(b)Root Mean Square (RMSSD)

The Root Mean Square is the square root of the mean square i.e., it is the arithmetic mean of the square of data. It is also known as the quadratic mean.

$$RMS = \sqrt{\frac{1(x_1^2+x_2^2+x_3^2+x_4^2+\cdots+x_n^2}{n}} \qquad (4)$$

(c)Percentile (pNN50, pNN20)

A percentile is a value of a variable indicating below which a certain percentage of observation fall. The percentile is given by,

$$n=(P/100)*N \qquad (5)$$

where P is the percentile,

N is the number of data points,

n-is ordinal rank when data is sorted in ascending order.

The following different features can be extracted from the percentile. Percentile of 10th, 20th, 90th, etc... of different values can be calculated. IQR-Inter quartile range is the difference between the 75th percentiles to the 25th percentile.

(d)SDSD

The standard deviation of the differences between successive NN intervals.

D. *Random Forest*

Random forest is an ensemble supervised machine learning algorithm which is a combination of two or more algorithm such as Decision Trees and Bagging algorithm. The Random forest combines the forecast made by each individual decision

DOI: https://doi.org/10.17762/ijritcc.v11i10s.7599
***Article Received**: 29 May 2023 **Revised**: 25 July 2023 **Accepted**: 10 August 2023*

trees to obtain a optimal solution. The random forest splits the nodes into sub-nodes randomly and selects the best features out of it to produce a better classification. After the best features are obtained from each decision tree , the Random Forest either uses a averaging or majority vote approach based on whether it is a regression model or a classification model . This process of averaging or majority voting is called aggregation. The decision tress provides low bias and high variance .When these decision trees are combined using majority vote process, low variance is obtained which in turn provides high accuracy and reduces over fitting. High dimensionality and varied feature types are well-handled by random forest. Crucial hyper parameters of the algorithm are "n estimators," "max depth," and "min samples split". The "n estimators" refers to the number of decision trees employed; maximum depth indicates the height of the tree; more splits, the more information about the data. To avoid over fit of the algorithm maximum depth should be optimal. .Figure 5 shows that each decision tree will be generated based on m distinct attributes .The best prediction from each tree will be stored [21].

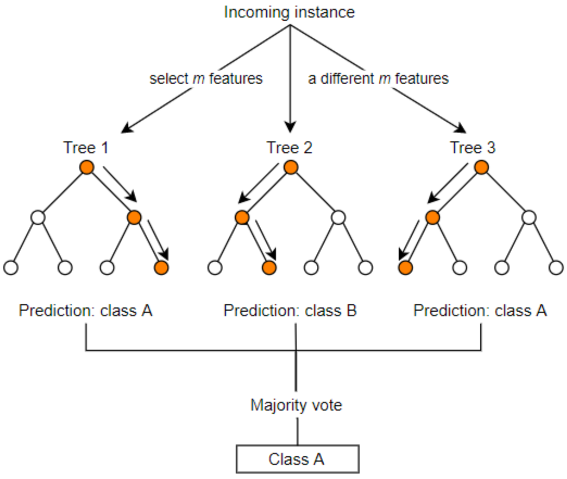


Figure 5. Workflow of random forest

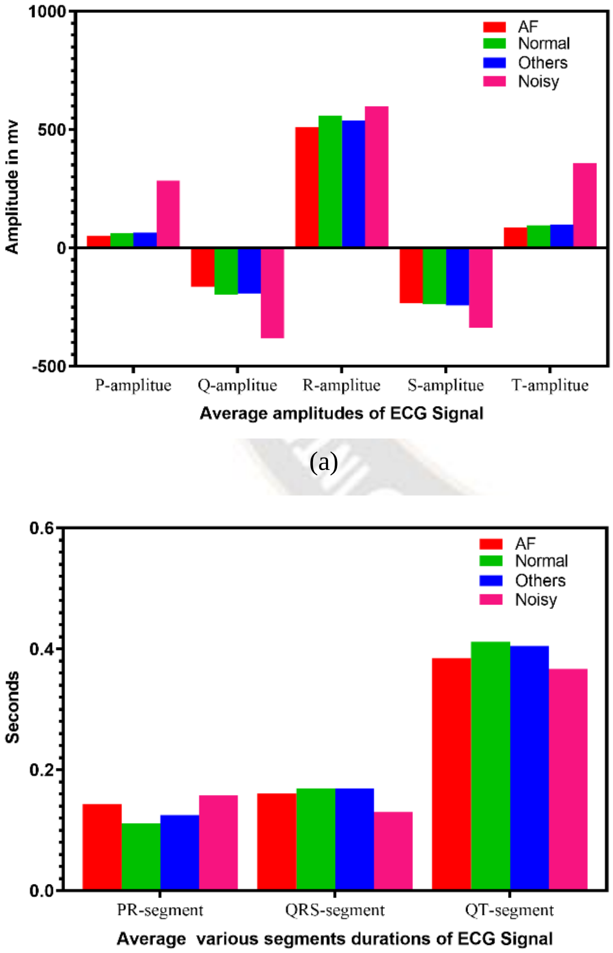
V. EXPERIMENTAL AND ANALYSIS

The ECG signals from Physionet 2017 challenge were analyzed and extracted morphological features, time-domain HRV and nonlinear HRV features. It is seen from Table1, that the heart rate in AF class is significantly higher as compared to normal and other arrhythmias. The amplitude and duration of various waves of the ECG signal as shown in Figure 6. The mean and standard deviation of P wave amplitude and PR segment duration is also very high in AF as compared to the other classes. This inferred that the electrical activities in the atrium were more irregular as compared to normal subjects and maybe the atrium is initiated by the AV node instead of the SA node. It is clear that the variations of linear and nonlinear features are much higher than the normal subjects and it may be due to abrupt activities of the heart in AF patients. When compared to other classes, all these differences between the derived properties of AF can be attributed to the irregular atrial contraction and activity that results in the typical electrical transmission to the AV node and ventricles. As a result, each of these aspects significantly aids in model training. The significance of the features was carried out using the ANOVA test (p≤0.01). Most of the extracted features were more significant in AF as compared to other classes.

TABLE 1. FEATURE EXTRACTION FOR DIFFERENT CLASSES

	AF		Normal		Others		Noisy	
Parameter	Average	SD	Average	SD	Average	SD	Average	SD
P-amplitude (mV)	51.511	43.517	61.689	39.668	65.008	61.490	284.321	254.266
Q-amplitude (mV)	-163.408	144.770	-197.878	172.773	-192.356	168.215	-381.757	374.596
R-amplitude (mV)	510.818	227.405	559.438	288.953	538.561	273.954	598.107	492.553
S-amplitude (mV)	-234.593	134.795	-237.371	150.785	-243.467	157.177	-337.399	258.584
T-amplitude (mV)	86.285	61.867	95.887	54.767	98.428	77.235	358.488	288.452
PR-segment (Sec)	0.143	0.026	0.111	0.030	0.125	0.033	0.157	0.022
QRS-segment (Sec)	0.161	0.024	0.169	0.026	0.169	0.028	0.130	0.029
QT-segment (Sec)	0.384	0.042	0.412	0.037	0.405	0.051	0.367	0.035
HR (bpm)	97.203	22.811	75.250	10.310	82.455	19.281	99.231	9.390
RMSSD (msec)	208.911	104.195	92.726	92.353	191.662	144.249	375.237	150.435
MeanNN intervals (msec)	685.180	155.657	833.265	114.559	801.971	168.816	696.011	80.566
SDNN	152.955	77.595	87.074	76.083	142.803	105.201	271.614	99.916
SDSD	211.984	106.164	94.108	93.976	194.495	146.625	380.924	152.525
CVNN	0.220	0.084	0.103	0.088	0.174	0.116	0.385	0.110
CVSD	0.301	0.117	0.110	0.109	0.235	0.169	0.533	0.173
MedianNN	667.748	165.910	851.694	129.665	818.100	197.562	643.551	95.394
pNN50	67.251	18.620	19.799	21.598	38.258	29.710	75.336	14.345
pNN20	81.223	16.048	39.559	26.522	50.486	30.767	84.147	13.290
SD1	149.895	75.069	66.544	66.451	137.528	103.680	269.354	107.851
SD2	152.413	81.986	100.104	87.337	140.865	114.149	269.031	101.893

SD1/SD2	1.079	1.314	0.647	0.358	1.138	1.939	1.018	0.195
HRV_MCVNN	0.187	0.104	0.055	0.067	0.107	0.137	0.334	0.137
HRV_IQRNN	184.684	112.798	80.812	110.331	150.663	185.213	317.138	143.083
HRV_TINN	637.622	328.826	365.572	291.466	588.610	393.944	1190.217	558.078
HRV_CSI	1.049	0.352	2.050	1.269	1.359	1.151	1.025	0.238
HRV_PIP	0.635	0.095	0.537	0.125	0.617	0.121	0.637	0.076
HRV_IALS	0.678	0.098	0.565	0.134	0.655	0.135	0.689	0.081
HRV_PSS	0.891	0.089	0.787	0.171	0.865	0.120	0.905	0.066
HRV_PAS	0.267	0.190	0.165	0.174	0.275	0.222	0.296	0.187
HRV_C1d	0.510	0.076	0.513	0.146	0.547	0.113	0.498	0.067
HRV_C1a	0.490	0.076	0.487	0.146	0.453	0.113	0.502	0.067
HRV_ApEn	0.361	0.196	0.342	0.165	0.384	0.163	0.331	0.181



(a)

(b)

Figure 6. The morphological features of ECG signals (a) peak amplitudes (b) durations

All the features listed in Table 1 are fed to the Random Forest classifier to classify in to four classes. Precision, recall, F1 Score, and accuracy are used to evaluate the classifier's performance. The classifier was trained and tested using 5-fold cross-validation to determine its accuracy. The classifier performance metrics of the four classes are depicted in Table 2. A total of 57 features were extracted from the ECG signals. The redundant and less significant features are removed by using principal component analysis (PCA) and it reduces the computational time from 22.95 seconds to 9.94 seconds. After the dimensionality reduction, we achieved the almost same accuracy as listed in Table 2.

TABLE 2. PRECISION, RECALL, F1-SCORE OF CLASSIFIERS.

Class	Precision	Recall	F1 Score
AF	0.94	0.98	0.96
Normal	0.83	0.87	0.85
Other Classes	0.86	0.76	0.81
Noisy	0.99	1.00	0.99
	Accuracy 0.91		

The plot of accuracy for training and testing of the classifier is shown in Figure 7. We can observe that there is not much difference in accuracy between training and testing and concluded that the model is not over fitted.

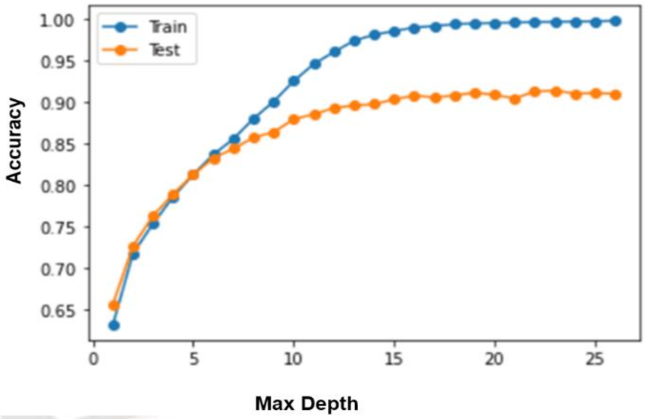


Figure 7. Plot of Accuracy vs. Max depth

VI. CONCLUSION

In current study, four kinds of ECG signals are categorized by a random forest classifier. Using wavelet transformations, amplitudes and corresponding durations of the ECG signals were estimated, and these features were provided as input to the classifier. After running a 5-fold cross-validation, the Random Forest classifier classifies normal, AF, and other rhythms with F1 scores of 0.97, 0.86, and 0.83, respectively, with an accuracy of 0.91. Additional feature extraction and network

***Article Received**: 29 May 2023 **Revised**: 25 July 2023 **Accepted**: 10 August 2023*

selection need to be considered in future research in order to increase accuracy.

REFERENCES

[1] García, Manuel, Juan Ródenas, Raúl Alcaraz, and José J. Rieta. "Atrial fibrillation screening through combined timing features of short single-lead electrocardiograms." In 2017 Computing in Cardiology (CinC), pp. 1-4. IEEE, 2017.

[2] Liu, Yang, Kuanquan Wang, Qince Li, Runnan He, Yong Xia, Zhen Li, Hao Liu, and Henggui Zhang. "Diagnosis of AF based on time and frequency features by using a hierarchical classifier." In 2017 Computing in Cardiology (CinC), pp. 1-4. IEEE, 2017.

[3] Zabihi, Morteza, Ali Bahrami Rad, Aggelos K. Katsaggelos, Serkan Kiranyaz, Susanna Narkilahti, and Moncef Gabbouj. "Detection of atrial fibrillation in ECG hand-held devices using a random forest classifier." In 2017 Computing in Cardiology (CinC), pp. 1-4. IEEE, 2017.

[4] Jiménez-Serrano, Santiago, et al. "Atrial fibrillation detection using feedforward neural networks and automatically extracted signal features." 2017 Computing in Cardiology (CinC). IEEE, 2017.

[5] Yaseen Alkubaisi, A. N. ., Abbas Abood, E. ., & Mohammed, M. H. . (2023). Computational Modelling Applications for the Optimal Design of Prefabricated Industrial Buildings According to the Harmonious Research Method . International Journal of Intelligent Systems and Applications in Engineering, 11(4s), 302–312. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2668

[6] Behar, Joachim A., et al. "Rhythm and quality classification from short ECGs recorded using a mobile device." 2017 Computing in Cardiology (CinC). IEEE, 2017.

[7] Andreotti, Fernando, et al. "Comparing feature-based classifiers and convolutional neural networks to detect arrhythmia from short segments of ECG." 2017 Computing in Cardiology (CinC). IEEE, 2017.

[8] Billeci, Lucia, et al. "Detection of AF and other rhythms using RR variability and ECG spectral measures." 2017 Computing in Cardiology (CinC). IEEE, 2017.

[9] Bin, Guangyu, et al. "Detection of atrial fibrillation using decision tree ensemble." 2017 Computing in Cardiology (CinC). IEEE, 2017.

[10] V. H. C. de Albuquerque, T. M. Nunes, D. R. Pereira, E. J. D. S. Luz, D. Menotti, J. P. Papa, and J. M. R. Tavares, "Robust automated cardiac arrhythmia detection in ECG beat signals". Neural Computing and Applications, pp. 1-15, 2016.

[11] Alickovic, E., & Subasi, A. Medical decision support system for diagnosis of heart arrhythmia using DWT and random forests classifier. Journal of medical systems, 40(4), 108. 2016.

[12] Shadmand, S., & Mashoufi, B. (2016). A new personalized ECG signal classification algorithm using block-based neural network and particle swarm optimization. Biomedical Signal Processing and Control, 25, 12-23.

[13] Mohapatra, S. K., Palo, H. K., & Mohanty, M. N. (2017). Detection of Arrhythmia using Neural Network. Annals of Computer Science and Information Systems, 14, 97-100.

[14] Mustaqeem, A., Anwar, S. M., Majid, M., & Khan, A. R. (2017, July). Wrapper method for feature selection to classify cardiac arrhythmia. In Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE (pp. 3656-3659). IEEE.

[15] Shensheng Xu, S., Mak, M. W., & Cheung, C. C. (2017, July). Deep neural networks versus support vector machines for ECG arrhythmia classification. In Multimedia & Expo Workshops (ICMEW), 2017 IEEE International Conference on(pp. 127-132). IEEE.

[16] Zellar, P. I. . (2021). Business Security Design Improvement Using Digitization. International Journal of New Practices in Management and Engineering, 10(01), 19–21. https://doi.org/10.17762/ijnpme.v10i01.98

[17] Pandey, S. K., & Janghel, R. R. (2019). ECG Arrhythmia Classification Using Artificial Neural Networks. In Proceedings of 2nd International Conference on Communication, Computing and Networking (pp. 645-652). Springer, Singapore.

[18] Kher, Rahul. "Signal processing techniques for removing noise from ECG signals." J. Biomed. Eng. Res 3 (2019): 1-9

[19] Rodrigues, Tiago, Sirisack Samoutphonh, Hugo Silva, and Ana Fred. "A Low-Complexity R-peak Detection Algorithm with Adaptive Thresholding for Wearable Devices." In 2020 25th International Conference on Pattern Recognition (ICPR), pp. 1-8. IEEE, 2021.

[20] Gutiérrez-Rivas, Raquel, Juan Jesus Garcia, William P. Marnane, and Alvaro Hernández. "Novel real-time low-complexity QRS complex detector based on adaptive thresholding." IEEE Sensors Journal 15, no. 10 (2015): 6036-6043.

[21] Shaffer, Fred, and James P. Ginsberg. "An overview of heart rate variability metrics and norms." Frontiers in public health 5 (2017): 258.

[22] Veerabhadrappa, S.T., Vyas, A.L. and Anand, S., 2015. Changes in heart rate variability and pulse wave characteristics during normal pregnancy and postpartum. International Journal of Biomedical Engineering and Technology, 17(2), pp.99-114.

[23] Kung BH, Hu PY, Huang CC, Lee CC, Yao CY, Kuan CH. An Efficient ECG Classification System Using Resource-Saving Architecture and Random Forest. IEEE J Biomed Health Inform. 2021 Jun;25(6):1904-1914. doi: 10.1109/JBHI.2020.3035191. Epub 2021 Jun 3. PMID: 33136548.