Proceedings of the International Conference on Smart Electronics and Communication (ICOSEC 2020)
IEEE Xplore Part Number: CFP20V90-ART; ISBN: 978-1-7281-5461-9

Premature Ventricular Contractions Classification using Machine Learning Approach

Jagdeep Rahul[1*]
[1]Department of Electronics and Communication Engineering, Rajiv Gandhi University
Itanagar, India
E-mail address: jagdeeprahul24@gmail.com

Marpe Sora[2]
[2]Department of Computer Science and Engineering, Rajiv Gandhi University
Itanagar, India
E-mail address: marpe.sora@rgu.ac.in

***Abstract*--In this paper, Premature Ventricular Contractions [PVCs] beat classification is proposed for detecting the ventricular arrhythmia. ECG arrhythmia records are considered from MIT-BIH AD and denoised by using the discrete wavelet transform (DWT). Thereafter, two stage median filter is used to eliminate the baseline wander to obtain the clean and smooth ECG signal. Proposed method has calculated the statistical features of extracted QRS complex of both PVCs and normal beats. KNN and SVM algorithms are used for performance evaluation of the proposed method. Overall SVM algorithm using Gaussian function with kernel scale =0.56 achieved the S_p = 99.71 %, S_e =99.80 %, $+P$ =99.71 % and Acc = 99.75 %. The results obtained have shown that the PVCs classification method is more accurate and reliable, and can be used for automatic classification of arrhythmia.**

Keywords--PVCs; classification; KNN; SVN; machine learning

I. INTRODUCTION

ECG reflects the electrical conduction analogous to the electromechanical activity of the heart. ECG signal consists of a sequence of P, QRS and T-wave, which comprises one cardiac cycle. Each wave's morphology in the cardiac cycle represents its own functionality [1]. Any change in the shape, amplitude and duration of a particular wave is identified as an abnormal ECG. P-wave represents the atrial depolarisation, QRS complex shows the depolarization of the ventricles, and T-wave determines the process of repolarization of the ventricles. QRS complex amplitude is the highest among the P and T-waves peaks [2]. In addition to this, QRS complex features make most widely studied components in the ECG signal analysis. The correct interpretation of the ECG signal can determine the status of the heart. Any abnormalities and arrhythmia present in the heart can be diagnosed with help of the ECG signal analysis [3].

In the cardiovascular system, heart is the main organ, which is responsible for pumping the blood throughout the body, where it is divided into two parts left and right. Each part consist the two chambers atrium and ventricle; atrium collects the blood and ventricle pump the blood throughout the body. An autonomous and centrally controlled nervous system node is situated at the wall of atrium is known as sino-atrial (SA) node, which is responsible for exciting the electrical impulse in the heart. The atrio ventricular (AV) node is situated near to the tricuspid wall between the atrium and ventricles chambers. The AV node collects the electrical impulse generated by the SA node, and propagates it to the bundle of his. From excitation to the propagation of electrical impulse from SA node to bundle of his via AV node generates the ECG signal. Any interruption in this electrical condition creates the abnormalities, and can be seen in the form of change in the morphology of the ECG signal. Such changes in the ECG signal are called as arrhythmias or cardiac arrhythmias [4-5].

Cardiac arrhythmia can be easily detected through the ECG signal analysis. Study on various features of the ECG signal can determine the presence of arrhythmia such as tachycardia, bradycardia, ventricular arrhythmia, and atrial fibrillation. When heart rate crosses the 100 beats per minute is known as tachycardia. Similarly when heart rate below the 60 beats per minute is said bradycardia. Moreover, when arrhythmia arises because of the interruption in the electrical conduction, that developed the abnormal rapid heart rate and prevent the filling of blood in the chambers is called ventricular arrhythmia. The ventricular arrhythmias are classified as ventricular tachycardia, ventricular fibrillation, premature ventricular contraction (PVC), and super ventricular arrhythmia. Whereas atrial fibrillations are also classified as Wolff-Parkinson-White (WPW), atrial flutter, premature atrial contractions (PACs), and heart block. In premature ventricular contraction (PVCs), the electrical impulse starts from the SA node and passing through the AV node. But, it does not propagate to the lower portion to activate the depolarization of the ventricles. PVC beat occurs when ventricles are activated by an ectopic site firing instead of the AV node. Most of the PVCs do not contains P-waves, although it was not activated by the atria. PVC complex is wider, taller, and deeper than the normal. Since, it is propagated through the myocytes of the heart [6-7].

Many methods using machine learning approach have been proposed by the researchers. ECG beat classification using statistical parameter was proposed in [8]. In [9] three level fusion technique was used for feature description and classification. CNN was used for automatic classification of arrhythmia in [10]. QRS complex statistical features are used for classification of Bundle branch block by using the KNN algorithm [11]. In [12], a method of PVCs classification was proposed using the entropy as a feature of the signal. In this work, an automatic classification of premature ventricular contractions beat in the ECG signal is proposed by using the machine learning approach. Two machine learning approach, k-Nearest Neighbors (KNN) and support-vector machine (SVM) are used for PVCs beat classification from the normal beats. This technique is applied on the MIT-BIH arrhythmia database (MIT-BIH-AD) for validation. The process flow of the methodology is shown in figure 1.

II. MATERIALS

ECG records used in this study are taken from the MIT-BIH AD. This database consist the total 48, where each records is of 30minutes duration. All records are samples at the rate of 360 samples per seconds and having the 11 bit resolution over the 10mV range. Records 100m, 105m, 106, 108m, 109m, 111m, 114m, 116m, 118m, 119m, 121m, 123, 124m, 200m, 201m, 202m, 203m, 205m, 207m, 208m, 209m, 210m, 213m, 214m, 215m, 217m, 219m, 221m, 223m, 228m, 230m, 231m, 233m, and 334m consisting the premature ventricular contractions beats have been used in this study.

III. METHODOLOGY

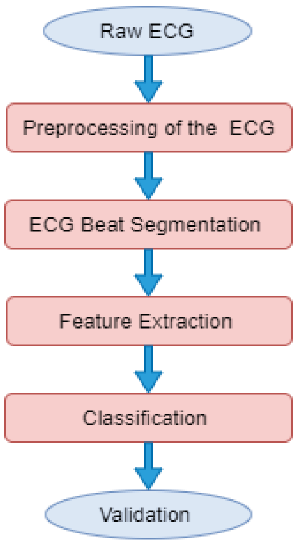

Fig.1 Process flow diagram

A. *Pre-processing of the ECG*

Pre-processing of the ECG signal is required at the initial stage due to presence of various types of noises in the raw ECG signal. The predominant noises which are present in the raw ECG signal can change the main characteristics of the ECG signal. In this work Discrete Wavelet Transform (DWT) is used to remove the high frequency noises such as motion artefacts and powerline interference. Wavelet transform is generally used in variety of compression and prepossessing applications. Wavelet is a mathematical tool, which cut down the signal into different frequency range and analyse the signal in its matching resolution scale [13]. After denoising, two stage median filters are used for baseline drift removal. Two median filters are connected in the cascade nature, where first stage median filter output is connected to the second stage median filter input. Finally output of the two stage median filter is subtracted from the initial input of the median filter. After that, the pre-processed ECG signal is obtained for further analysis [14].

B. *ECG Beat Segmentation*

The main component in the ECG signal is QRS complex, which carries lots of information about the heart status. The shape, duration, and amplitude of the QRS complex is used for determining the PVCs beat. Many methods are available for QRS detection in the literature work [15], [16], [17], and [18]. Since the QRC complex is most studied topic in the ECG signal analysis. The proposed study has detected and located the R-peaks by using the method proposed in [15]. PVCs beats are having the different morphology as compared to the normal QRS complex. The QRS complex detected as PVCs beats are wider, tall, and less pointer. But some aberrant PVCs beats are downward (Inverted), and wider in shape. Normal QRS complex is peaked, narrow shaped, and its duration is almost 100ms, where abnormal QRS complex having the duration is more than the 100ms. A window of 125ms is considered to capture the QRS complex in the ECG signal. The extracted QRS complex from the window in the ECG is marked and shown in the figure 2.

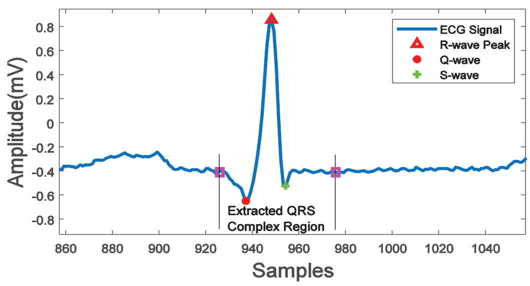

Fig.2. An example of showing extracted QRS complex region

C. *Feature Extraction*

When Premature Ventricular Contractions (PVCs) beat occurs that results change in the morphology of QRS complex. In PVC, electrical conduction through the AV node does not pass to the ventricles for the depolarization that time an ectopic site in the ventricles excite the impulse. In that situation, impulse is propagated through the myocytes of the ventricles. Thus it's widened the QRS complex due to the introduced delay, and increase the depolarisation time of the ventricles. The occurrence of PVCs in the ECG affects the shape of the QRS complex; also change the statistical features such as mean, variance, standard deviation, skewness, and kurtosis. These statistical features are used in the classification of the PVCs beats. The extracted QRS complex from the ECG signal is represented by the y[n]. The statistical features are calculated by using equations from (1) to (5), and are shown below.

$$\mu_y = E[X] = \frac{1}{N}\sum_{1}^{n} y[n] \tag{1}$$

$$\sigma_y^2 = E\left[\left(X-\mu_y\right)^2\right] \tag{2}$$

$$\sigma_y = \sqrt{E\left[\left(X-\mu_y\right)^2\right]} \tag{3}$$

$$s = E\left[\left(\frac{X-\mu_y}{\sigma_y}\right)^3\right] \tag{4}$$

$$k = \frac{E[\left(X-\mu_y\right)^4]}{\left(E\left[\left(X-\mu_y\right)^2\right]\right)^2} \tag{5}$$

Where, μ_y , σ_y , σ_y^2 , s ,and k represent the mean, standard deviation, variance, skewness, and kurtosis of the extracted QRS complex respectively.

The above statistical features are estimated from the selected QRS complex of the ECG signal. These features are applied in to the classifier along with label for further classification purpose.

D. Classifications

Classification is a process which categorizes the data in to class label. In this work classification is performed using the k-nearest neighbors (KNN) algorithm and support vector machine (SVM) to categorize the normal beat and PVCs beat. The KNN is a simple, slow learning, and non-parametric algorithm used for classification. But SVM is flexible and powerful supervised machine learning and can be used for classification. Aim of SVM to classify the data in to the classes to get maximum marginal hyper plane. The KNN has utilized the four different matric to evaluate the performance of the proposed method. But four different kernel functions are used in the SVM classifier to evaluate the performance of the technique. The 10-fold cross validation is used in training and validation for the classifier.

IV. RESULTS AND DISCUSSION

The proposed method for PVCs classification is evaluated using the two machine learning approaches. First the k-nearest neighbors (kNN) algorithm with four different distance metrics namely Cosine, City block, Euclidean, and Mahalnobis are used. Secondly the support vector machine (SVM) algorithm with different kernel functions namely Gaussian, Linear, Quadratic, and Cubic has been used in this classification method for the performance evaluation. The performance of the technique is evaluated using the parameters Specificity (S_p), Sensitivity (S_e), Positive predictivity ($+P$) and Accuracy (Acc). S_e % and S_p % are evaluated using the equation (6) and (7).

$$S_p = \frac{TN}{TN+FP} \times 100 \tag{6}$$

$$S_e = \frac{TP}{TP+FN} \times 100 \tag{7}$$

Where, TP, TN, FP, and FN are the true positive, true negative, false positive and false negative numbers respectively. These numbers are obtained from the classifier confusion matrices. Positive predictivity ($+P$) % and Accuracy (Acc) % are calculated using the equation (8) and (9).

$$+P = \frac{TP}{TP+FP} \times 100 \tag{8}$$

$$Acc = \frac{TP+TN}{TP+TN+FN+FP} \times 100 \tag{9}$$

Total 6114 PVCs effected QRS complex are extracted from the different arrhythmia records and similar number of 6114 normal QRS complex are extracted from the records of the MIT-BIH AD. Total 12228 instances with their statistical features are applied to the selected classifier.

TABLE 1 PROPOSED METHOD PERFORMANCE ON PVCS CLASSIFICATION

Classifier	Matric/kernel function	Category	Normal	PVC	S_e %	S_p %	$+P$ %	Acc %
k-NN	Euclidean	Normal	6097	17	99.72	99.54	99.54	99.63
		PVC	28	6086				
	Mahalnobis	Normal	6099	15	99.75	99.54	99.54	99.65
		PVC	28	6086				
	City block	Normal	6098	16	99.74	99.49	99.49	99.62
		PVC	31	6083				
	Cosine	Normal	6096	18	99.71	99.54	99.54	99.62
		PVC	28	6086				
SVM	Gaussian	Normal	6102	12	99.80	99.71	99.71	99.75
		PVC	18	6096				
	Linear	Normal	6106	8	99.87	99.04	99.04	99.45
		PVC	59	6055				
	Quadratic	Normal	6097	17	99.72	99.53	99.53	99.62
		PVC	29	6085				
	Cubic	Normal	6105	9	99.85	99.62	99.62	99.74
		PVC	23	6091				

Table 1 presents the performance of the proposed method using the KNN and SVM algorithms for classification of the PVCs beat from the normal beat. KNN algorithm utilised the value of k=1 and different distance functions to obtain the better results. SVM algorithm has used the different kernel function depends on the fitting of the available data. The KNN algorithm using Mahalnobis fitting distance function with k=1 has achieved the better results compared to the other distance function in the same algorithm. Whereas SVM obtained the better results using Gaussian function with kernel scale=0.56 compared to the other function in the same algorithm. Overall SVM algorithm using Gaussian function with kernel scale =0.56 achieved the specificity (S_p) = 99.71 %, Sensitivity (S_e) =99.80 % , Positive predictivity ($+P$) =99.71 % and Accuracy (Acc) = 99.75 %.

TABLE 2 PROPOSED METHOD PERFORMANCE COMPARISONS WITH OTHER EXISTING METHODS

Referenc	S_e %	S_p %	Acc %
[19]	91.4	97.7	98.6
[20]	98.8	99.7	99.59
[21]	81	84	86
This wor	99.8	99.7	99.75

Proposed method performance is compared with the other state of art existing methods is presented in the table 2. In [19], proposed a method based on the DNN and utilised the INCART database to obtain the results. The DWT for pre-processing and SVM with ANN were used in [20] to achieve these results. KNN and SVM algorithm are used in [21] to obtain the results. In overall comparison our proposed method for PVCs classification shows the significant improvement over the other existing methods.

V. CONCLUSION

This work presents a method for classification of the PVCs beat from the normal beat. The ECG signal is pre-processed using the DWT and median filter to get better accuracy. QRS complex of the both normal and PVCs beats are extracted from the MIT-BIH AD for feature extraction. Five different features are estimated from the selected QRS complex region. Two machine learning algorithm, namely KNN and SVM are used with different functions for performance evaluation. Overall SVM algorithm using Gaussian function with kernel scale =0.56 achieved the better results compared with the KNN algorithm. Obtained results have shown that PVCs classification method remains more accurate and reliable, and can be used for automatic arrhythmia classification.

REFERENCES

[1] Rangayyan, Rangaraj M. *Biomedical signal analysis*. Vol. 33. John Wiley & Sons, 2015.

[2] Rahul, Jagdeep, Marpe Sora, and Lakhan Dev Sharma. "An overview on biomedical signal analysis." *Int J Recent Technol Eng* 7 (2019): 206-9.

[3] Manlong, Nang Anija, et. al. "ST Segment Analysis for Early Detection of Myocardial Infarction." *International Journal of Computer Sciences and Engineering* 6 (2018): 1500-1504.

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

[5] Coviello, Jessica S. *ECG interpretation made incredibly easy!*. Lippincott Williams & Wilkins, 2020.

[6] Miller, William P. "Cardiac arrhythmias and conduction disturbances in the sleep apnea syndrome: prevalence and significance." *The American journal of medicine* 73.3 (1982): 317-321

[7] Ebrahimzadeh, Ataollah, and Ali Khazaee. "Detection of premature ventricular contractions using MLP neural networks: A comparative study." *Measurement* 43.1 (2010): 103-112.

[8] Celin, S., and K. Vasanth. "ECG signal classification using various machine learning techniques." *Journal of medical systems* 42.12 (2018): 241.

[9] Golrizkhatami, Zahra, and Adnan Acan. "ECG classification using three-level fusion of different feature descriptors." *Expert Systems with Applications* 114 (2018): 54-64.

[10] Zhai, Xiaolong, and Chung Tin. "Automated ECG classification using dual heartbeat coupling based on convolutional neural network." *IEEE Access* 6 (2018): 27465-27472.

[11] Sharma, Lakhan Dev, Ramesh Kumar Sunkaria, and Aman Kumar. "Bundle branch block detection using statistical features of qrs-complex and k-nearest neighbors." *2017 Conference on Information and Communication Technology (CICT)*. IEEE, 2017.

[12] Rizal, Achmad, and Inung Wijayanto. "Classification of Premature Ventricular Contraction based on ECG Signal using Multiorder Renyi Entropy." *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT)*. IEEE, 2019.

[13] Alfaouri, Mikhled, and Khaled Daqrouq. "ECG signal denoising by wavelet transform thresholding." *American Journal of applied sciences* 5.3 (2008): 276-281.

[14] Rahul, Jagdeep, and Marpe Sora. "A novel adaptive window based technique for T wave detection and delineation in the ECG." *Bio-Algorithms and Med-Systems* 16.1 (2020).

[15] Rahul, J., Sora, M. & Sharma, L.D. Exploratory data analysis based efficient QRS-complex detection technique with minimal computational load. *Phys Eng Sci Med* (2020). https://doi.org/10.1007/s13246-020-00906-y.

[16] Sharma, Lakhan Dev, and Ramesh Kumar Sunkaria. "A robust QRS detection using novel pre-processing techniques and kurtosis based enhanced efficiency." *Measurement* 87 (2016): 194-204.

[17] Dohare, Ashok Kumar, Vinod Kumar, and Ritesh Kumar. "An efficient new method for the detection of QRS in electrocardiogram." *Computers & Electrical Engineering* 40.5 (2014): 1717-1730.

[18] Yakut, Önder, and Emine Doğru Bolat. "An improved QRS complex detection method having low computational load." *Biomedical Signal Processing and Control* 42 (2018): 230-241.

[19] Rahhal, M. M. A., Ajlan, N. A., Bazi, Y., Hichri, H. A., & Rabczuk, T. (2018). *Automatic Premature Ventricular Contractions Detection for Multi-Lead Electrocardiogram Signal. 2018 IEEE International Conference on Electro/Information Technology (EIT).* doi:10.1109/eit.2018.8500197.

[20] P. N. Ranaware and R. A. Deshpande, "Detection of arrhythmia based on discrete wavelet transform using artificial neural network and support vector machine," *2016 International Conference on Communication and Signal Processing (ICCSP)*, Melmaruvathur, 2016, pp. 1767-1770, doi: 10.1109/ICCSP.2016.7754470.

[21] Sultana, N., Kamatham, Y., & Kinnara, B. (2016). *Performance Analysis of Artificial Neural Networks for Cardiac Arrhythmia Detection. 2016 IEEE 6th International Conference on Advanced Computing (IACC).* doi:10.1109/iacc.2016.84