

# Morphology-based Detection of Premature Ventricular Contractions

Rohit Hadia, Daniel Guldenring, Dewar D Finlay, Alan Kennedy, Ghalib Janjua, Raymond Bond, James McLaughlin

NIBEC, Ulster University, Jordanstown, United Kingdom

## Abstract

*Premature ventricular contraction (PVC) is the type of ectopic heartbeat, commonly found in the healthy population and is often considered benign. However, they are reported to adversely affect the accuracy of R-R variability based electrocardiographic (ECG) algorithms.*

*This study proposes a Principal Component Analysis (PCA) based algorithmic approach to detect the PVCs based on their morphology. The eigenvectors were derived from signal window around the R-peak, where signal window for the PVC (wPVC) and that of NSR (wNSR) were set to 0.55 seconds and 0.16 seconds respectively. We used 24 ECG recordings from MIT BIH arrhythmia database as training dataset and the remaining 24 ECG recordings as testing dataset.*

*Using the derived eigenvectors and the Linear regression (LR) analysis; complexes corresponding to the wNSR and wPVC were estimated from training and testing datasets. Four different classification methods were employed to differentiate between wPVC and wNSR, namely, Root mean squared error (RMSE), Pearson product-moment correlation coefficient comparison, Histogram probability distribution and k-Nearest Neighbour (KNN). All four methods were implemented individually to classify the wPVC and wNSR.*

*The performance of each of the classification approach was evaluated by computing sensitivity and specificity. With the sensitivity of 93.45% and specificity of 93.14%, KNN based classification method has shown the best performance. The method proposed in this study allows for an effective differentiation between NSR beats and PVC beats.*

## 1. Introduction

Premature ventricular contractions (PVCs) are early heartbeats that originate in the ventricles and are initiated by ectopic foci instead of sinoatrial node [1]. Automaticity in non-nodular tissues leads to the ventricular depolarisation at ectopic foci, resulting in PVCs [2].

PVCs are often considered benign and can be found in up to 4% of the general population. PVCs can be found in

up to 70% of the healthy population, on 24-48 hour Holter recordings [3]. PVC incidents increase with age. Based on occurrence frequency ( $>1000$  PVC/day) and width of the QRS complex of PVC beats ( $>0.16$  sec), PVCs are associated with left ventricular dysfunction and other heart diseases [4]. Additionally, PVCs can lead to unpleasant and in some cases incapacitating symptoms such as palpitations, chest pain, presyncope, syncope, and heart failure [2]. Furthermore, PVCs are also associated with compromising the performance of arrhythmia detection algorithms in ambulatory ECG monitoring applications [5].

In this study, we investigate the development of an algorithm, which will allow the identification of PVCs. The algorithm is based on ECG complex characterisation using principal component analysis (PCA) and data analysis approaches.

## 2. Method

### 2.1. Database

The MIT-BIH arrhythmia database was employed in this study [6]. This dataset consists of 48 ECG recordings each of which is 30 minutes' in duration. For this study, the data was limited to include only QRS complexes that reflected normal sinus rhythm (NSRs) and PVCs. These complexes were identified based on existing annotations provided along with the database. The database was divided into two equal datasets, namely a training dataset and testing dataset, each comprising of 24 ECG recordings. The proposed algorithm was trained using the training dataset, and its performance was measured on the testing dataset.

### 2.2. Proposed algorithm

As shown in Figure 1, the algorithmic approach used in this work was divided into five functional blocks.

- ECG signal pre-processing.
- Extraction of differently shaped ECG waveforms (PVCs and NSRs).
- Principal component analysis (PCA) on

differently shaped ECG waveforms.

- Waveform estimation using derived Principal components (PCs) and Linear regression analysis.
- Comparison between actual and estimated waveforms, and classification based on the results.

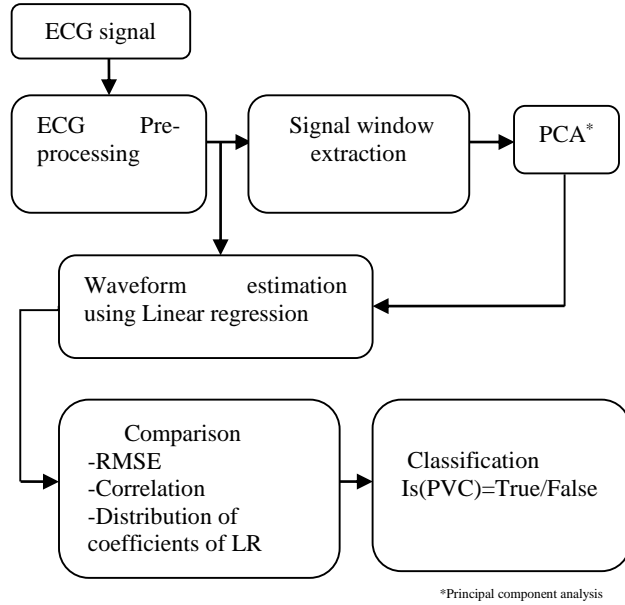


Figure.1 Block diagram PVC detection algorithm

### 2.2.1 ECG pre-processing

In the first functional block, a fourth order Butterworth bandpass filter with poles at 0.5 and 40 Hz is used to remove powerline interference along with other low-frequency noise components. The PVCs and the NSRs are then extracted from each ECG recording based on the provided annotations in the database. Based on the literature review, and our observations, the signal windows of the PVC beats (wPVC) and that of the NSR beats (wNSR) were extracted around the R-peak. The width of wPVC and wNSR was set to 0.55 seconds and 0.16 seconds in duration respectively. However, at the later stage of the research work, wPVC with width 0.16 seconds and wNSR with width 0.55 seconds were also used for the binary classifications.

### 2.2.2 Differently shaped waveform extraction

To reflect the possible variations in PVC waveform morphology and to avoid the biasing due to over representation of any specific PVC shape, a method to differentiate the PVC shapes based on Pearson's correlation coefficient (PCC) value comparison was implemented on each ECG recording of the training

dataset. Similarly, variation in the NSR waveform morphology is extracted using the same approach. Different wPVC morphologies were derived in the following three steps: (1) Linear correlation between the first wPVC of each ECG recording is measured with all the wPVC using PCC. (2) All the wPVC with the PCC values above 0.70 are averaged and stored as differently shaped wPVC in an array dwPVC. (3) wPVC above 0.70 PCC value are then discarded, and the step 1 and step 2 is performed with the remaining wPVC until all the different wPVC morphologies are derived and stored in the dwPVC. Similarly, different wNSR morphologies were derived using the same approach, same steps and stored in an array named dwNSR. Using this method, we have extracted and stored 88 and 40 different morphologies in dwPVC and dwNSR respectively.

### 2.2.3 Principal component analysis on dwPVC and dwNSR

Principal component analysis (PCA) is a statistical method that can be used to reduce the dimensionality of a data set containing correlated variables while retaining the maximum variance in the dataset [7]. Data is transformed into a new set of axes, by an orthogonal linear transformation which are mutually uncorrelated. These new axes are known as Principal components (PCs) or eigenvectors [8]. Each eigenvector contains some new information about the dataset in terms of its variance. However, in most cases, only the first few of these eigenvectors contribute to the maximum variance of the original dataset.

In this work, using PCA, we have computed the eigenvectors and corresponding eigenvalues from dwPVC and dwNSR. Based on the eigenvalues we were able to establish that the first four eigenvectors contributed to 84% of the variance in dwPVC data and 98% of the variance in dwNSR data.

### 2.2.4 Estimation of wPVC and wNSR waveforms

Using multiple linear regression and the derived eigenvectors we estimated the waveforms of wPVCs and wNSRs from the training dataset. The actual and the estimated waveforms of wPVCs and wNSRs are collectively represented as QR<sub>Sa</sub> and QR<sub>Se</sub> respectively.

$$QR_{Se} = \beta_t(0) + \sum_{i=1}^4 (\beta_t(i)Pc(i)) \quad (1)$$

Equation 1 is the mathematical representation of waveform estimation, where independent predictor variable Pc(i) represent the eigenvector corresponding to the maximum variance, the Regression coefficients are

represented by  $\beta_i(i)$ , and  $\beta(0)$  is the error term [10].  $\beta_i(i)$  and  $\beta_i(0)$  are derived by projecting  $Pc(i)$  on the QRSa using ordinary least squares regression as per equation 2

$$\beta_i = (X^T X)^{-1} X^T QRSa \quad (2)$$

where,  $X = \{\text{ones } |i|, Pc|i,4|\}$ ,  $i = \text{length of QRSa}$

### 2.2.5 Discrimination between estimated and actual waveforms.

Four methods were employed to differentiate the QRSa belonging to either the NSR or the PVC groups. A threshold based classification approach was used in the two methods, where, Root Mean Squared Error (RMSE) and Pearson- product-moment correlation was calculated between QRSe and QRSa. The other two methods used cluster based classification approaches, where Histogram Probability Distribution (HPD) and K-Nearest Neighbour (KNN) were computed to differentiate the NSR and PVC based on the distribution of  $\beta_{tr}(1)$  against  $\beta_{tr}(2)$ .

Root mean squared error:

RMSE is the commonly used method to measure the difference between the actual values and the values which are predicted using some mathematical model [9]. In this research work, we calculated RMSE between QRSe and QRSa as per the equation 3.

$$RMSE = \sqrt{\frac{(QRSa - QRSe)^2}{n}} \quad (3)$$

Where  $n = \text{number of samples in QRSa} = \text{number of samples in QRSe}$ . RMSE represents the sample standard deviation between QRSa and QRSe.

Correlation coefficient comparison:

The correlation coefficient is the measure of linear dependence between variables, its value ranges between -1 to 1, where negative value represents the negative linear correlation, and the positive value represents the positive linear correlation between the variables [10].

$$CORR = \frac{\text{Cov}(QRSa, QRSe)}{\sigma_{QRSa} \sigma_{QRSe}} \quad (4)$$

Linear dependency between QRSa and QRSe is measured in this section. Equation 4 is the mathematical representation of the computed correlation coefficient, CORR. Where, CORR is represented as the ratio of covariance of QRSa and QRSe to the product of standard deviations of QRSa,  $\sigma_{QRSa}$ , and QRSe,  $\sigma_{QRSe}$ .

Histogram probability distribution (HPD) of regression coefficients:

Distribution of coefficients of regression based on probability density is used as one of the methods to differentiate between PVCs and NSRs. Regression coefficients were derived using the equation 2 however, unlike in the previous two methods, training dataset was used to derive regression coefficients  $\beta_{tr}(1)$  and  $\beta_{tr}(2)$ . These coefficients were separated into two groups, namely, PVC\_coeff contains the  $\beta_{tr}$  belonging to the PVCs and NSR\_coeff contains the  $\beta_{tr}$  belonging to the NSRs. The probability density functions, Pr\_PVC for PVC\_coeff, and Pr\_NSRe for NSR\_coeff were calculated based on a bivariate histogram with uniform axis lengths, the number of bins, and the bin size. The x and y-axis represent the distribution of  $\beta_{tr}(1)$  with respect to  $\beta_{tr}(2)$ , and z-axis represents the number of PVCs and NSRs in an individual histogram bin. The  $\beta_i$  is used to estimate the probability of a QRSa to be a PVC or an NSR and store them in Pr\_PVCe and Pr\_NSRe as per equation 5 and 6 respectively.

$$Pr\_PVCe = f(Pr\_PVC, \beta_{tr}(1), \beta_{tr}(2)) \quad (5)$$

$$Pr\_NSRe = f(Pr\_NSR, \beta_{tr}(1), \beta_{tr}(2)) \quad (6)$$

Comparing the values of Pr\_PVCe and Pr\_NSRe values, the QRSa is determined to be a PVC if the estimated probability Pr\_PVCe is higher than Pr\_NSRe and vice versa.

k- Nearest Neighbour classification:

K-Nearest Neighbour (KNN) classification method was implemented to classify the QRSa to be a PVC or an NSR. From the training dataset, we have 46,500 observation pairs  $(\beta_{tr}(i), Y_i)$  with the response variable  $Y \in \{0=NSR, 1=PVC\}$  and the predictor vector  $\beta_{tr}(i)$  where  $I=1,2$ . Based on 6 fold cross validation using the training dataset, the value of the number of neighbour k was set to 5, and the Euclidian distance was used as the distance metric [11].

## 3. Results

The performance of the proposed method is measured using QRSa from the testing dataset. The performance is measured as the correctness of each of the four individual approaches mentioned in section 2.2.5, to identify PVC instances from QRSa. Hence, Sensitivity (Sn); the measure of the algorithm to correctly detect a PVC from all the QRSa corresponding to the PVC in testing dataset, and Specificity (Sp); the measure of the algorithm to correctly detect an NSR from all the QRSa corresponding to the NSRs in testing dataset, were calculated. Table 1 shows the derived Sn and Sp values for all the approaches.

Table 1. PVC classification performance

Method	Sensitivity (%)	Specificity (%)
RMSE	42.78	68.51
CORR	70.58	70.20
HPD	75.40	80.75
KNN	93.45	93.14

For the RMSE and the correlation coefficient based classification methods, the receiver operator curve (ROC) was plotted by varying the threshold RMSE and CORR values; the optimal threshold value was selected by computing the Youden's index. The sensitivity and specificity of the HPD and KNN classification model were computed based on the performance of the classification models on the testing dataset. Out of all the four approaches, the KNN classification model has shown the best performance with the 93.45% sensitivity and the 93.14% specificity.

#### 4. Discussion and conclusion

While comparing the performance of each method, one can say that RMSE has negligible adequacy to differentiate between PVCs and NSRs among the proposed morphology analysis methods. Also, differentiation based on correlation coefficient, and probability density values have shown comparatively better performance individually. However, of all the methods used, KNN classification method has shown the best performance to distinguish the PVCs from NSRs.

This study provides the information about the performance of individual methods mentioned in section 2.2.5 to distinguish between PVCs and NSRs based on morphological parameters. However, so far, an approach combining these methods has not been administered, further studies can be carried out to measure the performance to differentiate PVCs and NSRs based on the combination of different methods

#### Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676201.

#### References

[1] F. Del Carpio Munoz, F. F. Syed, A. Noheria, Y. M. Cha, P. A. Friedman, S. C. Hammill, T. M. Munger, K. L. Venkatachalam, W. K. Shen, D. L. Packer, and S. J. Asirvatham, "Characteristics of premature ventricular complexes as correlates of reduced left ventricular

systolic function: Study of the burden, duration, coupling interval, morphology and site of origin of PVCs," *J. Cardiovasc. Electrophysiol.*, vol. 22, no. 7, pp. 791–798, 2011.

[2] S. H. Sheldon, J. J. Gard, and S. J. Asirvatham, "Premature Ventricular Contractions and Non-sustained Ventricular Tachycardia: Association with Sudden Cardiac Death, Risk Stratification, and Management Strategies.," *Indian Pacing Electrophysiol. J.*, vol. 10, no. 8, pp. 357–71, 2010.

[3] G. A. Ng, "Treating patients with ventricular ectopic beats.," *Heart*, vol. 92, no. 11, pp. 1707–1712, 2006.

[4] J. B. Kostis, K. McCrone, a E. Moreyra, S. Gotzoyannis, N. M. Aglitz, N. Natarajan, and P. T. Kuo, "Premature ventricular complexes in the absence of identifiable heart disease.," *Circulation*, vol. 63, no. 6, pp. 1351–1356, 1981.

[5] A. Kennedy, D. D. Finlay, D. Guldenring, R. Bond, and J. Mclaughlin, "The Accuracy of Beat-interval based Algorithms for Detecting Atrial Fibrillation," pp. 893–896, 2015.

[6] S. G. Rockson, D. P. Lorenz, W. Cheong, and K. W. Woodburn, "Current Perspective," *Cardiovasc. Med.*, pp. 591–596, 2000.

[7] F. Castells, P. Laguna, L. Sörnmo, A. Bollmann, and J. M. Roig, "Principal component analysis in ECG signal processing," *EURASIP J. Adv. Signal Process.*, vol. 2007, 2007.

[8] I. T. Jolliffe, "Principal Component Analysis, Second Edition," *Encycl. Stat. Behav. Sci.*, vol. 30, no. 3, p. 487, 2002.

[9] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *Int. J. Forecast.*, vol. 22, no. 4, pp. 679–688, 2006.

[10] R. Fisher, *Statistical methods for research workers*, no. V. 1925.

[11] G. James, D. Witten, T. Hastie, and R. Tibshirani, "An introduction to Statistical Learning," in *Current medicinal chemistry*, vol. 7, no. 10, 2000, pp. 39–42.

Address for correspondence:

Rohit Hadia  
 NIBEC Building, Ulster University, Newtownabbey, Co Antrim,  
 BT37 OQB, United Kingdom  
 r.hadia@ulster.ac.uk