Effect of PCA Feature Reduction on Ventricular Ectopic Beat Classification

Avvaru Srinivasulu, Y Dileep Kumar

Abstract. Due to the cardiac diseases called Ventricular Tachycardia (VT) and Ventricular Fibrillation (VF) causes sudden cardiac death, it is crucial to recognize Ventricular Ectopic Beats (VEB) in Electrocardio- gram for the early diagnosis. There are many algorithms proposed earlier to classify the VEBs. Even though those algorithms achieved good accu- racy, the size of the feature set is large and not precise. In addition, earlier algorithms used feature reduction methods for reducing the feature set. Therefore, in this paper, we extracted only five features namely, Pre RR interval, post RR interval, QRS duration, QR slope, and RS slope. Later, we applied the Principle Component Analysis (PCA) for reducing the size of the feature set to observe the effect of Feature Reduction (FR) on the accuracy of VEB Classification. We applied different classifiers for classifying the cardiac beats in to normal and VEBs. Finally, using K-means Nearest Neighborhood (KNN) classifier and cubic Support Vec- tor Machine (SVM) classifier, we achieved 97.4% classification accuracy, 98.38%, 88.89% & 98.37% sensitivity, specificity & positive predictivity respectively. In addition, it is observed that by applying PCA-FR, the classification accuracy was reduced by a maximum of 3.7%.

Keywords: Electrocardiogram (ECG), ventricular ectopic beats (VEB), Feature Reduction (FR), K-means Nearest Neighborhood (KNN) clas- sifier, Support Vector Machine (SVM) classifier, Principle Component Analysis (PCA).

I. INTRODUCTION

Premature beats originating from ventricular myocardium are called Ventricular Ectopic Beats (VEB) [1]. It was identified as tall and wide beat precede P- wave in ECG [2][3]. The normal and VEB patterns are shown in Figure 1. VEB classification was done previously by extracting many different time domain, frequency domain, statistical and morphological features [4-12].
techniques were applied. But how much degree the FR technique will affect the classification accuracy is not studied. Basically, feature reduction (FR) techniques were used for reducing the feature vector size to reduce the computational time complexity without affecting the actual accuracy.
To find the most significant features among them, FR As the feature vector size is more, we can achieve good accuracy, so that the researchers will always try to get a greater number of features.

But the computational time will be more in such case. Therefore, the feature reduction should be applied in such a way that good accuracy is achieved with smaller feature vector size. In this paper, we have done such an experiment to find the effect of FR on the classification accuracy of VEBs.

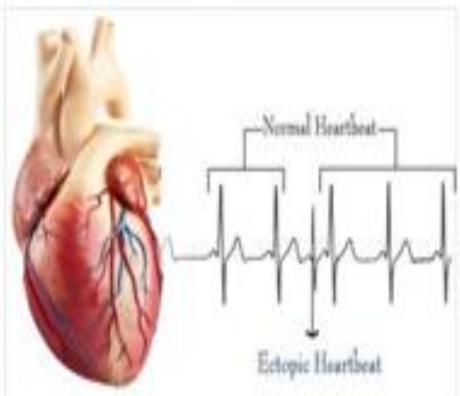

Fig. 1. Normal and Ventricular Ectopic Beats

II. METHODS

1.1 Database

In this paper, ECG data is taken from Physionet: MIT-BIH Arrhythmia database for normal and VEBs. Total 678 normal and 81 VEBs taken from 10 records (105, 109, 118, 119, 200, 202, 210, 214, 221 and 223).

1.2 Processing

Here, we applied, high pass filter with cutoff frequency 0.5Hz to remove the baseline drift present in ECG signal. As the ECG datasets were having only high frequency (HF) noise, we used a low pass filter with cutoff frequency of 30Hz to remove HF noise. Next, the Pan-Tompkins algorithm was applied to detect the R-peaks and difference operation method [13] was applied to find Q and S points on ECG beats.

1.3 Feature extraction

The following features were extracted from each beat of ECG to distinguish the VEBs from normal beats.

Table 1. Features extracted based on beat intervals

Feature	Description
pre_RR_ interval	previous RR interval (R(i) - R(i-1))
post_RR _interval	post (next) RR interval (R(i+1) - R(i))

Revised Manuscript Received on December 15, 2019.

Avvaru Srinivasulu, Department of Electronics and Instrumentation Engineering, GITAM, Bangalore campus, India, sreesri.avvaru@gmail.com.
Y Dileep Kumar, Bio-Signal Research Center, Department of Electronics and Instrumentation Engineering, Sree Vidyanikethan Engineering College, Tirupati, India. y.dileepkumar@yahoo.com

QRS duration	Time is taken for QRS complex
QR_slope	the slope between Q and R points
RS_slope	the slope between R and S points

1.4 Feature Reduction

In this paper, we used Principal component analysis (PCA) features reduction technique [14] for testing the effect of FR on classification accuracy. Principal component analysis method projects a set of points on a smaller dimension sub- space of best fit. It generates a new set of variables, called principal components. Each principal component is a linear combination of the initial variables. All the principal components are orthogonal to every different component, thus there is no redundant information.

1.5 Classifter

To verify the effect of FR on classification accuracy of VEBs, here we have applied 23 different classifiers. The list of classifiers is given in Table 2.

III. RESULTS

The ECG datasets taken from MIT-BIH Arrhythmia database are applied to Pan-Tompkins QRS detection algorithm to find R-peaks. Next, the Q, R, S points are found by a difference operation method. Therefore, the features: Pre- RR interval, post-RR interval, QRS duration, QR slope, and RS slope were cal- culated. The 5 features of 678 normal and 81 VEBs were applied to 23 different classifiers. PCA-FR reduced the dimension of the feature set from 5 to 3 (re- duction ratio [14] is 0.6). The beat detection rate without feature reduction and with PCA-FR are as shown in Table 3. The accuracy, Sensitivity, Specificity and Positive predictivity of all classifiers without and with FR are calculated by equations(1)-(4) and shown in Table 4 respectively.

$$Sensitivity(Se) = \frac{TP}{TP + FN} \quad (1)$$

$$Specificity(Sp) = \frac{TN}{TN + FP} \quad (2)$$

$$PositivePredictivity(Pp) = \frac{TP}{TP + FP} \quad (3)$$

Table 2. List of different classifiers applied in this paper

model	classifier Description
1.1	fine tree A decision tree with many leaves with max 100 fine decisions
1.2	medium tree A decision tree of medium flexibility with fewer leaves and max 20 decisions
1.3	coarse tree A decision tree of few leaves and max 4 decisions
1.4	fine KNN Number of nearest neighborhoods (k) = 1
1.5	medium KNN Number of nearest neighborhoods (k) = 10
1.6	coarse KNN Number of nearest neighborhoods (k) = 100
1.7	cosine KNN the cosine Distance metric is used
1.8	cubic KNN The cubic distance metric is used
1.9	weighted KNN The weighted distance metric is used
2.1	Linear SVM Linear separation using the linear kernel
2.2	Quadratic SVM Linear separation using quadratic kernel
2.3	Cubic SVM Linear separation using cubic kernel
2.4	Fine Gaussian Linear separation using Gaussian kernel with scale sqrt (# classes) / 4
2.5	medium Gaussian SVM Linear separation using Gaussian kernel with scalesqrt (# classes)
2.6	Coarse Gaussian SVM Linear separation using Gaussian kernel with scale sqrt (# classes)4
3.1	Boosted trees A decision tree with AdaBoost algorithm
3.2	Bagged trees A boosted-aggregated ensemble of the fine decision tree
3.3	subspace discriminant Discriminant classifier using random subspace algorithm
3.4	subspace KNN KNN classifier using random subspace algorithm
3.5	RUSBoosted trees A classifier with skewed data
4.1	Linear Discriminant Creates linear boundaries between classes
4.2	Quadratic Discriminant Creates elliptic, parabolic and hyperbolic boundaries
5	Logistic Regression The function of a linear combination of predictors

$$ClassificationAccuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

Retrieval Number: A10221291S52019/2019©BEIESP
DOI: 10.35940/ijeat.A1022.1291S52019

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Table 3. Beats detection

model	classifier	Before FR FN	FP	TN	TP	After FR FN	FP	TN	TP
1.1	fine tree	660	18	15	66	654	24	33	48
1.2	medium tree	660	18	15	66	654	24	33	48
1.3	coarse tree	653	25	25	56	661	17	42	39
1.4	fine KNN	667	11	9	72	654	24	23	58
1.5	medium KNN	661	17	15	66	656	22	19	62
1.6	coarse KNN	678	0	81	0	678	0	81	0
1.7	cosine KNN	652	26	20	61	651	27	24	57
1.8	cubic KNN	660	18	16	65	656	22	18	63
1.9	weighted KNN	665	13	13	68	658	20	20	61
2.1	Linear SVM	678	0	81	0	678	0	81	0
2.2	Quadratic SVM	671	7	13	68	663	15	24	57
2.3	Cubic SVM	667	11	9	72	660	18	22	59
2.4	Fine Gaussian SVM	676	2	36	45	664	14	25	56
2.5	medium Gaussian SVM	666	12	16	65	664	14	26	55
2.6	Coarse Gaussian SVM	677	1	68	13	678	0	80	1
3.1	Boosted trees	668	10	15	66	658	20	26	55
3.2	Bagged trees	662	16	11	70	662	16	24	57
3.3	subspace discriminant	678	0	81	0	678	0	81	0
3.4	subspace KNN	664	14	29	52	662	16	36	45
3.5	RUSBoosted trees	652	26	6	75	638	40	11	70
4.1	Linear Discriminant	673	5	69	12	675	3	70	11
4.2	Quadratic Discriminant	668	10	26	55	661	17	39	42
5	Logistic Regression	678	0	77	4	677	1	77	4

The scatter plot & Receiver Operating Characteristics (ROC) of fine KNN classifier are shown in Figure 2 & 3 respectively. Finally, the effect of PCA-FR on the accuracy of all classifiers are shown in Figure 4.

IV. CONCLUSION

As many classification systems involved the feature reduction concept for reducing the feature vector size after deriving many features from the signal, we tested the effect of feature reduction using PCA and presented in this paper. MIT-BIH Arrhythmia datasets were used for VEB classification by extracting Pre-RR interval, post-RR interval, QRS duration, QR slope, and RS slope. In this paper, the accuracy of VEB classification is tested with 23 different classifiers. It is observed that maximum classification accuracy of 97.4%

Table 4. Classification Performance

model	classifier	Before FR Acc	sp	se	PP	After FR Acc	sp	se	PP
1.1	fine tree	95.7	81.48%	97.35%	97.78%	92.5	59.26%	96.46%	95.20%
1.2	medium tree	95.7	81.48%	97.35%	97.78%	92.5	59.26%	96.46%	95.20%
1.3	coarse tree	93.4	69.14%	96.31%	96.31%	92.2	48.15%	97.49%	94.03%
1.4	fine KNN	97.4	88.89%	98.38%	98.67%	93.8	71.60%	96.46%	96.60%
1.5	medium KNN	95.8	81.48%	97.49%	97.78%	94.6	76.54%	96.76%	97.19%
1.6	coarse KNN	89.3	0.00%	100.00%	89.33%	89.3	0.00%	100.00%	89.33%
1.7	cosine KNN	93.9	75.31%	96.17%	97.02%	93.3	70.37%	96.02%	96.44%
1.8	cubic KNN	95.5	80.25%	97.35%	97.63%	94.7	77.78%	96.76%	97.33%
1.9	weighted KNN	96.6	83.95%	98.08%	98.08%	94.7	75.31%	97.05%	97.05%
2.1	Linear SVM	89.3	0.00%	100.00%	89.33%	89.3	0.00%	100.00%	89.33%
2.2	Quadratic SVM	97.4	83.95%	98.97%	98.10%	94.9	70.37%	97.79%	96.51%
2.3	Cubic SVM	97.4	88.89%	98.38%	98.67%	94.7	72.84%	97.35%	96.77%
2.4	Fine Gaussian SVM	95.0	55.56%	99.71%	94.94%	94.9	69.14%	97.94%	96.37%
2.5	medium Gaussian SVM	96.3	80.25%	98.23%	97.65%	94.7	67.90%	97.94%	96.23%
2.6	Coarse Gaussian SVM	90.9	16.05%	99.85%	90.87%	89.5	1.23%	100.00%	89.45%
3.1	Boosted trees	96.7	81.48%	98.53%	97.80%	93.9	67.90%	97.05%	96.20%
3.2	Bagged trees	96.4	86.42%	97.64%	98.37%	94.7	70.37%	97.64%	96.50%
3.3	subspace discriminant	89.3	00.00%	100.00%	89.33%	89.3	0.00%	100.00%	89.33%
3.4	subspace KNN	94.3	64.20%	97.94%	95.82%	93.1	55.56%	97.64%	94.84%
3.5	RUSBoosted trees	95.8	92.59%	96.17%	99.09%	93.3	86.42%	94.10%	98.31%
4.1	Linear Discriminant	90.3	14.81%	99.26%	90.70%	90.4	13.58%	99.56%	90.60%
4.2	Quadratic Discriminant	95.3	67.90%	98.53%	96.25%	92.6	51.85%	97.49%	94.43%
5	Logistic Regression	89.9	4.94%	100.00%	89.80%	89.7	4.94%	99.85%	89.79%

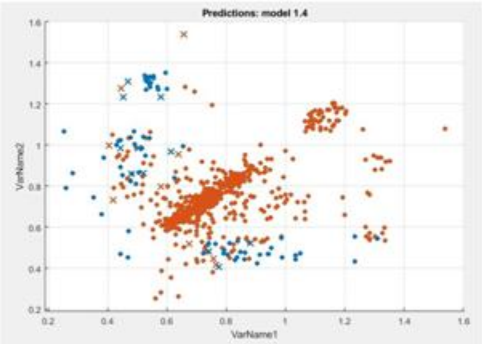

Fig. 2. Scatter Plot for Fine KNN classifier

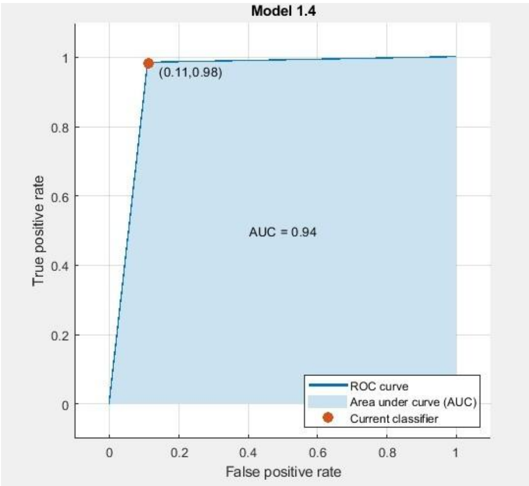

Fig. 3. Receiver Operating Characteristics (ROC) of Fine KNN classifier

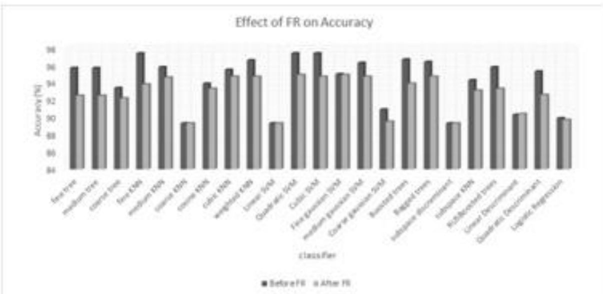

Fig. 4. Effect of PCA-FR on the accuracy of different classifiers

was achieved by fine KNN classifier and the accuracy was reduced by maximum 3.7% after applying PCA-FR. Therefore, it is concluded that the effect of PCA feature reduction on VEB classification accuracy is less significant, but it reduces the computational time when the feature vector size is very large.

REFERENCES

1. Ng, G. André,: Treating patients with ventricular ectopic beats. Heart 92.11, pp 1707-1712 (2006)
2. www.cardionetics.com/ventricular-ectopic-beats
3. James E Keany, et.al.,: Premature Ventricular Contraction. Medscape, Article, 761148, Updated: Jan 13 (2017)
4. [De Chazal, Philip: Detection of supraventricular and ventricular ectopic beats using a single lead ECG. In Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE, IEEE, pp. 45-48 (2013)
5. De Chazal, Philip, et.al.: Automatic classification of heartbeats using ECG mor- phology and heartbeat interval features. IEEE Transactions on Biomedical Engi- neering 51.7, pp. 1196-1206 (2004)
6. Sadiq, Ismail, and Shoab Ahmad Khan.: Heartbeat classification of ECGs us- ing morphology and beat intervals. In Bioinformatics and Biomedical Engineer- ing,(iCBBE) 2011 5th International Conference on, IEEE, pp. 1-5 (2011)
7. Hammed, Norhan S., and Mohamed I. Owis.: Patient adaptable ventricular ar- rhythmia classifier using template matching. In Biomedical Circuits and Systems Conference (BioCAS), IEEE, pp. 1-4 (2015)
8. Chow, H. S., Moody, G. B., & Mark, R. G.: Detection of ventricular ectopic beats using neural networks. In Computers in Cardiology, Proceedings of IEEE, pp.659- 662 (1992)
9. . Wiens, J., and John V. Guttag.: Patient-adaptive ectopic beat classification using active learning. In Computing in Cardiology, IEEE, pp. 109-112(2010)
10. Dotsinsky, Ivan A., and Todor V. Stoyanov.: Ventricular beat detection in single channel electrocardiograms. Biomedical engineering online 3, no. 1, pp. 3 (2004)
11. . Wiens, Jenna, and John V. Guttag.: Patient-specific ventricular beat classifica- tion without patient-specific expert knowledge: A transfer learning approach. In Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE, pp. 5876-5879 (2011)
12. Baali, Hamza, and Mostefa Mesbah.: Ventricular ectopic beats classification using sparse representation and Gini Index. In Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE, pp. 5821-5824 (2015)
13. Yeh Yun-Chi, et.al.: QRS complexes detection for ECG signal: The Difference Operation Method. Computer methods and programs in biomedicine 91.3, pp 245- 254 (2008)
14. Vasan, et.al.: Dimensionality reduction using Principal Component Analysis for network intrusion detection. Perspectives in Science 8, pp 510-512. (2016)