Identification of Arrhythmia Classes Using Machine-Learning Techniques

C. GURUDAS NAYAK^,1, G. SESHIKALA$, USHA DESAI$, SAGAR G. NAYAK#

^Dept. of Instrumentation and Control Engineering, MIT, Manipal University, Manipal, INDIA.
$School of Electronics and Communication Engineering, REVA University, Bengaluru, INDIA.

#Dept.of InformationTechnology ,University of Stuttgart , Germany - 70569

[1]cg.nayak@manipal.edu, cgurudasnayak@yahoo.co.in

Abstract: - The current paper, describes a machine learning-based approach for automatic detection of five classes of ECG arrhythmia beats using Discrete Wavelet Transform (DWT) features. Further, methodology comprises dimensionality reduction using Principal Component Analysis (PCA), *ten*-fold cross-validation and classification using Support Vector Machine (SVM) kernel functions. Using ANOVA significant features are selected and reliability of accuracy is measured by Cohen's kappa statistic. Large dataset of 110,093 heartbeats from 48 records of *MIT–BIH arrhythmia database* recommended by ANSI/AAMI EC57:1998, which are grouped into five classes of arrhythmia beats viz. Non-ectopic (N), Supraventricular ectopic (S), Ventricular ectopic (V), Fusion (F) and Unknown (U) are classified with class specific accuracy of 99.30%, 95.30%, 88.77%, 55.09% and 95.33%, respectively and an overall average accuracy of 97.48%, using SVM quadratic kernel. The developed methodology is an efficient tool, which has intensive applications in early diagnosis and mass screening of cardiac health.

Key-Words: - Analysis of Variance (ANOVA); Discrete Wavelet Transform; Electrocardiogram; Principal Component Analysis; Support Vector Machine.

1 Introduction

An electrical impulse generated in the sinoatrial node controls the rhythm of a heartbeat. Some disorders in the normal sinus rhythm are called as arrhythmia beats. Different arrhythmias can cause different ECG patterns. The arrhythmias such as ventricular as well as atrial fibrillations and flutters are life-threatening and may lead to stroke or sudden cardiac death. In patients suffered previously with a heart attack, the possibilities of arrhythmic beats will be more and also further the high risks of dangerous heart rhythms [1]. Heart disease remains the nation's as well as world's leading cause of death in both urban and rural areas [2]. Coronary heart disease is the most common type of heart disease, killing nearly 380,000 people annually. Risk factors of cardiovascular diseases are equally high for men as well as women [3].

ECG is a noninvasive, most accessible and cost-effective tool used for cardiac health decision making [4]. Visual interpretation of ECG is complicated and time consuming for large dataset and further may lead to inaccuracies due to the misclassification of beats [5]. Also, time-domain features itself cannot provide good discrimination among normal and abnormal classes [6]. These difficulties can be solved using machine intelligent diagnosis systems.

Martis et al. [7] applied the principal components (PCs) of DWT coefficients for arrhythmia diagnosis and achieved average accuracy of 97.23% using SVM. Same group using principal components reported classification accuracy of 98.11% using PCs of ECG [8]. Ventricular premature contraction (VPC) and atrial premature contraction (APC) beats are classified with 98.4% accuracy using higher order spectra (HOS) cumulant features of the wavelet packet decomposition and principal component analysis (PCA) [9]. Atrial flutter, atrial fibrillation and normal beats are classified with 97.65% accuracy using higher order spectra (HOS) bispectrum features and ICA [10]. Melgani and Bazi classified arrhythmia beats using particle swarm optimization (PSO) and SVM approach [11]. Oresko et al. developed wearable smart-phone device for ECG arrhythmia detection [12].

These reported methodologies are verified on smaller ECG dataset. The proposed system detects five classes of ECG arrhythmias for large dataset of 110,093 beats using DWT features and PCA for dimensionality reduction followed by *ten*-fold cross-validation, classification using SVM kernel functions and performance measure using class specific accuracy, overall accuracy and Cohen's kappa statistic.

Paper is organized as follows: Data description and methodology are explained in section II. Obtained results and discussion are presented in section 0. Finally section IV concludes the paper.

2 Materials and Methodology

2.1 Dataset Used

In current study, publicly available PhysioNet [13], *MIT–BIH arrhythmia database* [14] sampled at 360 Hz is used. Further, heartbeats from the entire dataset are categorized into five arrhythmia classes as recommended by ANSI/AAMI EC57:1998 standard [15]. The overall 110,093 heartbeats considered, in which 90575 Non-ectopic (N), 2972 Supraventricular ectopic (S), 7707 Ventricular ectopic (V), 1784 Fusion (F) and 7055 Unknown (U) beats are used for present experimentation. In this work, the methodology is implemented using MATLAB R2013A simulation tool. To implement Cohen's kappa: kappa toolbox designed by Giuseppe Cardillo [16] in MATLAB is connected to the developed programs.

Fig. 1 shows system approach of the proposed system and developed methodology is briefly explained as follows:

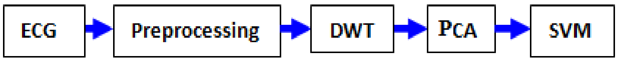

Fig. 1: Proposed system.

2.2 Preprocessing

ECG signals taken from *MIT–BIH arrhythmia database* [14] are denoised using DWT multi-resolution analysis [17] using Daubechies*db4* mother wavelet up to nine levels of sub-band decomposition. The 9th level approximation coefficient of frequency range 0–0.351 Hz, refers to baseline wander [18], are set to zero. Also, ECG does not contain any significant information above 45 Hz, so first two levels of detailed coefficients were set to zero. Remaining wavelet coefficients in the detailed sub-bands of 3rd, 4th, 5th, 6th, 7th, 8th and 9th level are reconstructed to get resultant denoised ECG. Further, denoised ECG R–peak is detected using Pan–Tompkins algorithm [19]. R–peak detected signal is segmented, such that each segment consists of 99 samples before R–peak and 100 samples after R-peak. Each of these 200 samples of cardiac beats of five arrhythmia classes are used in this study. Fig. 2 a) and b) gives typical plot of Normal (N) and Supraventricular (S) classes of ECG beats.

2.3 Discrete Wavelet Transform (DWT)

In the current study, each cardiac beat consisting of 200 samples is decomposed [17] into four sub-bands using Daubechies*db4* mother wavelet. Features are extracted at QRS–complex frequency range [20] from 3rd level detail of 22.25–45 Hz and 4th level detail coefficients of 11.25–22.5 Hz as shown in Fig. 3 and Fig. 4. These two sub-bands are applied independently for dimensionality reduction using PCA.

2.4 Principal Component Analysis (PCA)

PCA is a linear dimensionality reduction method, which identifies patterns in data and expresses by highlighting their similarities and differences [21]. PCA computation involves; finding covariance matrix from ensemble of heartbeats, eigenvalue and eigenvector decomposition of covariance matrix, sorting eigenvectors in the descending order of eigenvalues and finally projecting the original ECG data in the directions of sorted eigenvectors. The eigenvector with the highest eigenvalue is the principle component of the data set. Initial few components will represent the most of the variability present in the data. In this work, first 12 principal components (PCs) of PCA are subjected for pattern classification using SVM kernel functions. In current work, significance of initial 12 PCs is verified using *one*-way Analysis of Variance (ANOVA) test [24].

2.5 Support Vector Machine (SVM)

SVM separates a given set of binary labeled training features with a maximal margin from the hyper-plane. When linear separation is impossible, different kernel transformations can be applied for nonlinear mapping to a higher dimensional feature space [22]. Different kernels namely: quadratic, polynomial and radial basis function (RBF) kernels can be used.

3 Results and Discussion

ECG signals downloaded from *MIT–BIH arrhythmia database* [14] are subjected to 9 level sub-band decomposition [17] using DWT. Based on the reference annotations for QRS middle points as marked in the database, 110,093 ECG beats belonging to five classes (N, S, V, F and U) of arrhythmia are considered in this study. These beats are transformed using DWT based sub-band decomposition, from which 3rd and 4th level detail coefficients are extracted and subjected to PCA. Further, initial 12 principal components (six each) are tested using ANOVA for discrimination. *Ten*-fold cross validation is used during the classifier design and performance is measured using class specific accuracy, overall accuracy and Cohen's kappa statistic.

TABLE I and TABLE II lists the *F*-value and *p*-value statistic of the ANOVA test for 12 principal components (six each from 3rd and 4th level detail sub-bands). TABLE III represents Cohen's kappa statistic, class specific accuracy and overall accuracy using four different SVM kernels.

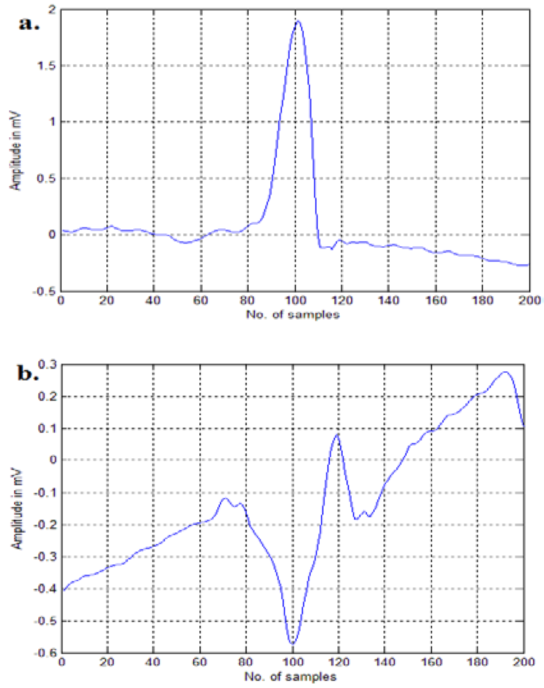

Fig. 2: Typical plot of a) N beat and b) S beat of ECG.

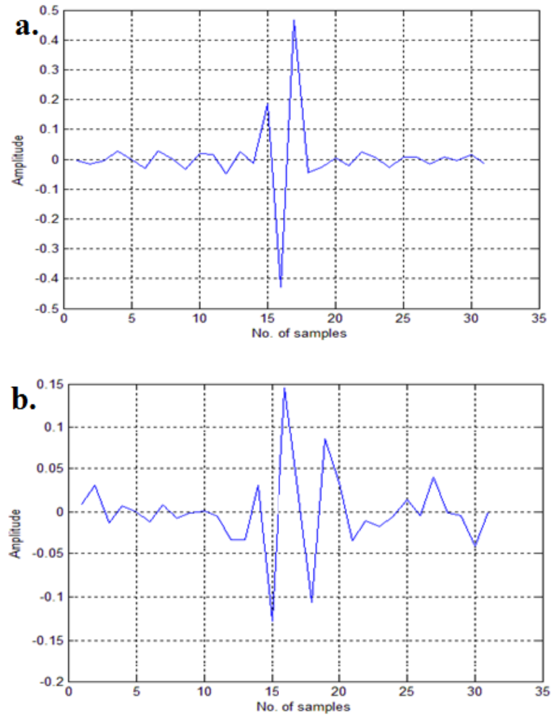

Fig. 3: DWT 3rd level detail coefficients a) N beat and b) S beat.

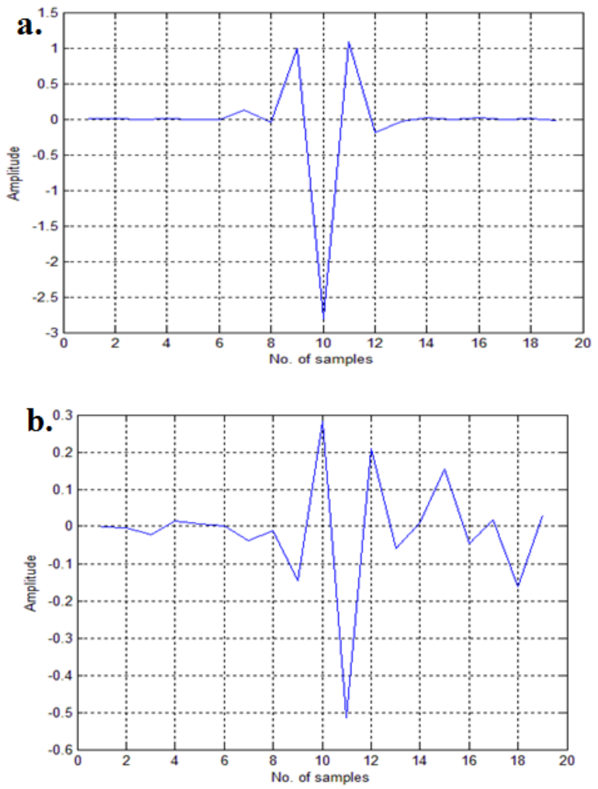

Fig. 4: DWT 4th level detail coefficients a) N beat and b) S beat.

It is observed from TABLE III that the SVM quadratic kernel provide highest average classification accuracy of 97.48% and a highest kappa coefficient of 0.9198. The kappa coefficient indicates the consistency of accuracy over ten-folds. Since we have obtained highest kappa coefficient for the classification accuracies of SVM quadratic kernel, which is consistent than other kernel functions. Hence, it can be further cross verified using box plots of accuracies for *ten*-folds, as shown in Fig. 5. The box width in the box plot accuracy of Fusion (F) beats is more than that of other classes, which indicates that less number of Fusion beats are detected and range of variation is more over *ten*-fold cross-validation. This is due to the skewed data (number of samples in five classes are unequal) considered in this study, which leads over fitting of the classification model. Hence, results in bias towards more data sample class. The line inside the box indicates the median value of *ten*-folds.

Recently, Martis et al. conducted work on large dataset of *MIT–BIH arrhythmia database* [14], using DCT [26] [29] and DWT [27] [28] features. The current paper classifies five classes of beats with overall and class specific accuracy and consistency of classifier is measured using kappa coefficient. The developed system is noninvasive and entirely computerized further can be extended in classification of tachycardia, coronary artery disease, autism, etc. Performance of method can be evaluated using different nonlinear methods and advanced classifiers can be applied in design.

Principal Components	***F*-value**	***p*-value**
PC1	4049.96	<0.0001
PC2	1700.06	<0.0001
PC3	110.71	<0.0001
PC4	926.65	<0.0001
PC5	1472.11	<0.0001
PC6	221.49	<0.0001

TABLE I. SIX PRINCIPAL COMPONENTS EXTRACTED FROM 3RD LEVEL DETAIL COEFFICIENTS

Principal Components	***F*-value**	***p*-value**
PC7	5177.32	<0.0001
PC8	5904.93	<0.0001
PC9	1478.62	<0.0001
PC10	448.47	<0.0001
PC11	472.61	<0.0001
PC12	707.78	<0.0001

TABLE II. SIX PRINCIPAL COMPONENTS EXTRACTED FROM 4TH LEVEL DETAIL COEFFICIENTS

TABLE III. RESULTS OF CLASS SPECIFIC ACCURACY AND COHEN'S KAPPA STATISTIC

SVM Kernel	N (%)	S (%)	V (%)	F (%)	U (%)	Overall (%)	Kappa Statistic
Linear	99.52	2.944	71.89	5.467	94.49	93.13	0.7606
Quadratic	**99.30**	**95.3**	**88.77**	**55.09**	**95.33**	**97.48**	**0.9198**
Polynomial	99.42	94.91	86.10	57.93	93.23	97.29	0.9131
RBF	98.24	85.26	89.56	42.27	99.53	96.46	0.8903

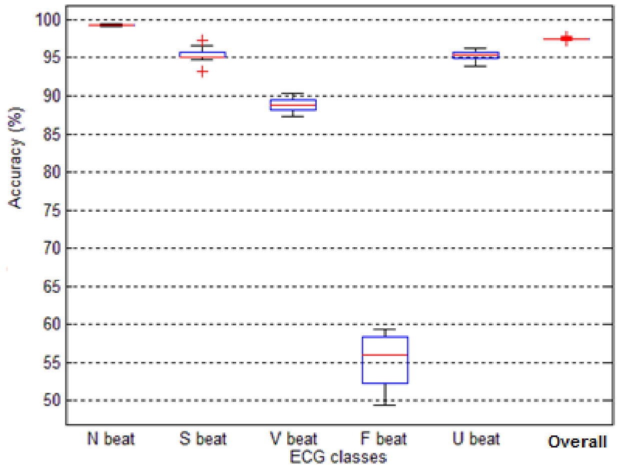

Fig. 5: Class specific and average accuracy of *ten*-folds for PCs of DWT using SVM quadratic kernel.

4 Conclusion

ECG is the electrical activity of heartbeat, which holds the hidden information about various conditions of cardiac health. In this present work, from DWT nonlinear features, five classes of cardiac arrhythmias are detected with good classification accuracy and class specific accuracy. Also, statistically significant features are selected and tested on large dataset using robust classifier SVM and *ten*-fold cross validation performed to avoid biasing in selection of training and testing feature sets. The current paper concludes, machine intelligent diagnosis can help in the objective visualization and extraction of hidden complex features with high discrimination.

References:

[1] D.H. Bennett, "*Bennett's cardiac arrhythmias, practical notes on interpretation and treatment,*" Wiley-Blackwell, 2012.

[2] World Health Organization: "Global status report on noncommunicable diseases," 2014 Update, *World Health Organization*, 2014, pp. 1-298.

[3] The Heart Foundation: Accessed on: 01/06/2015 URL:http://www.theheartfoundation.org/heart-disease-facts/heart-disease-statistics/

[4] R. Vecht, M.A. Gatzoulis, and Nicholas Peters, *ECG diagnosis in clinical practice*, Springer Science & Business Media, 2009.

[5] O. Faust, U.R. Acharya, and T. Tamura, "Formal design methods for reliable computer-aided diagnosis: A review," *IEEE Reviews Biom. Eng.*, 5, 2012, pp. 15-28.

[6] R.J. Martis, C. Chakraborty, and A.K. Ray, "An integrated ECG feature extraction scheme using PCA and wavelet transform," in: *IEEE India Annual Conference (INDICON), 2009*, 2009, pp. 1-4.

[7] R.J. Martis, M.M.R. Krishnan, C. Chakraborty, S. Pal, D. Sarkar, K.M. Mandana, and A.K. Ray, "Automated screening of arrhythmia using wavelet based machine learning techniques, *J. of Med. syst.*, *36* (2), 2010, pp. 677-688.

[8] R.J. Martis, U.R. Acharya, K. Mandana, A.K. Ray, and C. Chakraborty, "Application of principal component analysis to ECG signals for automated diagnosis of cardiac health, *Expert Syst. Appl.*, 39 (14), 2012, pp. 11792–11800.

[9] R.J. Martis, U.R. Acharya, A.K. Ray, and C. Chakraborty, "Application of higher order cumulants to ECG signals for the cardiac health diagnosis, in: *Annual International Conf. of the IEEE EMBC*, pp. 1697–1700, 2011.

[10] H. Prasad, R.J. Martis, U.R. Acharya, C.M. Lim, J.S. Suri, "Application of higher order spectra for accurate delineation of atrial arrhythmia," *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, 2013, pp. 57-60.

[11] F. Melgani, and Y. Bazi, "Classification of electrocardiogram signals with support vector machines and particle swarm optimization," *Inf. Tech. in Biom., IEEE Trans.*, 12 (5), 2008, pp. 667-677,.

[12] J.J. Oresko, Z. Jin, J. Cheng, S. Huang, Y. Sun, H. Duschl, and A. C. Cheng, "A wearable smartphone-based platform for real-time cardiovascular disease detection via

electrocardiogram processing," *Info. Tech. in Biom., IEEE Trans.*, 14 (3), 2010, pp. 734-740.

[13] G.B. Moody, R.G. Mark, and A.L. Goldberger, "PhysioNet: A web-based resource for study of physiologic signals," *IEEE Eng. Med. Biol. Mag.*, 20, 2001, pp. 70-75.

[14] G.B. Moody, R.G. Mark, The impact of the MIT-BIH arrhythmia database, *IEEE Eng. Med. Biol. Mag.* 20, pp. 45–50, 2001.

[15] ANSI/AAMI EC57: Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms (AAMI Recommended Practice/American National Standard), Order Code: EC57-293, 1998. <http://www.aami.org>.

[16] Giuseppe Cardillo: kappa http://www.mathworks.com/ matlabcentral/ fileexchange/15365-cohen-s-kappa.

[17] S.G. Mallet, "A theory for multiresolution signal decomposition: The wavelet representation." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11 (70), 1989, pp. 674-693.

[18] Thakor, Nitish V., and Yi-Sheng Zhu. "Applications of adaptive filtering to ECG analysis: noise cancellation and arrhythmia detection." *Biomedical Engineering, IEEE Transactions on* 38, no. 8, 1991, pp. 785-794.

[19] J. Pan, W.J. Tompkins, A real-time QRS detection algorithm, *IEEE Trans. Biomed. Eng.* BME-32 (3), 1985, pp. 230–236.

[20] B.-U. Kohler, C. Hennig, and R. Orglmeister, "The principles of software QRS detection," *IEEE Eng. Med. Biol. Mag.* 21 (1), 2002, pp. 42–57.

[21] I. Jolliffe, Principal Component Analysis. Springer Series in Statistics, Springer, 2002.

[22] S.S. Haykin, "*Neural networks: A comprehensive foundation*," 2nd ed., Prentice Hall, New Jersey, 1999.

[23] U. Desai, et al., "Decision Support System for Arrhythmia Beats using ECG Signals with DCT, DWT and EMD Methods: A Comparative Study," J. Mech. Med. Biol., 16 (1), 2016, 1640012.

[24] E. R. Girden, "*ANOVA: Repeated measures*," Sage Pub., Inc., 1991.

[25] U. Desai, et al., "Diagnosis of Multiclass Tachycardia Beats using Recurrence Quantification Analysis and Ensemble Classifiers." Journal of Mechanics in Medicine and Biology, 16 (1): 1640005, Feb. 2016. Martis, Roshan Joy, U. Rajendra Acharya, Choo Min Lim, and Jasjit S. Suri. "Characterization of ECG beats from cardiac arrhythmia using discrete cosine transform in PCA framework." *Knowledge-Based Systems* 45, 2013, pp. 76-82.

[26] 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 (5), 2013, pp. 437-448.

[27] R.J. Martis, U.R. Acharya, and Hojjat Adeli, "Current methods in electrocardiogram characterization"**,** *CBM*, 48 (Complete), 2014, pp. 133-149.

[28] U. Desai, et al., "Discrete Cosine Transform Features in Automated Classification of Cardiac Arrhythmia Beats." *Emerging Research in Computing, Information, Communication and Applications*, Springer India, 2015, pp. 153-162.