Automated ECG Delineation using Machine Learning Algorithms

[1]Indu Saini, [2]Dilbag Singh, [1]Arun Khosla

[1]Dept. of Electronics & Communication Engineering and
[2]Dept. of Instrumentation & Control Engineering,
Dr.B.R. Ambedkar National Institute of Technology, Jalandhar, India
Email: drdilbag@gmail.com

Abstract. *The aim of automated electrocardiogram (ECG) delineation system is the reliable detection of fundamental ECG components and from these fundamental measurements, the parameters of diagnostic significance, namely, P-duration, PR-interval, QRS-duration, QT-interval, are to be identified and extracted. In this work, two supervised machine learning algorithms, K-Nearest neighbour (KNN) and Support Vector Machine (SVM) have been applied for accurate and efficient delineation of ECG signals. The algorithms were evaluated on a standard database CSE DS-3. The mean and standard deviations of the basic intervals obtained by KNN and SVM algorithms have been calculated and compared with three 12-lead programs used in the CSE study from the combined program median. The results show that the proposed algorithms give a new direction of using KNN and SVM effectively for the identification and delineation of the ECG wave components.*

Keywords: ECG, KNN, SVM, gradient, classifier, durations, intervals.

1. Introduction

All ECG computer analysis programs are basically composed of two parts. The first part deals with the accurate identification and measurement of characteristic features of ECG signal whereas the second part deals with the diagnostic interpretation. The main task in the measurements is to find exact location of major reference points, i.e., onset and offset points of P, QRS and T-waves. In recent years, many algorithms have been developed for the detection and delineation of the ECG signals. Some of them are as:-

Li et al. in 1995 [1] proposed an algorithm based on Wavelet Transform (WT) for detecting ECG characteristic points. In [2] the authors used quadratic spline wavelet as prototype wavelet for detection of ECG significant points and the first four scales of the DyWT were analyzed. Martinez et al. in [3] have presented and validated a wavelet based ECG delineation system which performed QRS detection and provided the locations of the peak(s) of P, Q, R, S and T-waves, and their boundaries. Goutas et al. in 2005 [4] presented a new algorithm based on digital fractional order differentiation for P and T-waves detection and delineation. Mehta and Lingayat in [5] presented the application of SVM for QRS detection in single and 12-lead ECG using entropy and combined entropy criterion. Arzeno et al. in 2008 [6] analyzed traditional first-derivative based squaring function and Hilbert transform-based methods for QRS detection and proposed their modifications with improved detection thresholds. In [7], the authors presented an efficient method for detection of P and T-waves in 12-lead ECG using SVM. The authors in 2010 [8] proposed Bayesian inference to represent a priori relationships among ECG wave components.

This paper presents an application of KNN and SVM for delineation of ECG signal. The KNN method is an instance based learning method that stores all available data points (examples) and classifies new data points based on similarity measure [9].The goal of SVM is to produce a model which predicts target value of data instances in the testing set where only the attributes are given [10]. The SVM classification is based on supervised learning, where known labels help indicate whether the system is performing in a right way or not.

This paper is organized as follows: section II presents a brief outline of KNN and SVM-classifiers. Section III explains the methodology adopted for delineation. Section IV presents the results and Section V draws the conclusion.

2. Outline of KNN and SVM algorithms

KNN Algorithm
KNN is a non-parametric algorithm that does not make any assumptions on the underlying data distribution. It is also a lazy algorithm means it does not use the training data points to do any generalization. More exactly, all the training data is needed during the testing phase. One of the advantages of the KNN method in classifying the objects is that it requires only few parameters to tune: K and the distance metric, for achieving sufficiently high classification accuracy [9].

SVM Algorithm
SVM is an algorithm of machine learning introduced by Vapnik based on the structural risk minimization principle from statistical learning [11]. SVM is a method for finding a hyperplane in high dimensional space that separates training samples of each class which maximizing the minimum distance between that hyperplane and the training samples. A decision plane is one that separates between sets of objects having different class memberships [10]. The solution gives rise to the decision function of the form.

$$f(x) = \operatorname{sgn}\left[\sum_{i=1}^{l} y_i \alpha_i (x.x_i) + b\right] \qquad (1)$$

where α_i are Lagrange multipliers.

The accuracy of an SVM model is largely dependent on the selection of the kernel method applied. The input data are mapped to a higher dimensional space that a separating hyperplane is constructed to maximize the margin. There are number of kernel functions results in different kinds of SVMs with different performance levels. These include linear, polynomial, radial basis functions [10].

3. Methodology

In this section, we describe the proposed algorithm for the detection of P-wave, QRS-complex, T-wave, their onsets, offsets points, their durations and intervals in ECG signal using KNN and SVM classifiers.

Data Acquisition
The multi-lead ECG records have been acquired from CSE data libraries. All the data sets have been sampled at 500Hz. This dataset consists of ECG records of 125 patients whose record length is 10 seconds for each lead. The CSE committee has published measurement results of 25 records out of 125, hence proposed algorithm has been tested only for 25 records.

Method
Here, we describe the steps followed for developing the delineation algorithm;

In order to attenuate the noise in a raw ECG signal, the acquired signal is passed through a band-pass filter which is composed of cascaded high-pass and low-pass filters. The band-pass filter reduces the influence of muscle noise, baseline wander, 50 Hz and T-wave interferences [12].

For the feature extraction of filtered signal, the gradient of the signal is used as feature vector. Slope or gradient of the signal at every sampling instant is calculated. These slope values are then squared for normalization.

After obtaining the feature vector, the KNN and SVM classifiers are trained. In this phase a training matrix is formed, consisting of m training instances of n features. The number of training instances (m) is equal to the number of samples of selected portions of ECGs and the value of n is the normalized gradient value of each lead of the ECG at a training instance. If the training instance belongs to QRS region, the training label vector is set to 1 and if it belongs to non-QRS region it is set to -1.

After training of KNN and SVM, 25 number records of the CSE are tested for the detection of the QRS-complex. Here, testing labels are unknown; hence any random value can be used.

After testing, a train of 1's is obtained at the output of KNN and SVM classifiers. Then this train of 1's is picked and by using their duration, average pulse duration of 1's is evaluated. Those trains of 1's, whose duration turns out to be more than the average pulse duration are detected as QRS-complex and the other are discarded. The starting and end instances of QRS-complex are depicted as QRS-onset and QRS-offset points. QRS-complex duration has been calculated by using their onset and offset points.

The obtained QRS-complexes are removed from ECG signal by replacing them by baseline. The slope at every sampling instant of QRS less ECG signal is calculated and normalized. The KNN and SVM are again trained and tested for T-wave as explained above. The starting and end instances of T-wave are depicted as T-on and T-end points. QT-interval has been calculated from the onset point of QRS-complex and offset point of T-wave.

The obtained T-waves are removed from QRS less ECG signal by replacing them by baseline. The slope at every sampling instant of ECG signal without QRS and T-wave is calculated and normalized. The KNN and SVM are again trained and tested for P-wave as explained above. The starting and end instances of P-wave are depicted as P-onset and P-offset points. P wave duration is calculated by using onset and offset point of P-wave. PR-interval has also been calculated from onset of P-wave and onset of QRS-complex.

Mean and standard deviation of P-duration, PR-interval, QRS-duration and QT-interval for 25 records of CSE database has also been computed using KNN and SVM algorithms.

4. Results and Discussions

The quantitative values of the P-on, P-peak, P-off, QRS-on, R-peak, QRS-off, T-peak and T-end using KNN and SVM algorithms for 25 records of CSE database are computed and compared with referee results. The most of the values are well within the tolerance limits suggested by the CSE working party and the overall accuracy in the measurement of five wave fiducials is about 92.8 % and 94.4% for KNN and SVM algorithms respectively [13,14].

Based upon the onsets and offsets of P, QRS and T-wave, the important diagnostic parameters P-duration, PR-interval, QRS-duration and QT-interval have been calculated. The mean and standard deviations of the basic intervals obtained by KNN and SVM algorithms and three 12-lead programs used in the CSE study [15] from the combined program median is given in Table I.

Table 1. Comparison of Mean and Standard deviation of ECG wave intervals using KNN and SVM algorithms on CSE database.

Method	P-duration m±std(ms)	PR-interval m±std(ms)	QRS-duration m±std(ms)	QT-interval m±std(ms)
KNN based delineator	-2.5±9.3	5.4±9.9	-2.3±7.2	5.0±13.4
SVM based delineator	-0.1±7.9	2.3±7.1	-1.1±7.0	3.9±11.8
CSE Prog. 2 Marquette)	-12.0±17.6	-8.7±12.1	-0.8±7.2	6.2±15.4
CSE Prog. 11 Glasgow)	-2.4±12.6	5.1±12.5	-1.4±7.1	3.9±14.8
CSE Prog. 13 (Padova)	2.8±10.1	-2.8±8.3	1.8±7.3	-1.1±9.2

5. Conclusion

KNN and SVM based ECG delineators has been presented here for ECG delineation. Both the algorithms of delineation may be capable of enhancing specific rhythms in ECG signals, which may in turn, proves helpful in accurately detecting the P, QRS and T-wave components. These algorithms have been validated on 25 records of CSE DS-3 database. Mean and standard deviation has also been calculated and compared with the results of referees and tolerance limits accepted by cardiologists for CSE.

References

[1] Li C, Zheng C, and Tai C. Detection of ECG characteristic points using wavelet transforms. *IEEE Transactions on Biomedical Engineering*, 42 (1): 21–28, 1995.

[2] Martinez J P, Olmos S and Laguna P. Evaluation of a wavelet-based ECG waveform detector on the QT database. *Computers in Cardiology*, 27: 81-84, 2000.

[3] Martinez J P, Almeida R S, Olmos S, Rocha A P and Laguna P. A Wavelet-based ECG delineator: Evaluation on standard databases. *IEEE Transactions on Biomedical Engineering*, 51 (4): 570–581, 2004.

[4] Goutas A, Ferdi Y, Herbeuval J P, Boudraa M, Boucheham B. Digital fractional order differentiation-based algorithm for P and T-waves detection and delineation. *ITBM-RBM*, 26: 127–132, 2005.

[5] Mehta S S and Lingayat N S. Comparative study of QRS detection in single lead and 12-lead ECG based on entropy and combined entropy criteria using support vector machine. *Journal of theoretical and applied information technology*, 3(2): 8-18, 2007.

[6] Arzeno N M, Deng Z D, and Poon C S. Analysis of first-derivative based QRS detection algorithms. *IEEE Transactions on Biomedical Engineering*, 55(2): 478-484, 2008.

[7] Mehta S S and Lingayat N S. Application of support vector machine for the detection of P- and T-waves in 12-lead electrocardiogram. *Computer methods and programs in Biomedicine*, 93 (1): 46-60, 2009.

[8] Lin C, Mailhes C, and Tourneret J Y. P- and T-wave delineation in ECG signals using a bayesian approach and a partially collapsed gibbs sampler. *IEEE Transactions on Biomedical Engineering*, 57(12): 2840-2849, 2010.

[9] Thirumuruganathan S. A detailed introduction to K-Nearest Neighbor algorithm. 2010.

[10] Vikramaditya J. Tutorial on Support Vector Machine. School of EECS, Washington State University, Pullman, 2010.

[11] Vapnik V. The nature of statistical learning theory. New York: Wiley, 1988.

[12] Pan J and Tompkins W J. A real time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering*, 32: 230-236, 1985.

[13] Saini I, Singh D, Khosla A. P- and T-wave Delineation of ECG signal using Support Vector Machine. *IETE Journal of Research*, 59(5): 615-623, 2013.

[14] Saini I, Singh D, Khosla A. K-Nearest Neighbor-based algorithm for P and T-waves detection and delineation. *Journal of Medical Engineering and Technology*, 38 (3): 115-124, 2014.

[15] Willems JL. The CSE Multilead atlas manual-measurement results dataset-3. Printed in Belgium by ACCO, Brusselsestraat-118a, 3000 Leuven, Belgium, 1988.