❒ 6598

Feature extraction of electrocardiogram signal using machine learning classification

Sumanta Kuila[1], Namrata Dhanda[2], Subhankar Joardar[3]
[1,3]Department of Computer Science and Engineering, Haldia Institute of Technology, India
[2]Department of Computer Science and Engineering, Amity University, India

Article Info	**ABSTRACT**

In the various field of life person identification is an essential and important task. This helps for the investigation of criminal activities and used in various type of forensic applications like surveillance. For biometric recognition iris, face, voice and fingerprint have a limited fabrication and from there the exact decision regarding liveliness of the subject can be drawn. The aim of the approach is to construct a biometric recognition system based on ECG which processes the raw ECG signal. The entire process is supported by different filters for noise elimination and ECG characteristics waves gone through time domain analysis. Based on the analysis an efficient feature extraction model is developed where several best P-QRS-T signal parts are taken and the positions of the fragmented signals are normalized depends on the priorities of their positions. The calculation of domain features done 72 times. It checks the data sets (train and test) and from feature vector matching to each of the individual signal, separately. The performance and utility of the system are analyzed and feature vectors are examined by different classification algorithms of machine learning. The leading algorithms like K-nearest neighbor, artificial neural network and support vector machine are used to classify different features of ECG, and it is tested using standard cardiac database i.e. the MIT-BIH ECG -ID database.

Corresponding Author:

Sumanta Kuila,
Department of Computer Science and Engineering,
Haldia Institute of Technology,
Haldia, Midnapure (E), West Bengal, 721657, India.
Email: sumanta.kuila@gmail.com

1. INTRODUCTION

The ECG Signal is used as reliable and efficient document for medical applications. ECG is the biometric tool used for long period of time in different cardiac disease identification. The geometric and physiological characteristics of heart produces different signals and maintain uniqueness for different subjects [1]. The ECG signals of every individual person produce unique patterns and characteristics where the heart condition of a patient is decided by the outline of the waveform where function and rhythm of the heart produces effective information for the detection of heart disease [2, 3]. Each of the cardiac cycle of the ECG signal generates the characteristic wave of P-QRS-T. The features of the signal are categorized as Fiducial and non-Fiducial. For the study of ECG, the major uniqueness in the wave is finding of R-peak. So theare identified by making the location of R-peak as a standard [4]. The Figure 1 describes the standard ECG signal where different features are explained based on amplitude, distance, time, angle slope and some related features which supports fiducial based feature representation.For the design of biometric template fiducial based approach is important because it implements the difference between amplitude or temporal for successive fiducial points [5, 6].
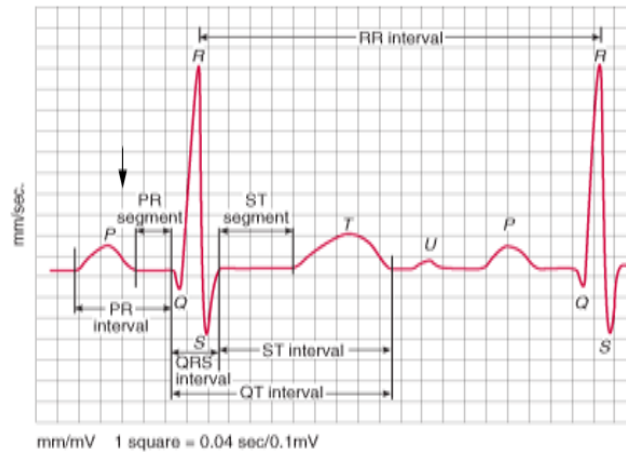
Figure 1. Standard ECG signal

2. FEATURE EXTRACTION ON ECG DATA

Several steps are followed for the feature extraction of ECG signal. The methodology that will help feature extraction and classification of ECG data will follow several procedures like pre-processing of data, feature extraction in logical and effective way and from that features the classification of ECG data. For ECG signal detection and classification initially the ECG data should be collected for preprocessing. For classification of ECG data P-QRS-T detection and feature extraction is required. In Figure 2, the whole process is described pictorially [7]. The steps are used for preprocessing and cleaning of data.
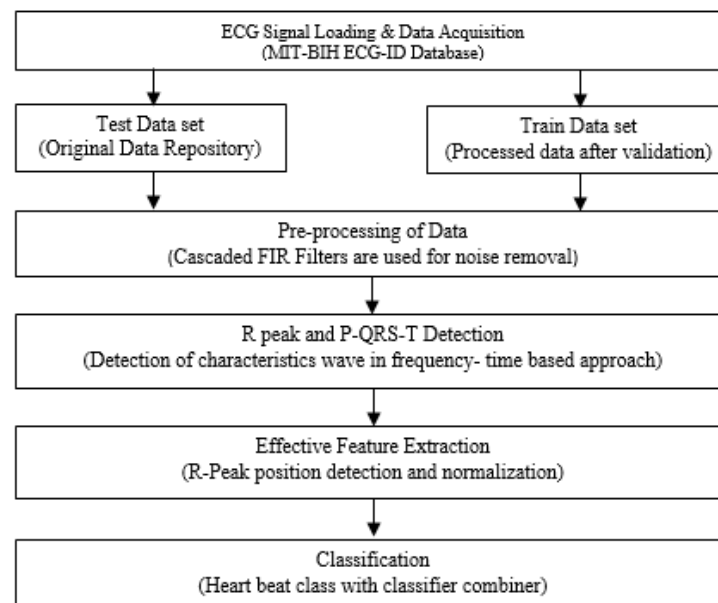

Figure 2. ECG feature extraction and classification system

2.1. ECG signal loading and data acquisition

The ECG data is available Phsiobank which contains individual patient's ECG recording and it is more than 90 thousand recordings. In this paper the raw ECG data that is used for feature extraction and classification is MIT-BIH Arrhythmia database which is already available in the PhysioNet [8]. The MIT-BIH Arrhythmia Database was created between 1975 and 1979 and it become operational in 1980. It contains large number of ECG recordings from different people (47 patients) both male and female (25 males and 22 females) of different age group (between 23 and 89 years) with various cardiac condition.

The ECG recording contains beat-by-beat annotations and it contains more than 200 hours recording which is publicly available in PhysioNet for research and commercial purpose [9, 10].

2.2. Pre-processing of data

The characteristics of raw ECG is that it is noisy and it hold deformation of different signal source. Interferences are generated by various frequency components while the process of acquisition done on the ECG signal recording procedure. The additional noise comes to the ECG signal from that interference and that unwanted data in the signal might fiddle the original data upon the ECG signal and brings the false result [11, 12].

2.3. Detection of wave characteristics (wavelet transform approach)

Different characteristics (P, Q, R, S and T) of the wave which are working on frequency and time based approach are analyzed here. The method primarily works with detection of R-peak using the method of wavelet decomposition. To obtain frequency and time information wavelet transform is used which delivers simultaneous and high-resolution wave signal. The operation done by compressing and dilating the wave signal. The high frequency components are extracted by fine wavelet and components containing lower frequency works with a stretched wavelet [13]. The wavelet transform is worked with the signal s(t) with the groups of time frequency component Ψ(t). That generates a coefficient set C(p,q) which can be implemented by :

$$C\text{(p,q)} = \frac{1}{\sqrt{p}} \int_{-\alpha}^{+\alpha} s(t)\, \psi^* \left(\frac{t-q}{p}\right) dt\; \frac{1}{\sqrt{p}} \int_{\alpha}^{\alpha} s(t)\, \psi^* \left(\frac{t-q}{p}\right) dt \tag{1}$$

where q represents the translation parameter of time location and p is the scale factor. The scale factor is represented as inversely proportional to the working frequency. The complex conjugate is denoted by * and the mother wavelet which works as analyzing wavelet is denoted by Ψ(t). The time scale illustration of digital signal is represented by discrete wavelet transform and in modern techniques of digital filtering are used to represent it. The discrete wavelet transform (DWT) capitulates wavelet transform of fast computation. The characteristics of DWT is that it is relatively easy to implement where different technique of translation is used to minimize the computation time. After final calculation the DWT can be represented as

$$C_{xy} = \frac{1}{\sqrt{p}} \int_{\alpha}^{\alpha} s(t)\, \psi_{xy}(t)\, dt \tag{2}$$

where the translated wavelet in dyadic scale can be describes as

$$\psi_{xy}(t) = 2^{-\frac{x}{2}} \psi \left(2^{-x} t - y\right) \tag{3}$$

The popular orthogonal properties of wavelet like Symlets, Daubechies, Discreate Mayer and Coiflets are used for the reconstruction of wave.

2.4. Analysis of time domain

For the detection of R-Peak frequency domain approach is used and for other characteristics of the wave time domain approach is used.

To calculate R-R interval the formula used $T_{\text{R-R}}(m) = \frac{Rlocation(m+1) - Rlocation(m)}{f_s}(\text{sec})$

where f_s represents the sampling frequency and Rlocation = location of R – peak.

The R-R interval between 65% and 95% is the time gap limit created by the window specifying time domain for recognizing P-wave. The window represents the maximum value for P-wave. Here the P wave is represented as,

u=Rlocation(j)-floor(0.2×f_s)
v= Rlocation(j)-floor(0.07×f_s)

where u and v are the window ranges for time domain.

To identify Q-wave the maximum value of time-based window is chosen and it taken 20ms on average to find the related R-peak in ideal case:

So for Q-wave   u=Rlocation(j)-floor(0.069×f_s)
                              v= Rlocation(j)

In the same way S-peak is recognized by choosing the minimum value of time-based window immediately after R-peak.

So for the S-wave   u= Rlocation(j)
                              v= Rlocation(j)-floor(0.07×f_s)

To identify the T-wave a window is created in time domain where the gap limits of R-R interval lies between 15% to 55% .

So for T wave   u = Rlocation(j)-floor(0.08×f_s)
                              v = Rlocation(j)-floor(0.4×f_s)

The windows from time domain are adaptive as they entirely depend on the values of R-R interval [14, 15]. Figure 3 shows the spotted peaks. Here the PQRST fragments are followed where the finest components are chosen. The minimum distance from the mean is checked with the Euclidian distance of the individual peaks. The best fragments which maintains these homogenous criteria are eligible for data extraction [16].
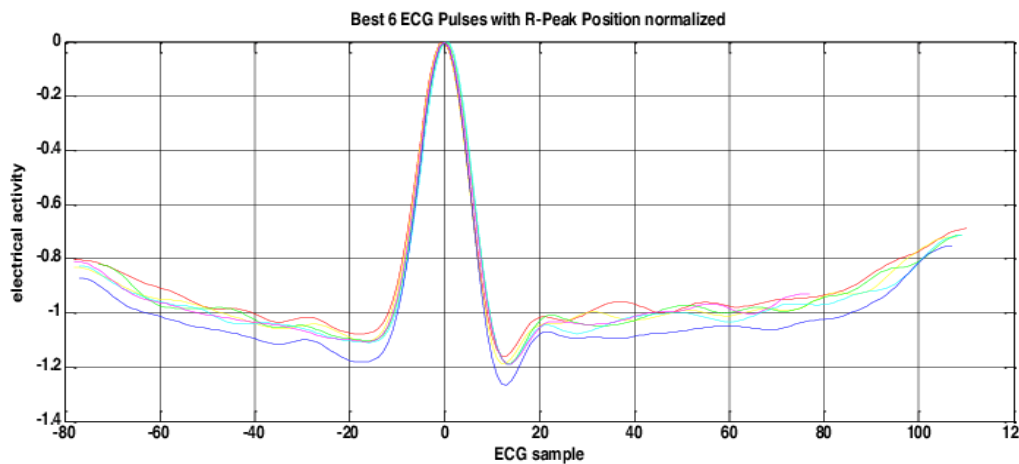

Figure 3. Spotted peaks of PQRST fragments

2.5. Efficient feature extraction

The features are extracted from the data set generated by time and amplitude of the ECG Signal. The nature of the feature describes the cardiac condition of a patient. The feature describes after the peaks are detected, i.e. P, Q, R, S also with QRS, ST and QT intervals have to check when the intervals have to be calculated. In the Figure 3 all PQRST fragments are studied and among them best six excellent fragments are chosen. The selection is based on the variation of the peak and from there lower and better fragments are taken for further analysis. The selected best six PQRST fragment of the electrocardiogram signal are divided into standalone pulses where the information of individual pulses can be compared with each other. Every individual pulse is splitted in the range of 'P' peak to 'T' peak. Figure 3 describes the PQRST beat and graphical representation of the of the data from where the comparison among the signal can be done. The data extracted from the pulses and before that all the pulses should meet a common point to make the position normalized for all the beat. Now comparing the signals, the features of the data can be analyzed. The time duration and amplitudes are calculated as we take the R-peak as the origin of the analysis [17, 18].

The Table 1 describes the various features of the ECG signal. In the work total 72 features are extracted from separate 7 pulses where it uses best six PQRST fragments. The following issues like time, amplitude, angle, slope, gap between time and gap between amplitude is to be discussed.

Table 1. Total set of features (timeandamplitude), 'R' as origin is taken

Feature	Description of Feature	Feature	Description of Feature
Time	1. Time difference of P w.r.t R(origin)	Distance	1. The distance between P and Q
	2. Time difference of Q w.r.t. R		2. The distance between Q and R
	3. Time difference of S w.r.t. R		3. The distance between R and S
	4. Time difference of T w.r.t. R		4. The distance between S and T
	5. Difference of time between P and Q		5. The distance between P and R
	6. Difference of time between P and S		6. The distance between Q and S
	7. Difference of time between P and T		7. STdist : Qsdistr
	8. Difference of time between Q and S		8. RSdist : QRdist
	9. Difference of time between Q and T		
	10. Difference of time between S and T		
	11. PTtime : QStime		
	12. QTtime : QStime		
Amplitude	1. Amplitude difference of P w.r.t. R	Slope	1. Slope of line PQ
	2. Amplitude difference of Q w.r.t. R		2. Slope of line QR
	3. Amplitude difference of Sw.r.t. R		3. Slope of line RS
	4. Amplitude difference of T w.r.t. R		4. Slope of line ST
	5. Difference of amplitude between P and Q		5. Slope of line PS
	6. Difference of amplitude between Q and R		6. Slope of line QS
	7. Difference of amplitude between R and S		7. Slope of line PT
	8. Difference of amplitude between S and T		8. Slope of line QT
	9. Difference of amplitude between P and S		9. Slope of line PR
	10. Difference of amplitude between Q and S	Angle	1. $\angle PQR$
	11. Difference of amplitude between P and T		2. $\angle QRS$
	12. Difference of amplitude between Q and T		3. $\angle RST$
	13. STamp : QSamp		4. $\angle RQS$
	14. RSamp : QRamp		5. $\angle RSQ$
	15. PQamp : QSamp		6. $\angle RTS$
	16. PQamp : QTamp	Miscellaneous	$\triangle QRS$
	17. PQamp : PSamp		$\triangle QRS : RSamp^2$
	18. PQamp : QRamp		$\angle QRS : \angle RST$
	19. PQamp : RSamp		$\angle QRS : QStime$
	20. RSamp : QSamp		$\angle RST : QTtime$
	21. RSamp : QTamp		$\angle RST : PQdist$
	22. STamp : PQamp		$\angle QRS : \angle RQS$
	23. STamp : QTamp		$\triangle QRS : QRamp$
			Perimeter of $\triangle QRS$
			Inradius of $\triangle QRS$
			Centroid of $\triangle QRS$ (Time)
			Centroid of $\triangle QRS$ (Voltage)

3. CLASSIFICATION OF THE EXTRACTED FEATURES

The classification is done based on the extracted features. Artificial neural network (ANN) has an important role in classification with the extracted features. Based on the training set data the classification is done by the classifiers. For pattern recognition artificial neural network (ANN), K-nearest neighbor (KNN) and support vector machine (SVM) are the classifiers used to implement the proposed methodology [19].

In this work, the standard MIT-BIH ECG-ID database has been used. The Massachusetts Institute of Technology-Beth Israel Hospital database consists of 48 records, of about 30 minutes, sampled at 360 Hz. Each record comprises of two signals: the modified lead II (MLII) and the second one is V1, V2, V3, V4 or V5 depending on the record. Following the inter-patient paradigm, the database was divided into 2 datasets, each dataset contains 22 records with a similar proportion of beat types:
- Training Set (Dataset 1): This dataset was used to train the classifier;
  101,106,108,109,112,114,115,116,118,119,122,124,201,203,205,207,208,209,215,223,225,230
- Testing Set (Dataset 2): This dataset was used to test the classifier;
  100,103,105,107,111,113,117,121,123,200,202,210,212,213,219,221,222,228,231,232,233,234
Notably, the Modified Lead II signal is used in our work, because it is present in every record [20].

3.1. Classification based on ANN

The ANN can be described as statistical learning model in the field of machine learning which has direct impact on biological neural network. Initially the neural network has to be tainted with electrocardiogram data of individual person and after that the biometric identification is done by the neural network which is generated by training data set [21, 22]. In this work MIT-BIH ECG ID database is used and its records are used for the purpose of training and testing. The facility of the database is that it contains more than one records for individual subjects. In this experiment 5 records are taken for every individual

patientand out of these five records first 2 records are used for train set and rest of the 3 records are used as test set. The experiment is done by taking 16 subjects, each of which contains 5 individual records. The records are taken from MIT-BIH ECG ID database and for convenience say, Person_09/rec_2 record can be renamed as H9. Here 70 features are extracted from double layer feed forward network. Softmax output neurons and Sigmoid hidden helps to classify the vectors which has enough neurons in the hidden layer. The classification works with 25 hidden layers and from there 20 output targets are chosen. Table 2 shows the confusion matrix for true classification of ECG pulse data. Each of the ECG record contains 6 PQRST fragments and one mean value are taken to prepare the confusion matrix. This makes total of 7 sets of pulse data generates the ECG record. The training set uses the pattern recognition network in the recognition phase and working with the test data extracts the features of a person arranging the sequence of training data. The probabilistic matrix also known as confusion matrix is generated by the entire simulation over the particular subject (The Person) where the probability produced by the simulation is utmost and it extracts the features of that person which is described in the confusion matrix [23, 24]. After the calculation it is found that the true positive rate (TPR) of the over all True classification is 71.18%.

True identification of the ECG pulse data is represented in Table 3. Misclassified pulses are defined by the basic difference between True classification and True identification [25, 26]. At the classification stage several pulses of a specific set may trend to misclassified. For the subject (the person) these misclassified pulses are analyzed and for true classified pulses (out of 7 pulses) it executes the maximum pulse set for true identification [27]. The confusion matrix generated by the testing, the TPR of total counted true identification of pulses is 81.75%, it is 9.95% extra compared to the pulses of true classification.

Table 2. ECG pulse data (true classification)

	H6	H7	H12	H24	H31	H41	H22	H52	H61	H35	H57	H68	H35	H44	H15	H37	H70
H6	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H7	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
H12	0	0	17	0	0	0	0	0	0	1	0	0	0	0	2	0	0
H24	0	0	0	19	0	0	0	0	0	0	0	0	2	1	0	0	1
H31	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0
H41	2	0	0	0	0	15	0	0	1	0	0	1	0	0	0	0	0
H22	0	0	0	0	3	0	12	0	0	1	0	0	0	1	0	0	0
H52	0	0	0	0	0	0	0	21	0	0	0	0	0	0	0	1	0
H61	0	0	0	0	0	0	0	0	7	0	0	0	1	0	0	0	0
H35	1	1	0	0	0	0	0	0	0	17	0	0	0	0	1	0	0
H57	0	0	0	1	1	2	0	0	0	0	21	0	0	0	0	0	0
H68	0	0	3	0	0	0	0	0	0	0	0	9	0	0	0	0	2
H35	0	0	0	0	0	0	1	0	0	0	0	0	11	0	0	0	0
H44	1	0	0	0	2	0	0	0	0	0	0	0	0	19	0	0	0
H15	0	0	0	0	0	0	0	0	1	2	0	0	0	0	21	0	0
H37	0	0	1	1	0	1	2	0	0	0	0	1	0	0	0	6	0
H70	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13
	83%	99%	84.2%	86%	44.3%	68.5%	63.7%	100%	34.8%	85.4%	100%	38.5%	49.5%	94.6%	95.2%	27.6%	54.7%

Table 3. ECG pulse data (true identification)

	H6	H7	H12	H24	H31	H41	H22	H52	H61	H35	H57	H68	H35	H44	H15	H37	H70
H6	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H7	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H12	0	0	18	0	0	1	0	0	0	0	2	2	2	0	0	0	0
H24	0	0	0	17	0	0	1	0	0	0	0	0	0	0	0	1	1
H31	0	1	0	0	16	0	0	0	0	0	0	0	0	0	0	0	0
H41	0	0	0	1	0	19	0	0	0	0	1	1	0	0	0	0	0
H22	1	0	1	0	0	0	16	0	0	0	0	0	0	1	0	0	0
H52	0	0	0	0	0	0	0	18	0	0	0	0	0	0	0	0	0
H61	0	0	0	0	0	1	0	0	21	0	0	0	1	0	0	0	0
H35	1	1	0	0	0	0	0	0	0	21	0	0	0	0	0	0	2
H57	0	0	0	0	1	2	0	0	0	0	17	0	0	0	0	1	0
H68	0	0	3	0	0	0	0	1	0	0	0	18	0	0	0	0	2
H35	2	2	0	2	0	0	1	0	0	0	0	0	14	0	0	0	0
H44	0	0	0	0	2	0	0	0	0	0	0	0	0	16	0	0	0
H15	0	0	0	0	0	0	0	0	0	0	0	3	0	0	21	0	0
H37	0	0	1	1	1	1	2	2	0	0	0	0	0	1	0	13	0
H70	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11
	81%	97%	87.2%	85%	74.3%	88.5%	73.7%	92.5%	100%	85.4%	85.3%	78.5%	69.5%	91.6%	100%	47.6%	52.7%

3.2. Overall result of the system

The experiment result is shown in the Table 4. Here the results of the parameters are shown. It is observed that the values of the performance paramaters reflects the characteristics of the samples of the ECG data taken.

Table 4. Overall performance of all parameters

Sl. No.	Performance Parameters	Value
1	Classification (True Positive Rate)	81.361%
2	Identification (True Positive Rate)	84.214%
3	Total Accuracy	95.245%
4	Total Recall	87.367%
5	Total Specificity	97.345%
6	Total Precision	93.234%
7	Total	87.341%

4. CONCLUSION

The paper proposes a systematic methodology of an ECG based biometric recognition system. The ECG data acquisition, preprocessing of data, detection of P-QRS-T, cardiac cycle classification and feature extraction are the proposed methodology. For human the ECG signal is universal and analyzing the signals unique identification of person is possible. Different machine learning classification like ANN, SVM, KNN classifiers are used to test the performance of the proposed system. For every ECG signal 6 best P-QRS-T portions are chosen by their minimum distance from the mean of the Euclidian distance where the concern peaks and their respective positions were normalized according to the R-peak position of the signal. For data extraction only these 6 fragments were considered in each record of MIT-BIH ECGID database. The obtained resultare very much supportive to implement ECG as one of the important biometric features.

REFERENCES

[1] R. M. Boole, et al., "Guide to Biometrics," *Springer-Verlag*, New York, 2004.
[2] H. Abbaspour, et al., "ElectrocardiogramBased Identification using a New Effective Intelligent Selection of Fused Features," *Journal of Medical Signal and Sensors*, vol. 5, no. 1, pp. 30-39, 2015.
[3] Y. N. Singh and S. K. Singh, "Evaluation of Electrocardiogram for Biometric Authentication," *Journal of Information Security*, vol. 3, no. 1, pp. 39-48, 2012.
[4] M. Tantawi, et al., "ECG Signals Analysis for Biometric Recognition," *International Conference on Hybrid Intelligent Systems (HIS)*, pp. 169-175, 2014.
[5] F. Agrafioti, et al., "Heart Biometrics: Theory, Methodsand Applications," *Biometrics*, pp. 199-216, 2011.
[6] L. Biel, et al., "ECG Analysis: A New Approach in Human Identification," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 3, pp. 808-812, 2001.
[7] P. de Chazal, et al., "Automatic Classification of Heartbeats Using ECG Morphology and Heartbeat Interval Features," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 7, pp. 1196-1206, 2004.
[8] "MIT-BIH Arrhythmia Database," 2005. [Online]. Available: https://physionet.org/physiobank/database/mitdb/
[9] G. B. Moody and R. G. Mark, "The MIT-BIH Arrhythmia Database on CD-ROMand Software for Use with IT," *Proceedings Computers in Cardiology*, pp. 185-188, 1990.
[10] V. H. C. De Albuquerque, et al., "Robust AutomatedCardiac Arrhythmia Detection in ECG Beat Signals," *Neural Computing and Applications*, vol. 29, pp. 679-693, 2018.
[11] P. M. Agante and J. P. M. de Sa, "ECG Noise Filtering Using Wavelets with Soft-thresholding Methods," *Computers in Cardiology*, vol. 26, pp. 535-538, 1999.
[12] Y. Sun, et al., "Characteristic wave detection inECG signal using morphological transform," *BMC Cardiovascular Disorders*, vol. 5, no. 1, pp. 29-35, 2005.
[13] S. Saminu, et al., "Wavelet Feature Extraction for ECG Beat Classification," *2014 IEEE 6th International Conference on Adaptive Science and Technology*, pp. 978-983, 2014.
[14] L. Williams and Wilkins, "ECG interpretation made incredibly easy!," Philadelphia, Wolters Kluwer, 2015.
[15] K.K. Patroand, P.R. Kumar, "A Novel Frequency-Time Based Approach for the Detection of Characteristic Waves in Electrocardiogram Signal," *Microelectronics, Electromagnetics and Telecommunications*, pp. 57-67, 2016.
[16] Y. Wang, et al., "Analysis of Human Electrocardiogram for Biometric Recognition," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, pp.1-11, 2008.
[17] C. Ye, et al., "Heartbeat Classification Using Morphological andDynamic Features of ECG Signals," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 10, pp. 2930-2941, 2012.
[18] M. I. Owis, et al., "Study of Features Based on Nonlinear Dynamical Modeling in ECG Arrhythmia Detection and Classification," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 7, pp. 733-736, 2002.

[19] R. V. Andreão, et al., "ECG Signal Analysis ThroughHidden Markov Models," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 8, pp. 1541-1549, 2006.

[20] S. M. Jadhav, et al., "Artificial Neural Network Based Cardiac Arrhythmia Classification Using ECG Signal Data," *International Conference on Electronics and Information Engineering*, pp. 228-231, 2010.

[21] B. Anuradhaand, V. C. V. Reddy, "ANN for Classificationof Cardiac Arrhythmias," *ARPN Journal of Engineering and Applied Sciences*, vol. 3, no. 3, pp. 1-6, 2008.

[22] S. N. Yuand K. T. Chou, "Integration of independent component analysis and neural networks for ECG beat classification," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2841-2846, 2008.

[23] J.C Platt, et al., "Large Margin DAGs for MulticlassCristianiniass Classification," *Advances in Neural Information Processing Systems 12 (NIPS 1999)*, pp. 548-553, 1999.

[24] J. A. Nasiri, et al., "ECG Arrhythmia Classification with Support Vector Machines and Genetic Algorithm," *Third UKSim European Symposium on Computer Modeling and Simulation*, pp. 187-192, 2009.

[25] A. Fratini, et al, "Individual identification via electrocardiogram analysis," *Bio Medical Engineering Online*, Aug. 2015.

[26] Syed Muhammad Anwar, et al, "Arrhythmia Classification of ECG Signals Using Hybrid Features," *Computational and Mathematical Methods in Medicine*, Nov. 2018.

[27] Fahim Sufi1, et al, "Polynomial distance measurement for ECG based biometric authentication," *Security Communication Networks*, *Wiley Online Library*, pp. 303-319, 2008.

BIOGRAPHIES OF AUTHORS

Mr. Sumanta Kuila is currently working as an Assistant Professor in the Deptartment of Computer Science and Engineering, Haldia Institute of Technology, Haldia-721657, India. He has 16 years of experience in the field of Computer Science and Engineering. He is doing Ph.D in Computer Science and Engineering from Amity University Uttar Pradesh, Lucknow campus. He did his masters (M.E) from west Bengal University of Technology in 2005.

Dr. Namrata Dhanda is presently Professor in the Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University Uttar Pradesh, Lucknow campus. She received her Ph.D from University of Petroleum and Energy Studies, Dehradun, India in 2015. She has 18 years of teaching experience in different reputed engineering and technical institutes in India. Her current research interest includes Data Mining, Big Data Analytics, Big Data, Data Wire housing Internet of Things etc.

Dr. Subhankar Joardar is presently Professor and Head in the Department of Computer Science and Engineering, Haldia Institute of Technology, Haldia-721657, India. He received his Ph.D degree from Birla Institute of Technology, Mesra, Ranchi, India in 2016. He did his masters (M. Tech and MCA) both from BIT, Mesra, Ranchi in 2009 and 2002 respectively. He has published more than 14 technical papers in the referred journals/conferences. He has served as Organizing Chair of international conference (ICITAM 2017).