CISTER
Research Center in
Real-Time & Embedded
Computing Systems

Conference Paper

ECG Arrhythmia Classification Using Mahalanobis-Taguchi System in a Body Area Network Environment

Aftab Ali

Nur Al Hasan Haldar

Farrukh Aslam Khan

Sana Ullah*

*CISTER Research Center

CISTER-TR-160304

2015/12/06

ECG Arrhythmia Classification Using Mahalanobis-Taguchi System in a Body Area Network Environment

Aftab Ali, Nur Al Hasan Haldar, Farrukh Aslam Khan, Sana Ullah*

*CISTER Research Center

Polytechnic Institute of Porto (ISEP-IPP)

Rua Dr. António Bernardino de Almeida, 431

4200-072 Porto

Portugal

Tel.: +351.22.8340509, Fax: +351.22.8321159

E-mail: sauah@isep.ipp.pt

http://www.cister.isep.ipp.pt

Abstract

Arrhythmia is caused by improper and irregular sinus rhythm or heartbeats. In order to diagnose cardiac arrhythmia, electrocardiogram (ECG) beat classification and analysis is very necessary. The efficiency and accuracy of any classification model highly depends on selecting the most relevant features. The aim of this study is to classify different arrhythmic beats with a reduced set of relevant-only ECG features. To optimize the ECG feature selection process and increase the classification accuracy, a Mahalanobis-Taguchi System (MTS) based classifica- tion and analysis scheme is proposed. MTS is a multi-dimensional pattern recognition system which dynamically selects important features for further analysis. Arrhythmia can occur at any time and thus requires proper and continuous monitoring of the patient to reduce sudden heart attacks. The proposed MTS- based classification scheme is integrated with a Wireless Body Area Network (WBAN) for pervasive monitoring. The proposed scheme is analyzed and compared with a state-of-the-art scheme in terms of sensitivity, specificity, and accuracy. The results show that the proposed scheme performs significantly better than the other scheme by achieving high sensitivity, specificity, and classification accuracy for different arrhythmic heartbeats i.e., Left Bundle Branch Block (LBBB), Premature Ventricular Contraction (PVC), Right Bundle Branch Block (RBBB), and Atrial Premature Contraction (APC).

ECG Arrhythmia Classification using Mahalanobis-Taguchi System in a Body Area Network Environment

Aftab Ali[†,‡], Nur Al Hasan Haldar[†], Farrukh Aslam Khan[†,‡], Sana Ullah[*,§]
[†]Center of Excellence in Information Assurance (CoEIA), King Saud University, Riyadh, Saudi Arabia
[‡] National University of Computer and Emerging Sciences, A. K. Brohi Road, H-11/4, Islamabad, Pakistan
[*] CISTER Research Center, ISEP, Polytechnic Institute of Porto (IPP), 4200-135 Porto, Portugal
[§] Department of Computer and Software Technology, University of Swat, Odigram, Pakistan
Email: {afkhan.c, nhaldar.c, fakhan}@ksu.edu.sa, sauah@isep.ipp.pt

Abstract—**Arrhythmia is caused by improper and irregular sinus rhythm or heartbeats. In order to diagnose cardiac arrhythmia, electrocardiogram (ECG) beat classification and analysis is very necessary. The efficiency and accuracy of any classification model highly depends on selecting the most relevant features. The aim of this study is to classify different arrhythmic beats with a reduced set of relevant-only ECG features. To optimize the ECG feature selection process and increase the classification accuracy, a Mahalanobis-Taguchi System (MTS) based classification and analysis scheme is proposed. MTS is a multi-dimensional pattern recognition system which dynamically selects important features for further analysis. Arrhythmia can occur at any time and thus requires proper and continuous monitoring of the patient to reduce sudden heart attacks. The proposed MTS-based classification scheme is integrated with a Wireless Body Area Network (WBAN) for pervasive monitoring. The proposed scheme is analyzed and compared with a state-of-the-art scheme in terms of sensitivity, specificity, and accuracy. The results show that the proposed scheme performs significantly better than the other scheme by achieving high sensitivity, specificity, and classification accuracy for different arrhythmic heartbeats i.e., Left Bundle Branch Block (LBBB), Premature Ventricular Contraction (PVC), Right Bundle Branch Block (RBBB), and Atrial Premature Contraction (APC).**

I. INTRODUCTION

An electrocardiogram (ECG) is used to record the electrical activity and muscular function of the heart so that heartbeat irregularities, also known as arrhythmia, can be diagnosed. Some types of arrhythmia may be life-threatening and need to be detected earlier. Detection of life-threatening arrhythmia is challenging because an arrhythmic beat can occur infrequently and the deviation of such beat pattern is very minute. Continuous monitoring of the ECG signal is necessary because an arrhythmia can occur at a random location and time. Similarly, another challenge is the optimization of the arrhythmia detection process to reduce the unnecessary computation burden.

Much of the literature is dedicated to the detection of arrhythmia and cardiac abnormalities. In this context, [1] uses morphological and time-frequency domain analysis for cardiac arrhythmia detection. Similarly, a hidden Markov model (HMM) is used in [2] to automatically detect and analyze the ECG data. Neural network based classification schemes are used in [3] and [4] to classify and detect cardiac arrhythmia in an ECG. Another scheme which uses a neural network and a type-2 fuzzy c-means clustering (T2FCM) algorithm for the analysis and classification of ECG arrhythmia is presented in [5]. Similarly, a supervised neural network-based algorithm is used in [6] to classify ECG records between normal and ischemic beats. These schemes all use clustering techniques to classify normal and abnormal ECG beats. A cluster analysis scheme is presented in [7], which is based on simple Mahalanobis distance [8] to automatically classify cardiac arrhythmia.

In the above discussion, all of the neural network-based schemes suffer from convergence to local and global minima. Therefore, these schemes are very expensive to use in continuous health monitoring systems. Furthermore, most of these schemes are supervised, which means that manual labelling is required for classification. Therefore, these schemes cannot be used in live health monitoring systems as human involvement is not practical in such cases. The cluster analysis scheme discussed above uses static features, which reduce the energy efficiency and performance of the overall system. In other words, once features are selected, they will always be used for each and every case or record of the ECG arrhythmia database during the detection process. It is possible that some features might not work as efficiently for a particular record as it worked for some other record. This approach increases the chances of useless features in the analysis process, which adds burden on the whole detection mechanism. For example, if a particular record requires only a few features to detect cardiac abnormality, then there is no need to use useless features in the detection process for that particular record.

The Mahanalobis-Taghuchi System (MTS) [9] is a diagnosis and predictive method for analyzing patterns in multivariate cases that can be used in a range of applications including the recognition of liver diseases. In [10], MTS is used for the detection and classification of breast cancer data with nine attributes. The experiments showed that MTS performed better than the neural network based technique used on the same dataset. In [11], MTS is combined with Multifractal analysis scheme to predict the condition of a chemical industry complex system. Multifractal analysis is used to extract nonlinear features from the complex data, and then MTS is applied to classify important features of the multi-variable environment.

This paper presents an MTS-based cardiac arrhythmia

978-1-4799-5952-5/15/$31.00 ©2015 IEEE

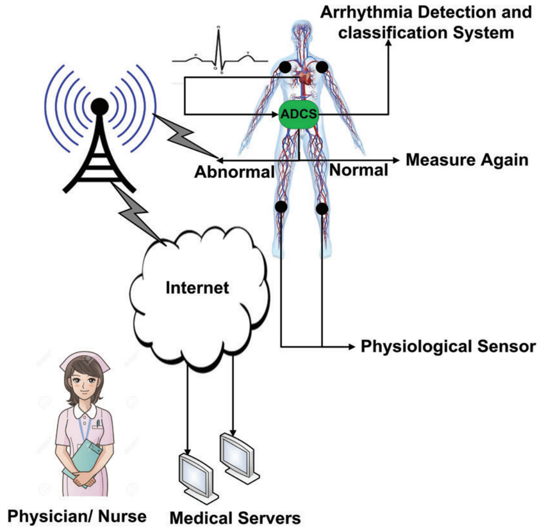


Fig. 1: The Overall Architecture of Arrhythmia Detection System

classification and detection system using continuous and ubiquitous monitoring in a Wireless Body Area Network (WBAN) environment. The proposed work uses dynamically selected features for arrhythmia detection and classification. In other words, the feature utilization varies from record to record, so if the requirements of detection and classification can be fulfilled for a particular record using minimum features, then the scheme uses only those features which are best suited for that particular record. This behavior of the proposed scheme makes it very efficient and robust for cardiac arrhythmia detection. Since the ECG signal is non-stationary in nature, the arrhythmia can occur at any time with irregular intervals. To handle this uncertainty of arrhythmia occurrence, the WBAN-based architecture is proposed. The integration of WBAN will make the detection scheme ubiquitous and continuous [12-16], which will help in correctly locating the occurrence of cardiac arrhythmia. The end-to-end communication process [17] of continuous monitoring for cardiac arrhythmia detection using a WBAN environment is shown in Fig. 1. A comparison of the proposed scheme with the cluster analysis scheme [7] shows that the proposed scheme outperforms the cluster analysis scheme in terms of sensitivity, specificity, and accuracy for each of the five beats classifications. To the best of our knowledge, the proposed work uses MTS for the first time for the arrhythmia detection problem.

The remainder of the paper is organized as follows: Section II elaborates the proposed work. Section III presents the results and analysis, and section IV concludes the paper.

II. PROPOSED WORK

The Mahalanobis-Taguchi System (MTS) consists of the Taguchi method and Mahalanobis Distance (MD) method, where the Taguchi method is used to find optimal system parameters, and Mahalanobis Distance measures the statistical distance between a point and the reference dataset. The details of the MTS algorithm are available in [18]. The overall system description is presented below.

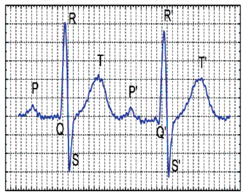


Fig. 2: PQRST Wave of ECG Heartbeats

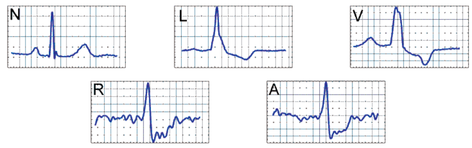


Fig. 3: Normal, LBBB, PVC, RBBB, and APC Heartbeat Patterns

A. ECG Data Collection

The ECG data used for the experiments was collected from the MIT-BIH Arrhythmia database [19]. This database contains ECG records for 48 unique individuals. Each record is 30 minutes long. In most cases, the upper signal is a modified limb lead II (MLII), which is available for most of the data records collected for our experiments. For every person, 650,000 records were found where the time difference between each record was 0.00277778 ($\sim$ 0.003s) seconds. Out of the 48 records available in MIT-BIH arrhythmia database, 43 records were used for our experiments. The other 5 records (i.e., Record ID 201, 219, 102, 104, 107) demonstrated internal inconsistencies. Therefore, these records were excluded from our experiments.

B. RR Interval and QRS Detection

By using the Waveform Database (WFDB) software package [20], the R to R wave (RR) interval for each person is extracted from his/her data records. After obtaining the RR intervals and R peak nodes, the QRS complex is identified using single-channel QRS detector WQRS [21]. This algorithm is based on length transformation unlike other available QRS detection methods, which are based on slope detection. By using WQRS, the temporal location of the assumed QRS candidate is determined [22]. An example of P, Q, R, S, and T node positions is shown in Fig. 2.

C. ECG Attribute Selection

Different arrhythmia have different characteristics and natures. An arrhythmic beat can differ from Normal or other types of beats in terms of a significantly different range of attribute values. In comparison with NORM (Normal) beats, the QRS interval, QRS area, R-amplitude, QRS-ratio, etc. are higher in Premature Ventricular Contraction (PVC) arrhythmic beats, whereas RR-ratio and RR-interval values are significantly lower. Similarly, in the case of Atrial Premature Contraction (APC), the RR-interval and RR-ratio are lower (not as much lower as PVC) in comparison to Normal beats. There is a significant difference in the QRS area, QRS duration, and R-amplitude in APC and PVC types of arrhythmia. QRS duration is also higher in Left Bundle Branch Block (LBBB)

TABLE I: ECG FEATURE DESCRIPTION IDENTIFIED AS INITIAL FEATURES

Attribute Number	Attribute Symbol	Description of Attribute	Units
A	RR-int	Duration gap between two consecutive R peaks	ms
B	RRa/b	The ratio of RR interval of previous beat (RRa) with current beat (RRb)	....
C	RRavg(a,c)-b	Difference between average of RRa and RRc from RRb	ms
D	QRS-dur	The time duration between Q and R nodes of QRS complex	ms
E	QRS a/b	Ratio of QRSa (QRS previous) duration and QRSb (QRS current) duration	....
F	RR-ratio	Ratio of RR interval with average RR interval	....
G	QRS-ratio	Ratio of QRS duration with average QRS interval	....
H	R-amp	The amplitude value of R peak	mV
I	QRS-area	QRS complex area	mV * ms

and Right Bundle Branch Block (RBBB) than in Normal beats, but this attribute difference is significantly more pronounced in the case of LBBB than it is in RBBB. Hence, RR interval, QRS duration, R-amplitude, QRS-area, RR-ratio with respect to Average RR interval of all records, and QRS ratio with respect to average QRS interval are chosen as initial attributes. The last two attributes are calculated by dividing the main attribute value of corresponding beats with the average beat value of the entire dataset. The preceding ratios of RR interval and QRS duration are also chosen as initial features. Let us assume the RR interval of the current beat (b) is RR_b, the RR interval of the immediately preceding beat (a) is RR_a, and the preceding ratio of current beat b, $RR_{a/b}$ = (RR_a/ RR_b). In the same way, the preceding ratio of QRS duration is calculated. These two attributes are important for the detection of sudden changes in the beat sequence. Another feature, based on RR interval differences, is taken as an initial attribute in our proposed system. This feature can compare the RR interval with the average value of the preceding and following interval values. This feature is symbolized as $RR_{avg(a,c)-b}$. Let b be the current beat and a, c be the previous and next beats, respectively, then $RR_{avg(a,c)-b}$ = (RR_a + RR_c)/2 - RR_b. In a sequence of the same type of ECG beat, the value of $RR_{avg(a,c)-b}$ should be very near to 0; irregular beats will cause the value to differ from 0. Hence, this feature, which is based on RR interval, can also differentiate arrhythmic beats. Fig. 3 shows the pattern of the above-mentioned five classes. Hence, in our proposed system, a total of 9 attributes are selected based on their importance in arrhythmia detection. The 9 attributes are mentioned in Table I.

D. Arrhythmic Heartbeat Detection

The proposed system deals with arrhythmic beat identification using ECG data in a ubiquitous environment. ECG signals will be collected by the ECG sensors attached to the upper part of the patient's body. Some sample data that contain different arrhythmic and normal conditions are essential for creating the training dataset. MTS is used on both the training and testing datasets. For experimental purposes, the arrhythmia dataset available at MIT-BIH [19] was used. This dataset is 30 minutes long divided into 5 minutes of training data and 25 minutes of testing data.

TABLE II: L_{12} ORTHOGONAL ARRAY AND GAIN CALCULATED FROM SN RATIO

Run#	Attributes										
	A	B	C	D	E	F	G	H	I		
1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	2	2	2	2	2	2
3	1	1	2	2	2	1	1	1	2	2	2
4	1	2	1	2	2	1	2	2	1	2	2
5	1	2	2	1	2	2	1	2	1	1	1
6	1	2	2	2	1	2	2	1	2	1	1
7	2	1	2	2	1	1	2	2	1	1	1
8	2	1	2	1	2	2	2	1	1	2	2
9	2	1	1	2	2	2	1	2	2	1	1
10	2	2	2	1	1	1	1	2	2	2	2
11	2	2	1	2	1	2	1	1	1	2	2
12	2	2	1	1	2	1	2	1	2	1	1
Gain	25.52	15.36	12.03	12.21	8.05	9.98	9.96	5.66	7.50		

E. Construction of Full Model MTS Measurement Scale

For constructing a full model MTS measurement scale (MS), Normal data needs to be collected from the training dataset. In case if there is no Normal data in the training sample, other annotated data are used as reference data. The Normal (healthy) dataset is standardized by using mean and standard deviation of each corresponding attribute. The correlation matrix and inverse correlation is calculated using the reference data. Abnormal attributes of a particular arrhythmia type (e.g. PVC) are considered to validate the full MTS model. The Normal (reference) and Abnormal data are collectively referred to as the training dataset. By using the mean and standard deviation of the Normal (reference) sample, MD is calculated for the entire training dataset (Normal and Abnormal together).

As an example, for construction of the full MTS model of person number 119 in the MIT-BIH Dataset, a total of 326 sample beats (246 Normal and 80 Abnormal) were used. These samples were standardized using the mean and standard deviation of the Normal data. MD was calculated for the dataset using the Normal dataset as a reference. The MDs related to Abnormal observations were larger than MDs corresponding to the Normal sample. This scenario defines the concept of a "good" measurement scale.

F. Important Variables Identification

Once the full model for the MTS measurement scale is built, the dataset needs to be filtered so that irrelevant variables are sorted out and only relevant variables remain. By using Orthogonal Array (OA) and SN ratios, which measure the significance of each data attribute, the most important variables are selected. In our proposed system, a total of 9 features were initially selected. With an OA L_{12} and level 2 (2^{11}), we evaluate each row to decide whether the attribute should be included or excluded. Each attribute is represented in a column. The last two columns of OA (grey shaded) in Table II cannot be used as only 9 attributes are available for the initial construction of the system.

Table II shows the attributes in columns, whereas the combination of including and excluding attributes as runs are presented in each row. Using L_{12} OA, abnormal test data (PVC sample before 5 minutes in the case of person 119) are used to calculate the SN ratio. The SN ratio, which measures the accuracy of any measurement scale, is calculated using the MDs of each run as shown in equation (1).

$$SN_q = -10log\left[\frac{1}{p}\sum_{j=1}^{p}\frac{1}{MD_j}\right], \quad (1)$$

where j=1,2,...,p and p is the number of abnormalities

SN_q represents the SN ratio of q_{th} run of the OA. The average SN ratios for level 1 (when level 1 characteristics are included as a variable) and level 2 (when level 2 characteristics are included as a variable) are calculated for each characteristic. The Gain in SN ratio for each characteristic is calculated as follows:

$$Gain = (Average\ SN\ Ratio)_{Level1} - (Average\ SN\ Ratio)_{Level2} \quad (2)$$

Each gain value is assigned against the attribute. The most positive valued attributes A (RR-int), B ($RR_{a/b}$), D ($RR_{avg((a,c)-b)}$), and C (QRS-dur) are selected (for person 119), and lesser valued attributes are discarded. The selected set of relevant attributes is not fixed for all other persons. The number of selected attributes may vary from person to person. In our experiments, the number of most relevant attributes ranged from three to five.

G. Construction of Reduced Model of MTS Scale

The normal sample of training data with a reduced set of attributes is used to construct MS. In the case of person number 119, only four attributes (A, B, C and D), which are determined from the previous step as most relevant, are used for further processing.

H. Future Diagnosis with Relevant Variables

For future diagnosis, the most relevant attributes are used to calculate the MD for each testing beat. If the calculated MD for each beat is less than a threshold, then the entry will be included in the normal cluster, otherwise the particular entry will be treated as abnormal. The labelling of each cluster depends upon the properties of the cluster elements. A cluster which contains more similar features out of the five groups (NORM, PVC, APC, LBBB, and RBBB) will be labelled accordingly. For evaluation and testing purposes, the ECG beats which are available in the last 25 minutes of each person's data record are treated as the test dataset. The threshold value for clustering normal and abnormal beats for person 119 is identified as 23.7. In Fig. 4, the sudden increase of MD value at this threshold is shown. On the other hand in Fig. 5, the corresponding MD of each heartbeat is plotted against the test sample for the same person. The red horizontal line is the threshold value, which was identified from Fig. 4. The beats which have MDs less than this threshold are treated as normal, otherwise the beats are regarded as arrhythmic. According to our results, the number of normal beats is identified as 1,297, and the remaining 364 beats are identified as arrhythmic. The green-colored points in Fig. 5 are identified as Normal heartbeats, whereas the blue-colored points are marked as abnormal. The red line which is parallel to the X- axis is separating the Normal and Abnormal data points.

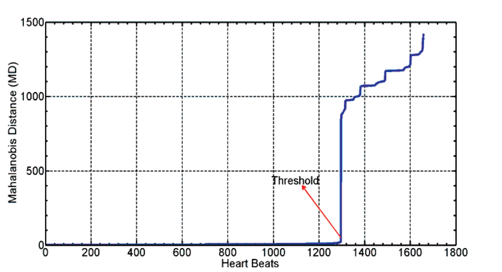


Fig. 4: Threshold Identification using MD Values for Person 119

III. EXPERIMENTS AND RESULTS

A. Experimental Setup

In order to perform different experiments and analysis, a Java-based programming environment and MATLAB based simulation environment were used. The dataset of the ECG records was collected from the MIT-BIH arrhythmia database [7], which contains 48 different ECG recordings, from which we selected 43 records for our experiments on the basis of training data availability. For in-depth analysis, we examine the performance and efficiency of the proposed scheme in terms of sensitivity, specificity, and accuracy. This work considers different types of arrhythmia like LBBB (L), PVC (V), RBBB (R), and APC (A). The classification of Normal (N) in the presence of these arrhythmia is the main objective of the proposed system. The results are calculated and obtained for each arrhythmia type individually.

B. Results and Discussion

Sensitivity and specificity are the statistical measures used to analyze the performance of a binary classification test. In the context of this work, sensitivity of a test refers to the ability of a test to correctly identify those records of the ECG that contain any of the five types of arrhythmia. Similarly, specificity refers to the ability of the test to correctly identify those patients without the disease. The accuracy also has been measured to validate the usefulness of the proposed scheme.

1) Detection of N beats:

The average sensitivity of normal beats (N) is found to be 98.07%, which clearly demonstrates that the scheme detects 98.07% of the records with the N (true positives), but 1.94%

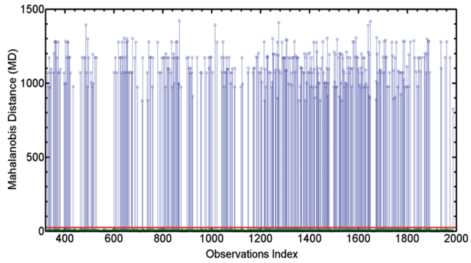


Fig. 5: Heartbeat Index-wise Mahalanobis Distance Plot

TABLE III: SENSITIVITY, SPECIFICITY, AND ACCURACY OF N BEATS CLASSIFICATION

Detection of N in ECG data							
Record ID	FP (%)	FN (%)	TP (%)	TN (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
100	0.38	0.53	99.47	97.57	99.47	99.62	99.54
101	0.20	0.53	99.47	98.95	99.47	99.80	99.64
103	0.06	0.17	99.83	99.17	99.83	99.94	99.88
105	0.85	0.75	99.25	96.57	99.25	99.13	99.19
106	2.33	1.62	98.38	97.53	98.38	97.67	98.03
108	0.55	0.62	99.38	97.12	99.38	99.44	99.41
112	0.29	0.28	99.72	98.62	99.72	99.71	99.71
113	0.00	0.13	99.87	99.32	99.87	100.00	99.93
114	2.08	0.13	99.87	99.38	99.87	97.95	98.91
115	0.00	0.00	100.00	100.00	100.00	100.00	100.00
116	3.05	2.14	97.86	92.34	97.86	96.81	97.35
117	0.00	0.23	99.77	98.82	99.77	100.00	99.88
119	0.00	0.15	99.85	99.71	99.85	100.00	99.92
121	0.00	0.26	99.74	98.71	99.74	100.00	99.87
122	0.00	0.05	99.95	99.76	99.95	100.00	99.98
123	0.00	0.00	100.00	100.00	100.00	100.00	100.00
200	8.50	4.94	95.06	94.25	95.06	91.73	93.37
202	2.92	0.33	99.67	98.25	99.67	97.11	98.38
203	9.75	4.90	95.10	89.49	95.10	90.18	92.65
205	1.81	0.47	99.53	98.16	99.53	98.19	98.86
208	25.79	15.06	84.94	89.32	84.94	77.60	81.01
209	2.15	2.56	97.44	93.98	97.44	97.76	97.60
210	3.65	2.74	97.26	92.09	97.26	96.19	96.74
212	9.58	4.91	95.09	98.04	95.09	91.10	93.02
213	8.29	0.86	99.14	98.21	99.14	92.22	95.57
215	2.66	0.86	99.14	96.83	99.14	97.33	98.24
220	1.18	0.56	99.44	98.63	99.44	98.82	99.13
221	4.33	2.76	97.24	93.90	97.24	95.59	96.42
222	16.85	7.14	92.86	86.69	92.86	83.73	88.21
223	9.57	5.31	94.69	91.51	94.69	90.53	92.60
228	6.73	4.79	95.21	90.72	95.21	93.10	94.17
230	0.05	0.11	99.89	99.50	99.89	99.95	99.92
231	0.98	0.00	100.00	100.00	100.00	99.03	99.51
233	3.42	1.13	98.87	98.31	98.87	96.64	97.75
234	0.31	0.67	99.33	97.17	99.33	99.68	99.50
Average	3.67	1.94	98.07	96.53	98.07	96.47	97.25

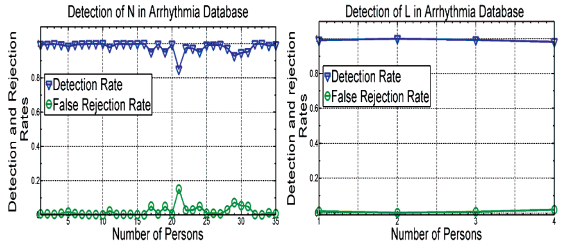
of the records containing N go undetected (false negatives). The sensitivity of N beats classification is high due to the scheme having very low false negatives (FN) and very high true positives (TP) as seen in Table III.

The specificity of N beats classification is high, ranging up to 96.47%, as shown in Table III. This demonstrates that the proposed scheme has better true negatives (TN) (96.53%) and very low false positives (FP) (3.67%) for N beats classification. The reason behind the high detection results of N is the availability of sufficient training data in each patient record of the dataset.

The scheme has 97.25% accuracy for N beats classification because it has very high TP (98.07%) and TN (96.53%). Similarly, accuracy also depends on the FP and FN; lower values of FN and FP assure higher accuracy. The average detection rate for N beats classification is very high, as can be seen in Fig. 6(a), because the scheme has a very high TP and very low FN.

2) Detection of L beats:

For L beats classification, the average sensitivity is 99.11%, while the misclassification rate is 1.86%. The sensitivity of L beats classification is again high due to high TP and very low FN as shown in Table IV. The specificity of L beats



(a) Detection of N Beats **(b)** Detection of L Beats

Fig. 6: N and L Beats Detection in Arrhythmia Database

classification is also high (up to 96.80%). This high specificity demonstrates that the proposed scheme has better TN (96.03%) and very low FP (3.29%) for L beats classification. The accuracy of L beats classification is 97.94%; again, this is high due to the high values of TP (99.11%) and TN (96.03%) and the low values of FP (3.29%) and FN (0.89%). Our scheme delivers a very high detection rate and very low false rejection rate for L beats classification because it exhibits high TP and low FN as shown in Fig. 6(b).

TABLE IV: SENSITIVITY, SPECIFICITY, AND ACCURACY OF L BEATS CLASSIFICATION

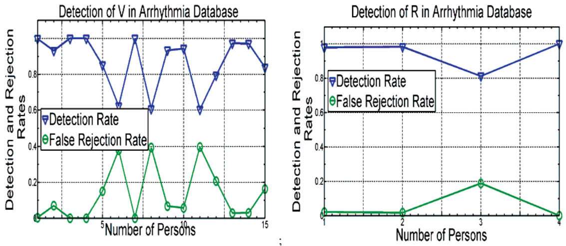
Detection of L in ECG data							
Record ID	FP (%)	FN (%)	TP (%)	TN (%)	Sensitivity %)	Specificity (%)	Accuracy (%)
109	0.44	0.92	99.08	96.07	99.08	99.55	99.31
111	0.06	0.06	99.94	94.00	99.94	99.94	99.94
207	5.35	0.73	99.27	98.97	99.27	94.87	97.03
214	7.32	1.86	98.14	95.07	98.14	92.85	95.46
Average	3.29	0.89	99.11	96.03	99.11	96.80	97.94

3) Detection of V beats:

For V beats classification, the average sensitivity is 94.53%, which means that our scheme misclassifies 5.47% of the records having V. The high sensitivity of V beats classification is due to the scheme producing very low FN (5.47%) and very high TP (94.53%), as shown in Table V. Similarly, on average the specificity of the V beats classification is also very high (96.12%). The higher specificity for V beats classification is the result of having very high TN (98.67%) and very low FP (4.31%). Moreover, due to the higher TP and TN, the V beats classification also has a higher accuracy, confirming the overall superior detection rate and better performance of the proposed scheme for V beats classification as can be seen in Fig. 7(a).

TABLE V: SENSITIVITY, SPECIFICITY, AND ACCURACY OF V BEATS CLASSIFICATION

Detection of V in ECG data							
Record ID	FP (%)	FN (%)	TP (%)	TN (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
105	3.57	6.9	93.1	99.92	93.1	96.55	94.85
106	0	0	100	100	100	100	100.00
119	0	0	100	100	100	100	100.00
200	4.94	14.86	85.14	94.81	85.14	95.04	90.09
207	20	0	100	100	100	83.33	90.91
221	0.65	2.85	97.15	99.58	97.15	99.35	98.25
228	3.3	2.98	97.02	99.49	97.02	96.79	96.90
233	2.03	16.19	83.82	95.52	83.82	97.92	90.78
Average	4.31	5.47	94.53	98.67	94.53	96.12	95.22



(a) Detection of V Beats **(b)** Detection of R Beats

Fig. 7: V and R Beats Detection in Arrhythmia Database

4) Detection of R beats:

As discussed above, the sensitivity depends upon TP. A lower TP in a particular record results in degradation of sensitivity for that particular record and eventually affects the overall average sensitivity. For example, the sensitivity of R beats classification is 94.34%, which is a little lower than expected because of insufficient training data for record number 212. If a record contains only 2 or 3 R beats in 5 minutes of training data, then the scheme will be unable to detect the R beats in that particular record properly because the scheme is not well trained for the R beats. This scenario can be seen in Table VI, where the TP is 81.15% and the FN is 18.85% for record number 212.

The scheme produces 93.54% of specificity for R beats classification because it has very high TN values for all of the records except record number 212. This decrease in TN reduces the overall specificity of the scheme, but not severely because the average is affected by only one record. Since TP is a key component in the accuracy calculation, lower TP has the effect of degrading the reported accuracy.

Fig. 7(b) shows a relatively good detection rate for R beats classification due to high TP and low FN in some cases, while other cases have lower performance due to low TP. For example, person number 3 in Fig. 7(b) shows a low detection rate and a high false rejection rate because of the lack of sufficient training data for R beats in that particular case.

TABLE VI: SENSITIVITY, SPECIFICITY, AND ACCURACY OF R BEATS CLASSIFICATION

Detection of R in ECG data							
Record ID	FP (%)	FN (%)	TP (%)	TN (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
118	2.09	2.09	97.91	92.40	97.91	97.79	97.85
124	3.86	1.70	98.30	93.57	98.30	98.30	97.18
212	22.98	18.85	81.15	81.73	81.15	78.05	79.57
231	0.00	0.00	100.00	100.00	100.00	100.00	100.00
Average	7.23	5.66	94.34	91.93	94.34	93.54	93.65

5) Detection of A beats:

The classification of A beats is high due to the high detection rate and low false rejection rate produced by the proposed scheme. As can be seen in Table VII, the sensitivity of the proposed scheme is high due to high TP. Similarly, the specificity is also very high due to high TN. As a direct result, accuracy, which depends on both TP and TN, is also high. Figure 8 shows the detection and false rejection rate for A beats classification. This results in the proposed scheme demonstrating high performance for the A beats classification.

TABLE VII: SENSITIVITY, SPECIFICITY, AND ACCURACY OF A BEATS CLASSIFICATION

Detection of A in ECG data							
Record ID	FP (%)	FN (%)	TP (%)	TN (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
100	0.000	0.000	100.00	100.000	100.000	100.000	100.000
118	11.8	12.8	87.2	89.6	87.2	88.4	87.8
232	0.086	0.086	99.914	99.837	99.914	99.914	99.914
Average	3.96	4.30	95.703	96.48	95.70	96.10	95.90

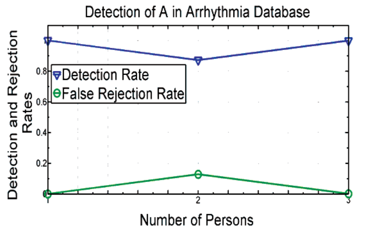


Fig. 8: Detection of A Beats in Arrhythmia Database

The respective results for average sensitivity, specificity, and accuracy are 98.07%, 96.47%, and 97.25% for N classified beats; 99.11%, 96.80%, and 97.94% for L classified beats; 94.53%, 96.12%, and 95.22% for V classified beats; 94.34%, 93.54%, and 93.65% for R classified beats; and 95.70%, 96.10%, and 95.90% for A classified beats. The lower results for sensitivity, specificity, and accuracy in some cases are due to insufficient training data availability. For example, patient record ID 118 contains only 10 A beats in 5 minutes of training data. This reduces the accuracy to 87.8%, resulting in a decrease of the overall A beats classification to 95.90%. Similarly, in the classification of V beats, there are certain records in the training data which do not contain sufficient information for the system to be well trained. This lack of sufficient training data ultimately degrades the average accuracy of the system to 95.22% for V beats classification.

6) Comparison with Cluster Analysis Scheme:

The proposed scheme is compared with the cluster analysis scheme [7] in terms of sensitivity and accuracy. The cluster analysis scheme uses only 14 records from MIT arrhythmia database (person IDs 103, 113,123, 234, 111, 214, 118, 212, 231, 200, 221, 233, 222, and 232). Out of these 14 records, record numbers 103, 113, 123, and 234 are used as the Normal (N) case. Of the remaining records, two (111 and 214) are of the LBBB beats, three (118, 212, and 231) are of the RBBB beats, three (200, 221, and 233) are of the PVC database, and two (123 and 234) are of the APC database [7].

In order to compare the sensitivity and accuracy of the proposed scheme with that of the cluster analysis scheme, the results are calculated for 14 records. The same evaluation parameters are also calculated for a total of 43 persons from the same MIT database. The comparison in Table VIII shows that the proposed scheme demonstrates markedly improved results in all cases of classification. The overall sensitivity in the case of PVC is higher, though it is low for the particular 14 records used in our comparison. The cluster analysis scheme achieved 94.30% accuracy for 14 records, while the proposed scheme calculated 96.59% for the same 14 records accurately. It is also notable that the overall accuracy for the broader sample of 43 patients is 95.52%. This clearly demonstrates that the

TABLE VIII: COMPARISON OF PROPOSED AND THE CLUSTER ANALYSIS SCHEME [7]

	# of Records	Sensitivity %					Accuracy %
Scheme		N	L	V	R	A	
Cluster Analysis [7]	14	95.59	91.32	94.51	90.50	93.77	94.30
Proposed Scheme	14	99.76	99.04	88.70	93.02	99.91	96.59
Proposed Scheme	43	98.07	99.11	94.53	94.34	95.70	95.52

accuracy of our scheme is higher while detecting arrhythmia in a prominent number of persons. These findings also prove the efficacy and robustness of the proposed method.

Another major contribution is that the proposed scheme has a dynamic approach for feature selection, while the cluster analysis scheme has static features (i.e., the same four features were used in the analysis of all records). The proposed scheme has a maximum of five features. For some records, it utilizes all the five features, while for other records, it uses only three or four features. This dynamic feature selection and utilization make the proposed scheme well-suited for energy-aware environments like WBANs. Similarly, it also helps in improving the overall performance of the system by selecting the most appropriate features for the detection process based on the unique characteristics of the specific record being analyzed.

IV. CONCLUSION

The main purpose of this study was to identify different types of arrhythmia by using an automated system with a relevant and reduced set of attributes. The Mahalanobis-Taguchi System (MTS) was used to accurately analyze and classify patient ECG data records while minimizing unnecessary performance load and implementation cost by avoiding the use of irrelevant features. It is also designed to filter out the isolated factors which have the least impact on ECG beat classification. The proposed architecture can fit in real-time systems, where ECG sensor nodes attached to the human body can extract ECG readings based on time. The RR interval and QRS complex measure the feature values while the Mahalanobis Distance calculated using filtered feature values is used for future cardiac event prediction. The number of relevant feature set as the output of the MTS system varies from 3 to 5 from one patient to the next due to person-specific divergences in the ECG signal. By applying the proposed system on the MIT-BIH arrhythmia dataset, the sensitivity for Normal (NORM), LBBB, PVC, RBBB, and APC was calculated as 98.07%, 99.11%, 94.53%, 94.34%, and 95.70%, respectively. A separate statistical analysis confirms that the use of MTS for ECG beat classification yields better accuracy than other recent methods proposed by different researchers. In future, we intend to make the proposed scheme more robust and optimized by incorporating some dynamic threshold selection mechanism.

REFERENCES

[1] I. Christov, G. Gømez-Herrero, V. Krasteva, I. Jekova, A. Gotchev, K. Egiazarian, “Comparative study of morphological and time-frequency ECG descriptors for heartbeat classification”, Medical Engineering & Physics, vol. 28, no. 9, pp. 876-887, 2006.

[2] A. Koski, “Modelling ECG signals with hidden Markov models”, Artificial Intelligence in Medicine, vol. 8, no. 5, pp. 453-471, 1996.

[3] H. G. Hosseini, D. Luo, and K. J. Reynolds, “The comparison of different feed forward neural network architectures for ECG signal diagnosis”, Medical Engineering & Physics, vol. 28, no. 4, pp. 372-378, 2006.

[4] Y. Jongmin, F. Ibrahim, S. A. L. Narainasamy, and R. Omar, “Heart beat detection using estimated regular intervals from electrocardiography and blood pressure”, The 18th IEEE International Symposium on Consumer Electronics (ISCE 2014), 22-25 June, 2014, Jeju, Korea, pp. 1-2.

[5] R. Ceylan, Y. Ozbay, and B. Karlik, “A novel approach for classification of ECG arrhythmias: type-2 fuzzy clustering neural network”, Expert Systems with Applications, vol. 36, no. 3, pp. 6721-6726, 2009.

[6] A. De Gaetano, S. Panunzi, F. Rinaldi, A. Risi, and M. Sciandrone. “A patient adaptable ECG beat classifier based on neural networks”, Applied Mathematics and Computation, vol. 213, no. 1, pp. 243-249, 2009.

[7] Y.-C. Yeh, C. W. Chiou, and H.-J. Lin, “Analyzing ECG for cardiac arrhythmia using cluster analysis”, Expert Systems with Applications, vol. 39, no. 1, pp. 1000-1010, 2012.

[8] C. Su and T. Li, “A Mahalanobis distance based classifier for diagnosis of diseases”, Journal of Chinese Institute of Industrial Engineers, vol. 19, no. 5, pp. 41-47, 2002.

[9] G. Taguchi and R. Jugulum, “The Mahalanobis-Taguchi strategy: a pattern technology system”, John Wiley and Sons Press, New York, p. 256, 2002.

[10] E. A. Cudney, J. Hong, R. Jugulum, K. Paryani, K. Ragsdell, and G. Taguchi, “An evaluation of Mahalanobis-Taguchi system and neural network for multivariate pattern recognition”, Journal of Industrial and Systems Engineering vol.1, no. 2, pp. 139-150, 2007.

[11] Y. Lv, and J. Gao, “Condition prediction of chemical complex systems based on multifractal and mahalanobis-taguchi system”. Proc. International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering(ICQR2MSE), 17-19 June, 2011, Xi’an, China, pp. 536-539.

[12] A. Ali and F. A. Khan, “Energy-efficient cluster-based security mechanism for intra-WBAN and inter-WBAN communications for healthcare applications”, EURASIP Journal on Wireless Communications and Networking, vol. 2013, no. 1, pp. 1-19, 2013.

[13] A. Ali and F. A. Khan,“A broadcast-based key agreement scheme using set reconciliation for wireless body area networks”, Journal of Medical Systems, vol. 38, no. 5, pp. 1-12, 2014.

[14] E. Kartsakli, A. Antonopoulos, A. S. Lalos, S. Tennina, M. Renzo, L. Alonso, C. Verikoukis, “Reliable MAC design for ambient assisted living: moving the coordination to the cloud”, IEEE Communications Magazine, vol. 53, no. 1, pp. 78-86, 2015.

[15] A. Ali and F. A. Khan, “An improved EKG-based key agreement scheme for body area networks”, Proc. 4th International Conference on Information Security and Assurance (ISA 2010). 2010, Miyazaki, Japan, CCIS Vol. 76, pp. 298-308.

[16] S. Irum, A. Ali, F. A. Khan, H. Abbas, “A hybrid security mechanism for intra-WBAN and inter-WBAN communications”, International Journal of Distributed Sensor Networks, 2013:11, 2013, Article ID 842608.

[17] F. Ghavimi and H. Chen, “M2M communications in 3GPP LTE/LTE-A networks: architectures, service requirements, challenges, and applications”, IEEE Communications Surveys & Tutorials, vol. 17, no. 2, pp. 525- 549, second quarter 2015.

[18] J. C. Jiang and T. Y. Lin, “A comparison of the Mahalanobis-Taguchi system to a selective Naïve bayesian algorithm for semiconductor chemical Vapor deposition process”, AISS: Advances in Information Sciences and Service Sciences, vol. 5, no. 1, pp.720-729, 2013.

[19] MIT-BIH Arrhythmia Database. Available: http://www.physionet.org/physiobank/database/mitdb/ (last accessed 05 Feb., 2015)

[20] WFDB Software Package. Available: http://physionet.org/physiotools/wfdb.shtml#library (last accessed 05 Feb., 2015)

[21] WQRS command for QRS detection. Available: http://www.physionet.org/physiotools/wag/wqrs-1.htm (last accessed 31 Mar., 2015)

[22] Y. Jongmin, T. Jeon, and M. Jeon, “Heart beat detection using estimated regular intervals from electrocardiography and blood pressure”, The 18th IEEE International Symposium on Consumer Electronics (ISCE 2014), 22-25 June, 2014, pp. 1-2.