Classification Algorithm for Feature Extraction using Linear Discriminant Analysis and Cross-correlation on ECG Signals

Muhammad Fahad Shinwari, Naveed Ahmed, Hassan Humayun, Ihsan ul Haq, Sajjad Haider and Atiq ul Anam

Department of Engineering, International Islamic University, Islamabad, Pakistan
National University of Modern Languages, Islamabad, Pakistan

engr.mfahadshinwari@gmail.com, project_naveed@yahoo.com, hasanshah@numl.edu.pk, ihsanulhaq@iiu.edu.pk, sajjad@numl.edu.pk, atiqulaman@ciit.net.pk

Abstract

This paper develops a novel framework for feature extraction based on a combination of Linear Discriminant Analysis and cross-correlation. Multiple Electrocardiogram (ECG) signals, acquired from the human heart in different states such as in fear, during exercise, etc. are used for simulations. The ECG signals are composed of P, Q, R, S and T waves. They are characterized by several parameters and the important information relies on its HRV (Heart Rate Variability). Human interpretation of such signals requires experience and incorrect readings could result in potentially life threatening and even fatal consequences. Thus a proper interpretation of ECG signals is of paramount importance. This work focuses on designing a machine based classification algorithm for ECG signals. The proposed algorithm filters the ECG signals to reduce the effects of noise. It then uses the Fourier transform to transform the signals into the frequency domain for analysis. The frequency domain signal is then cross correlated with predefined classes of ECG signals, in a manner similar to pattern recognition. The correlated co-efficients generated are then thresholded. Moreover Linear Discriminant Analysis is also applied. Linear Discriminant Analysis makes classes of different multiple ECG signals. LDA makes classes on the basis of mean, global mean, mean subtraction, transpose, covariance, probability and frequencies. And also setting thresholds for the classes. The distributed space area is divided into regions corresponding to each of the classes. Each region associated with a class is defined by its thresholds. So it is useful in distinguishing ECG signals from each other. And pedantic details from LDA (Linear Discriminant Analysis) output graph can be easily taken in account rapidly. The output generated after applying cross-correlation and LDA displays either normal, fear, smoking or exercise ECG signal. As a result, the system can help clinically on large scale by providing reliable and accurate classification in a fast and computationally efficient manner. The doctors can use this system by gaining more efficiency. As very few errors are involved in it, showing accuracy between 90% - 95%.

Keywords: *ECG (Electrocardiogram), HRV (Heart Rate Variability), LDA (Linear Discriminant Analysis), Cross-correlation*

1. Introduction

According to medical definitions, electrocardiogram, ECG measures micro voltages in the heart muscle over a specific time period. It is composed of several parameters: P, Q, R, S and T waves [1].

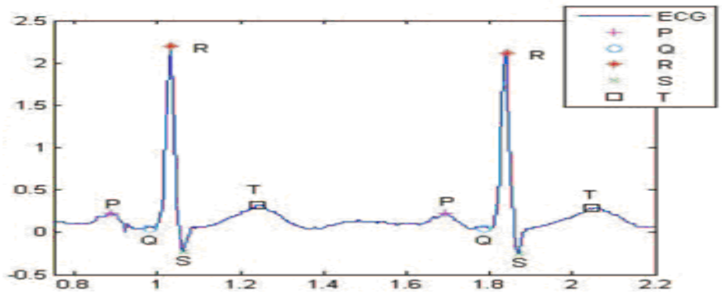


Figure 1. Illustrates a Typical ECG Waveform with P, Q, R, S and T Waves

ECG is most commonly used for identifying abnormal heart activity. The electrocardiogram wave is of paramount importance in clinical practice to assess the cardiac status of patients. The ECG signals are considered to be quasi periodic signals and extracting important information from it is quite difficult for doctors. Its Heart Rate Variability understanding requires experience and careful attention which might not be achieved efficiently at busy hospitals. There are more chances of errors due to patients load and inexperienced young doctors. Its incorrect interpretation could lead to life threatening situations.

Differences in normal and abnormal ECG signals can't be easily determined especially with human eyes. The difference between normal and abnormal ECG waveforms is shown below in Figure 2.

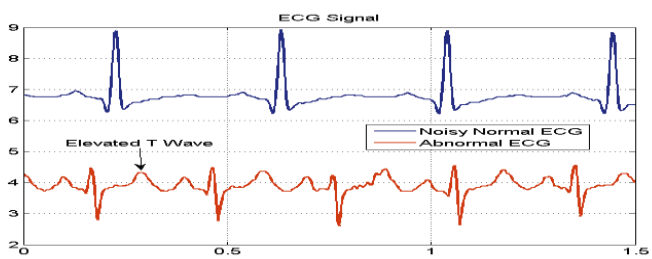


Figure 2. Normal Vs Abnormal ECG Waveforms

In normal ECG wave p, q, r, s and t waves are in normal shape despite of noise present in it. And the wave repeats itself after some constant predictable time. While abnormal ECG signal shows abnormal and disrupted shape of ECG parameters. The wave repeats itself with irregular time interval. And is not quiet matching with normal ECG waveform. Thus there we need a system which could allow the doctors to quickly extract the information from ECG accurately. So a precise and convenient system is needed that consumes less time and achieves higher accuracy.

Many complex algorithms and methods in signal processing field had been proposed and implemented in the past few years to extract features from ECG waveforms. Such as, improving Classification performance of Linear Feature Extraction [2], correlation dimension and largest Lyapunov exponent [3], using wavelet transform and time intervals methodology [4], multiple signal classification algorithm [5], and efficient formation of morphological wavelet transform features together with the temporal features of the ECG signal [6]. Also some research techniques are used to classify ECG signals including the following, support vector machines for cardiac beat detection [7] , Artificial Neural Network (ANN) for classification [8, 9], Fisher Linear Discriminate Analysis (FLDA) technique [10], selforganized map [11], heartbeat interval combined with the shape and morphological properties of the ECG parameters [12] etc.

These above techniques mentioned, persists of drawbacks. The main drawback of these techniques is the significant amount of computation and processing time for extracting the desired features and making classification on certain features. Therefore, in this paper, we propose an algorithm that is capable to autonomously extract features from an ECG wave. First of all we acquire data. The ECG is of lengthy time intervals and there exists different noises in it that may alter the parameters of ECG waveform. So we de-noise it with filtering method. And then apply different feature extraction techniques. Furthermore, manual analysis of ECG wave is to be considered time consuming and lead to errors scenarios. So the importance of automatic recognition and analysis of the ECG signals is solved. It provides quick and accurate (with less error) classification of multiple ECG signals (fear, normal, exercise and normal) by extracting different features from them.

For easy understanding the paper is divided in phases. In phase- I, our main cynosure is to understand and get familiar with the theme of this paper. And then in phase- II we discuss proposed architecture on the basis of data acquisition, noise reduction, transformation and dimensionality reduction. Phase- III allows us to know the feature extraction techniques based on Cross-Correlation and Linear Discriminant Analysis for classification of multiple ECG signals. Phase- IV deals with simulated results and discussions. And the last Phase- V consists of conclusion that is concluded upon the above research findings and results. In the last, certain future work is recommended for the researchers.

2. Proposed Architecture

The system architecture for the designed algorithm is shown in Figure 3. Data in the form of ECG signals are acquired at different state e-g fear, normal etc. and then are processed. The signals are de-noised by filtering techniques. And Fourier transform is applied to analyze data in frequency domain. Those Fourier points are used for simulation purpose to extract important features from ECG signals. For clear and easy analysis, each signals dimensions are reduced. Feature Extraction techniques are applied to determine the unknown ECG signal.

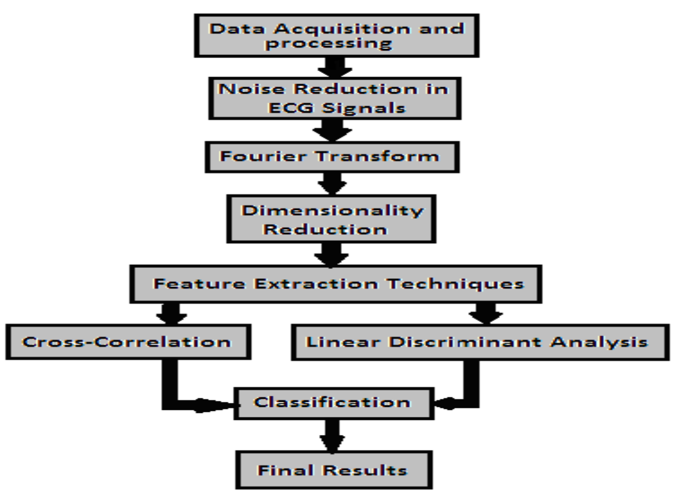


Figure 3. Block Diagram Illustrating the Designed Algorithm

2.1. Data Acquisition and Processing

ECG signals were acquired at different states i-e fear, normal, smoking etc. with the help of "Alive Heart and Activity Monitor". Its an medical equipment designed to record ECG signals in real time domain. It uses wireless Bluetooth technology to transmit ECG to a computer or central monitoring center etc. [13]. The sampling frequency of this electronic cardio-equipment was set to 300 Hz, satisfying the Nyquist criteria. As the maximum frequency of ECG signal is nearer to 140 Hz. The ECG signals were acquired from three subjects. Normal, exercise, smoking and fear ECG signals were acquired from each subject. The ECG signals were transmitted to computer. The signals were converted to ".mat" extension, so it can be simulated in matlab software. The signals are processed and then plotted.

2.2. Noise Reduction

ECG is an electrical activity of the heart that carries of carries various kinds of noises such as power-line interference, baseline wander, shot noise, electromyography noise (EMG), electrode motion artifacts etc. [14,15]. The noises are given in Table 1.

Table 1. Different Noises in ECG Signals

Qualities Contaminant	Frequency Range	Time Duration
Power line	Narrowband 50 or 60±2Hz	Continuous
Baseline Drift	0.15Hz-0.3Hz	Transient or Continuous
EMG Noise	10K Hz	50 ms
Electrosurgical Noise	100Khz-1Mhz	1-10 sec
Electrode Pop	Narrowband	1 sec
Electrode Motion Artifacts	Broadband	0.5 sec
Quantization Noise	Broadband	Continuous

These noises alter the ECG signal parameters (p, q, r, s and t waves). And the information couldn't be easily extracted. There was a need to de-noise the weak ECG signals contaminated with several noises. So the subtle and important information couldn't be lost in the ECG waveform. For such specific purpose, we simply use Band-pass filter, and its proposition deals with different types of noises to greater extend. We simulate to reject certain noises and find less noisy ECG signal with SNR (Signal to Noise Ratio) increased to greater extend. At last, ECG signal is de-noised with optimum results. Band-pass filter, with the upper and lower cut-off frequencies are set to 0 and 0.8/fs respectively, to de-noise majority of noises. After band pass filter we obtain clean ECG signal. And it shows the optimal effect with less computation.

2.3. Dimensionality Reduction and Transformation

From each corresponding signal, first 1000 samples were selected. In this way, dimensions were reduced from ECG waveform. It is done for easy understanding and feature extraction.

Then they are converted from time domain to frequency domain with the help of Fourier transform. In time domain information can't be extracted, so we transform the signals to frequency domain and analyze the signal. And further we see pedantic changes for comparison with each other after simulations.

3. Feature Extraction Techniques

The proposed architecture is a novel implementation of various techniques based on the state-of-the-art of Linear Discriminant Analysis and cross correlation (pattern recognition) process, with the help of these techniques, the algorithm is enhanced for feature extractions methodology. And could be used as multi-function purpose. It distinguishes and determines different classes of ECG signals in a more convenient way.

Here we proposed two techniques for feature Extraction. They are discussed below.

3.1. Cross-correlation

Cross correlation is an excellent tool to match images and signals with each other. It is robust to noise, and can be normalized for pattern matching. It's a measure of similarity of two waveforms as a function of a time-lag applied to one of them.

The mathematics for discrete functions, f and g, is defined as:

$$(f * g)[n] \stackrel{\text{def}}{=} \sum_{m=-\infty}^{\infty} f * [m] g[n+m].$$

Where f * represents the complex conjugate of f.

The cross-correlation of two quantities is similar to the convolution of two functions [16]. So here we use cross correlation as a pattern recognition technique, in which any unknown input ECG signal was cross-correlated with all the predefined ECG signals (exercise, fear, smoking and normal). It would determine its specified class (exercise, fear, smoking or normal) on the basis of co-efficient values. The thresholds are decided for each certain group of signal with simple "if else" function. After cross correlating unknown input ECG signal, coefficients are generated for each cross-correlation. All the signals are normalized between range of 0 and 1. The co-efficient value for each cross correlation would possess highest value of 1 and minimum value of 0. So after passing through cross correlation, the coefficients' values amongst all cross correlated coefficients are taken in account. The value which is nearest to "1" or is highest, it is considered to be its specified class upon the value occurring in certain threshold range as defined for each signal. And thus the input unknown ECG signal is determined that it belongs to which category (exercise, fear, smoking or normal).

3.2. Linear Discriminant Analysis

Then Linear Discriminant Analysis is used as a feature extraction technique. LDA is used to classify objects into groups, based on the set of features that describe the specific object. In our case we use multi-class Linear Discriminant Analysis to classify multiple unknown ECG signals into multiple groups (exercise, smoking, normal and fear) based on the feature of calculations of mean, global mean, mean subtraction, transpose, covariance, probability, frequencies and at the end defining thresholds for each class on the distributed space area [17] .

The mathematical Derivation of the LDA is given below. And it is described by [18].

$$d_i(x)\ln(P(C_i)) + x^T C^{-1} m_i + \frac{1}{2} m_i^T C^{-1} m_i$$

Where m_i is the N length of mean vector for ith class

C_i is the $N \times N$ covariance matrix for the ith class,

$P(C_i)$ is the prior probability of class C_i.

The selected class is the one that has the highest value of $d_i(x)$.

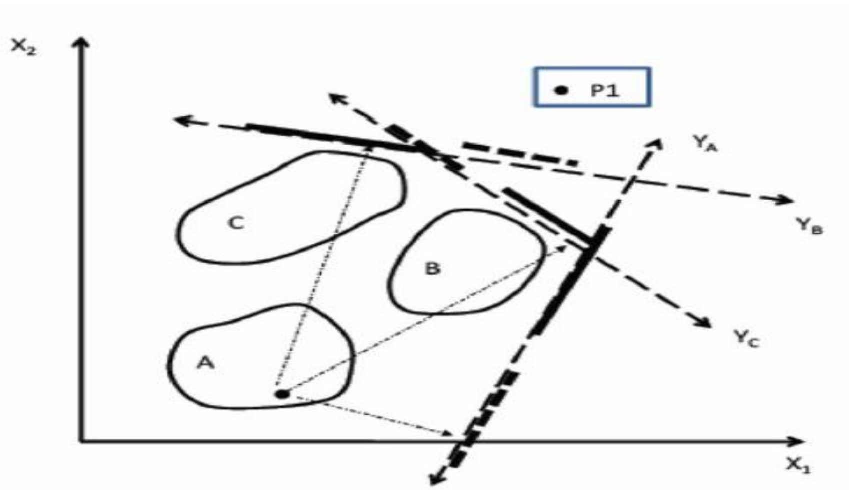


Figure 4. Multiple Classes (A, B and C) of Different Groups on Distributed Space Area after Passing through Multi-class LDA

In multi-class LDA process, the covariance for each ECG signal shows that the input signal differs from the predefined ECG signal. And critical points are placed in Rejected matrix of critical points and accepted matrix of critical points during probabilistic approach. Moreover different frequencies present in the ECG signals, are plotted according to their frequencies on the distributed space area. And then two threshold levels are defined along each x-axis and y-axis for each group or class (exercise, smoking, normal or fear ECG signal). Thus displaying the corresponding name of the unknown ECG signal after passing through LDA process.

4. Simulation Results and Discussions

After simulation, we get following results with the help of feature extraction technique of Cross-correlation.

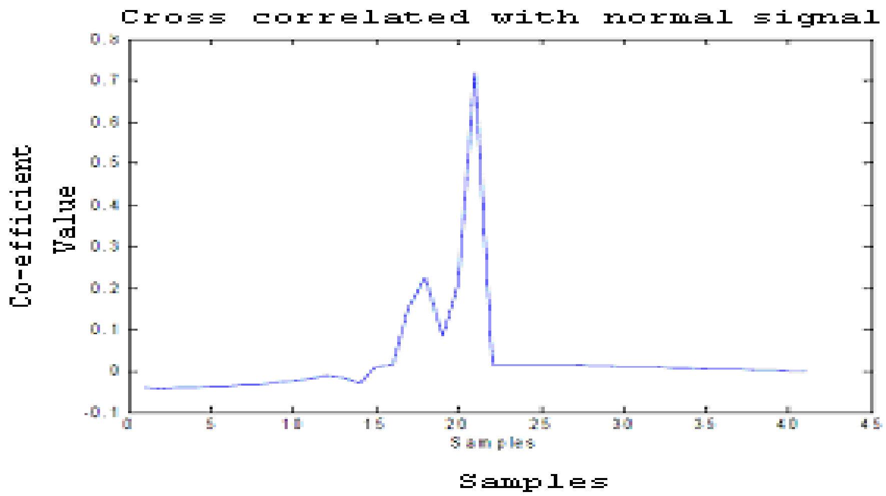


Figure 5. Unknown ECG Signal Cross-correlated with Predefined Normal Signal

When any unknown ECG signal is cross correlated with predefined normal signal, it generates a coefficient of certain value. The value lies between domain of 0 to 1. In Figure 5, we can see along y-axis that the peak value (coefficients value) of cross correlated output signal is equal to 0.7.

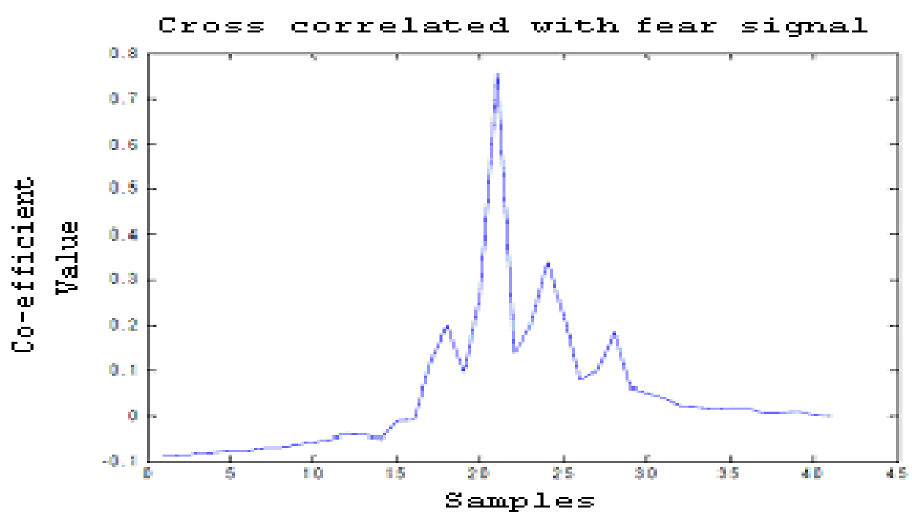


Figure 6. Unknown ECG Signal Cross-correlated with Predefined Fear Signal

In above Figure 6, unknown ECG signal is cross correlated with predefined fear signal, again it generates a coefficient of value 0.75 along y-axis.

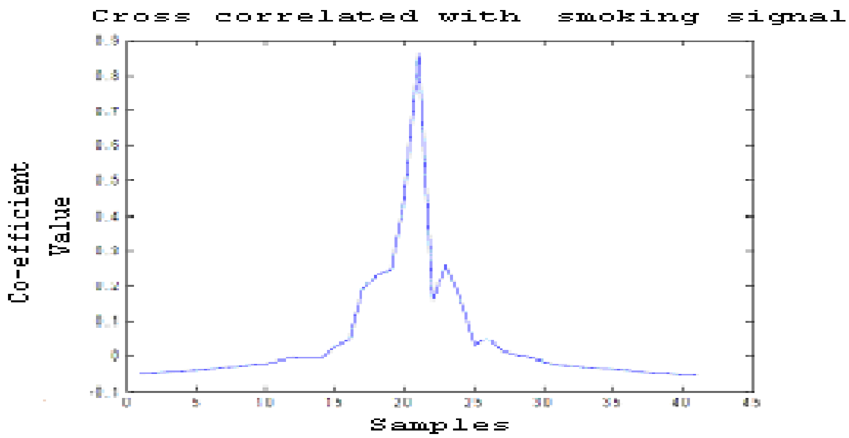


Figure 7. Unknown ECG Signal Cross-correlated with Predefined Smoking Signal

Also in Figure 7, the signal is cross correlated with predefined smoking signal. It generates coefficient of 0.85 at y- axis.

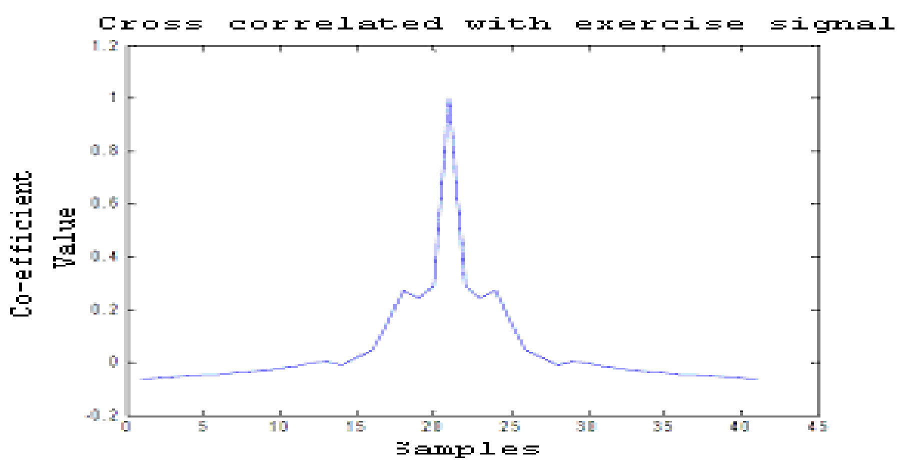


Figure 8. Unknown ECG Signal Cross-correlated with Predefined Exercise Signal

In Figure 8, same process is repeated and we observe the greater value coefficient along y-axis is present that is equal to 1. It is the maximum value generated after cross correlation of unknown ECG signal with predefined exercise signal. The value is maximum amongst all cross correlated coefficients, so it shows maximum pattern is matched. Thus the unknown signal is considered to be exercise signal.

Further simulation, gives the Linear Discriminant Analysis output graph on Distributed Space Area. Following results help in feature extraction technique for classification.

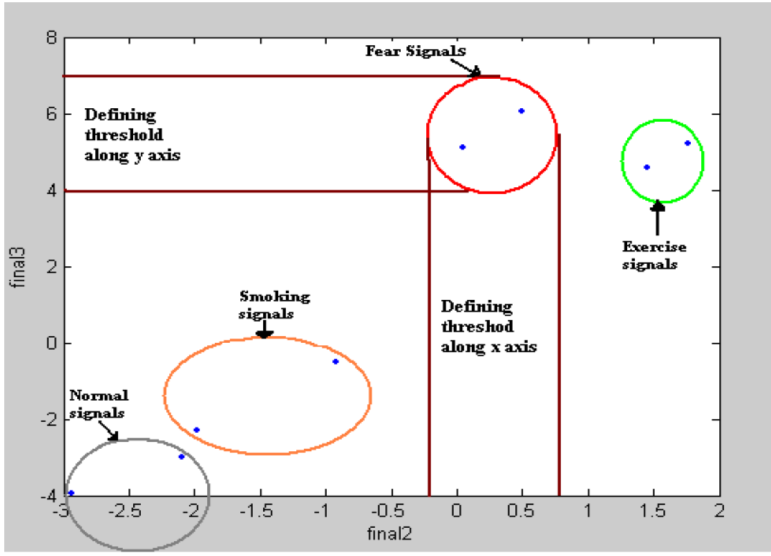


Figure 9. Output Graph after Linear Discriminant Analysis, it shows Multiple classes (exercise, fear, normal and smoking) of different groups on Distributed Space Area after Passing through Multi-class LDA Each Class is Defined by Thresholds along x and y-axis

In above Figure 9, the output graph of Linear Discriminant Analysis is shown. In which feature extraction technique is used (LDA). After giving any unknown ECG signal it plots the signal in the predefined class. Classification is done by making multiple classes on several set of features as discussed earlier in our feature extraction techniques. In case of error, the signal would lie outside the class boundary area. And the algorithm would be unable to distinguish the given signal. If such case occurs then it is counted as an error.

As defined for each signal. And thus the input unknown ECG signal is determined that it belongs to which category (exercise, fear, smoking or normal).

4.1. Improvement Achieved by the Proposed Framework

As earlier algorithms were proposed to determine only normal and abnormal ECG signals. So these primitive algorithms were constrained to certain limitations. But our designed system is more improved algorithmic system in a sense, that it not only distinguishes between normal and abnormal ECG signals. It also determines unknown multiple ECG signals (exercise, fear and smoking ECG signals) comparatively to normal ECG signal. The exercise, fear and smoking signals are to be considered abnormal ECG signals. Thus it further elaborates the abnormal ECG signals into multiple classes. And the methodology adopted is simple compared to other algorithms [19].

4.2. Accuracy

Our algorithm is more efficient because it achieves greater accuracy with fewer errors. As empirical approach was adopted, by giving numerous ECG signals to the system. The Features were remarkably extracted from multiple signals through LDA and Cross-correlation techniques. The accuracy results for Cross-correlation and LDA are shown below in Figure 10 and Figure 11. Both achieve overall accuracy of 92%.

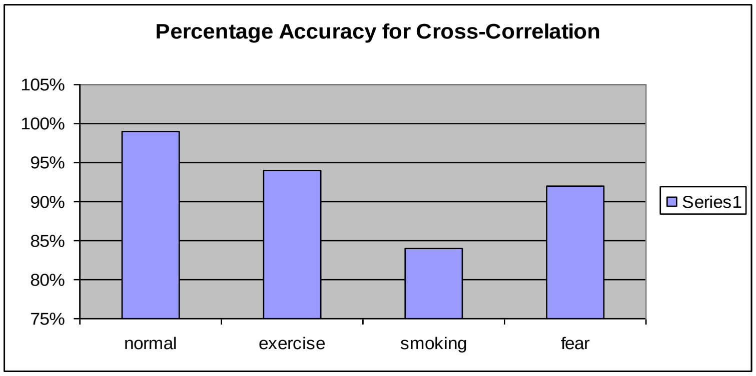


Figure 10. Graph Showing Accuracy of the Cross-correlation Algorithm for each Category of ECG Signal

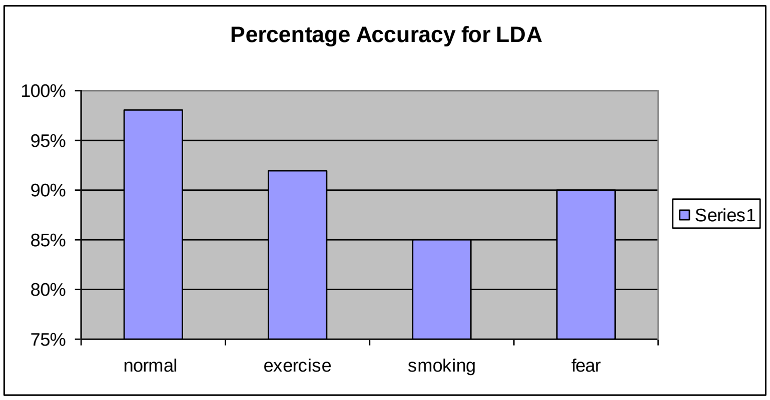


Figure 11. Graph Shows Accuracy of the LDA Algorithm for each Category of ECG Signal

5. Conclusion

Thus in this paper novel framework is achieved and an algorithm based system is designed. It involved feature extraction techniques like Linear Discriminant Analysis and Cross-correlation. The doctors can now quickly extract the information from ECG signal from the heart patient with less reading errors. It would also reduce the doctors work load. So a precise and convenient system was achieved that consumes less time and achieves higher accuracy between 90-95 %

The Future works recommended are that, one can revive and enhance our algorithmic system in the following ways:

a) Increasing the number of classes.

b) Introduce Graphical User Interface (GUI) system in the above algorithm for low tech users, by making it compatible with hospital Personal Computers. It would provide easy understanding for doctors.

c) Design a micro controller based electronic hardware that displays the result on the screen after simulation.

Acknowledgements

I am deeply indebted to my research contributors Assistant Prof. Dr. Tariq Qaisrani, Assistant Prof. Sajjad Haider and Assistant Prof. Hassan Shah. Their continuous help, stimulating suggestions and experience helped me in all the times of study and analysis of the research paper in the pre and post research period. I am also extremely grateful to Assistant Prof. Naveed Ahmed and Assistant Prof. Atiq ul Inam, as they went to extremes, to deliver their best by giving all of their precious time, knowledge and encouragement. It was exquisitely good learning experience while working under them.

References

[1] R. M. Rangayyan, "Biomedical Signal Analysis: A Case-Study Approach", Wiley–Interscience, New York, **(2001)**.

[2] R. Acharya, J. S. Suri, J. A. E. Spaan and S. M. Krishnan, "Advances in Cardiac Signal Processing", ISBN-13 978-3-540-36674-4, Springer Berlin Heidelberg New York, **(2007)**.

[3] M. Owis, A. Abou-Zied, A. B. Youssef and Y. Kadah, "Robust feature extraction from ECG signals based on nonlinear dynamical modeling", 23rd Annual International Conference IEEE Engineering in Medicine and Biology Society, (EMBC'01), vol. 2, **(2001)**, pp. 1585-1588.

[4] O. T. Inan, L. Giovangrandi and G. T. A. Kovacs, "Robust neuralnetwork- based classification of premature ventricular contractions using wavelet transform and timing interval features", IEEE Transaction on Biomedical Engineering, vol. 53, no. 12, **(2006)**, pp. 2507-2515.

[5] A. R. Naghsh-Nilchi and A. R. K. Mohammadi, "Cardiac Arrhythmias Classification Method Based on MUSIC, Morphological Descriptors, and Neural Network", EURASIP Journal on Advances in Signal Processing, vol. 2008, Article no. 202, **(2008)**.

[6] T. Ince, S. Kiranyaz and M. Gabbouj, "A Generic and Robust System for Automated Patient-specific Classification of Electrocardiogram Signals", IEEE Transactions on Biomedical Engineering, vol. 56, no. 5, **(2009)** May.

[7] S. S. Mehta and N. S. Lingayat, "Support Vector Machine for Cardiac Beat Detection in Single Lead Electrocardiogram", IAENG, International Journal of Applied Mathematics, **(2007)**, pp. 1630-1635.

[8] B. M. Z. Asl and S. K. Setarehdan, "Neural Network Based Arrhythmia Classification Using Heart Rate Variability Signal", Proceedings of the 2nd International Symposium on Biomedical Engineering, Bangkok, Thailand, **(2006)** November.

[9] B. Anuradha and V. C. V. Reddy, "ANN for classification of cardiac arrhythmias", ARPN Journal of Engineering and Applied Sciences, vol. 3, no. 3, **(2008)** June.

[10] R. P. W. Duin and M. Loog, "Linear dimensionality reduction via a heteroscedastic extension of lda: the chernoff criterion", IEEE Trans. PAMI, vol. 26, no. 6, **(2004)** June, pp. 732–739.

[11] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt and L. Sörnmo, "Clustering ECG Complexes Using Hermite Functions and Selforganizing Maps", IEEE Transaction on Biomedical Engineering, vol. 47, no. 7, **(2000)** July, pp. 838-848.

[12] P. de Chazal, M. O'Dwyer and R. B. Reilly, "Automatic Classification of Heartbeats Using ECG Morphology and Heartbeat Interval Features", IEEE Transaction on Biomedical Engineering, vol. 51, no. 7, **(2004)** July, pp. 1196- 1206.

[13] www.alivetec.com.

[14] Y. Der Lin and Y. Hen Hu, "Power-line interference detection and suppression in ECG signal processing", IEEE Trans. Biomed. Eng., vol. 55, **(2008)** January, pp. 354-357.

[15] N. V. Thakor and Y. S. Zhu, "Applications of adaptive filtering to ECG analysis: noise cancellation and arrhythmia detection", IEEE Transactions on Biomedical Engineering, vol. 38, no. 8, **(1991)**, pp. 785-794.

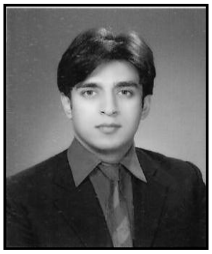
[16] http://en.wikipedia.org/wiki/Cross-correlation.

[17] L. J. Hargrove, E. J. Scheme, K. B. Englehart and B. S. Hudgins, "Multiple Binary Classifications via Linear Discriminant Analysis for Improved Controllability of a Powered Prosthesis", IEEE Trans. Neural Syst. Rehabil. Eng., vol. 18, no. 1, **(2010)** February, pp. 49-57.

[18] R. Duda, P. Hart and D. Stork, "Pattern Classification", 2nd ed. New York: Wiley- Interscience, **(2000)**.

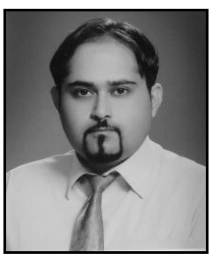
[19] K. Waseem, A. Javed, R. Ramzan and M. Farooq, "Using Evolutionary Algorithms for ECG Arrhythmia Detection and Classification", Seventh International Conference on Natural Computation, **(2011)**.

Authors



Engr. Muhammad Fahad Shinwari, serving as an Lecturer at the Department of Engineering in National University of Modern Languages (NUML), Islamabad, Pakistan since 2011. I completed my Bachelors in Electronics Engineering from COMSATS Institute of Information Technology, Abottabad in 2010. Currently, engaged in research of Masters in Electronics Engineering from International Islamic University, Islamabad. My research interest lies in the signal processing, power and energy fields. Recently, my several papers got accepted in International Journals and International Conferences, which are in publication process.

Naveed Ahmad is working as Assistant Professor in Faculty of IT/Engineering at National University of Moderrn Languages, Sector H-9 Islamabad since 2006. He did his MS in Computer Engineering from COMSATS Institute of Information Technology Islamabad in 2005.His research interests includes Ad hoc wireless sensor networks and disaster Management. Currently he is also involved in Research work on Wireless Ad hoc Sensor Networks for Disaster Management.



Hassan Humayun is working as Assistant Professor in Faculty of IT/Engineering at National University of Modern Languages Sector H-9 Islamabad since 2007. He did his MS in High Speed Mobile and Networks from Oxford Brooks University, Oxford UK in 2006 .His research interests includes Computer Networks and wireless networks for Mobile communication.

Sajjad Haider is Assistant Professor at the Department of Information Technology in National University of Modern Languages (NUML), Islamabad, Pakistan. He holds Bachelors and Masters Degrees in Computer Science and is a PhD Scholar in Shaheed Zulfiqar Ali Bhutto Institute of Science and Technology (SZABIST), Islamabad Campus. Before joining NUML in 2002, he was working as Network Manager in National Institute of Electronics (NIE), a Research and Development Organization working under the Ministry of Science and Technology, Government of Pakistan.

Sajjad's research interests lie mainly in the area of Parallel and Distributed Computing, High Performance Computing, Fault Tolerant Computing and Network Security. He has many publications in National and International Conferences and in refereed Journals. He is also the paper reviewer of many International Conferences and is a member of IACSIT and IAENG.

Engr. Atiq-ul-Inam is Assistant Professor at the Department of Engineering in Comsats Institute of Information and Technology, Abottabad, Pakistan since 2003. He holds Bachelor's Degree from De La Salle University, Manila in Communication Engineering and possess double Master Degrees in Computer Engineering from military college of Engineering, Risalpur. His research interests include signal processing and communication. He has many publications in National and International Journals. And he is also paper reviewer of many International Conferences.

Dr. Ihsan-ul-Haq is serving as Assistant Professor in Department of Engineering in International Islamic University, Islamabad, Pakistan. He completed his Phd in (Electronic and Information Engineering) from Beihang University,Beijing China. He has research papers published in international journals and conferences. Currently his research area lies in communication and signal processing.