International Journal of Power Control Signal and Computation(IJPCSC)
Vol3. No1. Jan-Mar 2012 ISSN: 0976-268X
www.ijcns.com

Mobile Based CVD Detection and Diagnosis from Compressed ECG by K-Means Clustering

K.Dhivya[1] R.Deepa[2]
[1]M.E Embedded Systems, [2]Assistant Professor/EIE,
Bannari Amman Institute of Technology,
Sathyamangalam,TamilNadu,India.
[1]divyaece89.k@gmail.com, [2]rangarajdeepa@gmail.com

Abstract

This paper deals with the innovative data mining technique in the detection of CVD from compressed ECG a real time approach. The decompression of ECG signal and diagnosis of disease cause delay which will lead to death of the patient. Innovative technique with data mining that performs real-time classification of CVD from compressed ECG packets has been demonstrated. By this real time application any cardiac abnormalities found can be informed to emergency by means of SMS/MMS/e-mail automatically. The proposed system uses data mining techniques, such as attribute selection from the compressed ECG packets and K-based clustering. A set of constraints are generated in the hospital server for each abnormalities. These constraints are received by patients mobile phone and abnormal beats are identified in real-time. This innovative data mining technique on compressed ECG packets enables faster identification of cardiac abnormality directly from the compressed ECG, helping to build an efficient telecardiology diagnosis system.

Index Terms— Cardiac abnormality detection, faster cardiovascular diagnosis, m-health, medical data mining, mobile telecardiology.

1. INTRODUCTION

In a medical environment, there are several signals which must be constantly or periodically supervised. Some of the most common are the temperature, the concentration of oxygen in blood, the arterial pressure or the electrocardiogram waveform. It is under this scenario that this thesis is developed. In this case, there is an implemented system of acquisition of electrocardiogram (ECG) and phonocardiographic (PCG) signals, which must be wirelessly and error-free sent to the required medical location.

The introduction of telecommunication technologies in the health care environment has led to an increase in the accessibility to health care providers, to more efficient tasks and to a higher overall quality of health care services. However, many challenges including medical errors and a partial coverage of health care services in rural and underdeveloped areas still exist worldwide.

Many medical errors occur due to a lack of correct and complete information at the location and time it is needed, and it may result in wrong diagnosis. The required medical information can be made available at any place any time using sophisticated devices and widely deployed wireless networks. Nevertheless, wireless technologies cannot avoid or eliminate all medical errors, as some of them might have been originated before sending the information.

In order to avoid possible errors while compressing data before sending it through the wireless network, there are some existing algorithms for lossless compression of ECG signals where the original ECG waveform can be exactly reconstructed after the procedures of compression, transmission and decompression. Moreover, after the phase of compression of the required medical signals, wireless technologies can be effectively used by matching infrastructure capabilities to health care needs.

ECG signal has been intensively used by cardiac specialists to efficiently diagnose cardiovascular diseases (CVD) for the last seven decades. Apart from diagnosing CVD, ECG is also used for monitoring breathing pattern, mental stress, and condition of autonomous nervous system. In addition to monitoring different physiological states, ECG can also reveal the identity of a person using ECG-based biometric techniques. CVD being the number one killer of the modern era, researchers are providing wireless cardiovascular monitoring facilities to save life. Since ECG signals are enormous in size, usage of compression technology makes whole telecardiology-based diagnosis faster and efficient.

A faster solution is of crucial importance for diagnosis and treatment of CVD, as delay of every second counts toward patient's mortality. Even though ECG compression enables faster transmission, it introduces a slight delay as the compressed ECG needs to be decompressed before performing any diagnosis.

One of the major benefits of using digital information is that it can be compressed. The goal when compressing and transmitting data over the network is to make the data sent as small as possible. The smaller the data is, the faster it can be transmitted over the network. Moreover, apart from decreasing the size of the original data, as much of the original information as possible must be retained when dealing with medical information. Since it can be required to record an ECG signal during 24 hours, the computer storage may arise up to several GBytes. Considering the several million ECGs annually recorded for the purposes of comparison and analysis, the need for effective ECG data compression techniques is becoming increasingly important.

There are two types of compression: lossless and lossy compression. In lossless data compression, the

original data can be exactly reconstructed while the lossy approach always involves a loss of information. Due to the diagnostic uses of medical images, and since a small detail may be very important, medical image compression techniques have primarily focused on lossless methods.

There are many available algorithms for lossless compression, and each algorithm has several variants. With so many choices, it is important to select the algorithm that better it's the requirements of the system, as they form the basic specifications for the system. By following these specifications, the system should work as desired for this application. Even though this kind of methods allows the identical reconstruction of the data, lossless methods can only provide limited compression factors, usually ranging between 1:2 and 1:3.7.

To achieve better compression rates, one must know how data is structured and which compression method is the most appropriate to use. More information about how the data is organized gives higher probabilities to achieve reliable results. In section II the system to be implemented and its methodology is presented, in section III its software implementation has been shown, in section IV hardware description has been explained, results have been shown in section V and section VI gives the conclusion.

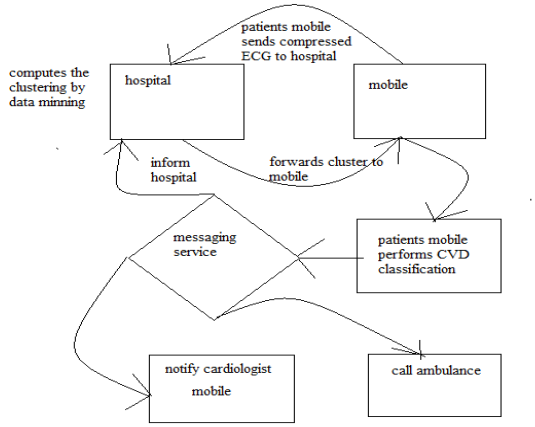


Figure 1 Mobile phone alert mechanism

2. SYSTEM AND METHOD

The patients mobile phone receives the ECG signal from the ECG sensors attached to the patient's body by bluetooth protocol. The received ECG packets are compressed by mobile phone before transmission for efficient transmission. The compressed packets are send to the hospital server by public network. If the data's are send in continuous manner it generates traffic and causes internet expenses and computational burden on hospital server.

After receiving the compressed ECG packets, the hospital server performs the disease recognition task with data mining techniques such as attribute selection and K-based clustering. It should be mentioned that the attributes of the attribute selection process are the frequencies of different characters used for encoding the original ECG to compressed ECG. After the reduction of attribute set (with attribute selection process), K performs the clustering in the hospital server and generates a range for the attribute set. These ranges of attributes determine the affinity of a compressed packet toward different clusters (e.g., normal, premature ventricular contraction (PVC), atrial fibrillation, atrial premature beat (APB), etc.). these cluster ranges are sent to the patient's mobile phone, so that the patient's mobile phone can perform efficient real-time recognition of a particular disease with a rule-based system. If the mobile phone recognizes any life-threatening disease, it can directly notify emergency services for initiating lifesaving protocols.

A. Attribute Subset Selection on Hospital Server

The feature subset selection process reduces the dimensionality of the data to be analyzed, speeds up execution of learning algorithms, improves the performance of data mining techniques (e.g., learning time, predictive accuracy, etc.), and improves the comprehensibility of the output.

In this paper, we have adopted a correlation-based feature subset (CFS) selection technique. CFS considers the following criteria for performing attribute selection task for our telecardiology scenario:

1) The attribute's affinity (or utility) toward a particular class (i.e., what frequency/value range for a particular character can allocate a particular compressed ECG to a specific disease).

2) The attribute's correlation with other attributes (i.e., if the value ranges for character *a*, *b*, and *c* have the same impact for allocating a compressed ECG packet to be under five different CVD clusters, then we might just consider character *a* for attribute selection, as having *b* and *c* would be redundant).

The utility of an attribute can be represented using Pearson's coefficient for correlation, where the variables are standardized as in

$$r_{xy} = \frac{\Sigma(x_i-\overline{x})(y_i-\overline{y})}{(n-1)\sigma_x\sigma_y} \quad (1)$$

$$Us = \frac{C\overline{r_{ap}}}{(C+C(C-1)\overline{r_{aa'}})^{1/2}} \quad (2)$$

where *xi* and *yi* are the sample mean calculated from the data, σx and σy are the standard deviations, a, $a \in S$,$C \leftarrow |S|$, $rxy \leftarrow$ average correlation between features x and y . For a subset S of C features, the utility function calculates how much the features(a, $'a$) are related rap to the predicted class p, while being less correlated with each other $ra'a$. The utility function minimizes the effect of irrelevant attributes as they are less correlated with the predicted class. It also omits redundant attributes as they are highly correlated with each other. During experimentation, the attribute selection process was executed on the full dataset. Since there are $2n$ possible subsets from n attributes, a heuristic that executes fast and uses limited computational resources needs to be sought. Therefore, a greedy best first algorithm to search through the candidate subsets for a locally optimal solution was utilized. The algorithm initiates its operation with an empty

subset, adding one attribute at a time and evaluating the utility function, to determine the correlation of the subset with the predicted class. The next attribute is added as long as the utility value does not reduce for the best subset. If there is a reduction, then the algorithm selects the next best subset and commences adding attributes to it. In some datasets, where there are groups of features that are locally predictive to the predicted class, the attributes that were initially discarded while building the best subset need to be investigated further. In this case, after the best subset has been generated, the algorithm thoroughly looks into the rejected list of attributes and compares its correlation with the predicted class against the average correlation with the subset. If its correlation with the class is higher than its correlation with the attribute subset, indicating a stronger affinity to the class than the subset, then the attribute is included in the subset.

Using the smaller subset of attributes, we can now produce clusters from the normal as well as different abnormal compressed ECG patterns. After the cluster formation task, the hospital server can determine the cluster mean and the cluster ranges. It should be mentioned that the process given in this paper works solely on the compressed ECG character frequency, and does not even require decompression, which would take valuable extra time from the patient's life.

B. Cluster Formation on Hospital Server

Clustering process sorts a set of objects in a way that similar objects remain under specific groups (or clusters). Although there are several available techniques to build multidimensional clusters, we have chosen a statistical clustering technique called K-means to cluster compressed ECG data, since it can be used to find the correct number of clusters automatically.

Clustering algorithms can be classified into 4 broad categories, that is, exclusive, overlapping, hierarchical and probabilistic. In exclusive clustering, data is classified such that no single element belongs to two different groups. Overlapping clustering on the other hand is the exact opposite, where an element can belong to different clusters. Hierarchical clustering classifies data into groups in a hierarchical way with some clusters being a part of another. Probabilistic clustering is beyond the scope of this text. Among these categories the k-means clustering algorithm belongs to the exclusive clustering variant.

The K-Means clustering algorithm is an effective exclusive procedure to cluster 'N' M-Dimensional data into 'K' clusters. The algorithm starts with 'K' points selected at random, chosen as the centroids for the data set. Once this is accomplished, each point in the sample space, is associated to the centroid whose distance to the current point is minimum. After which the winning centroid re-adjusts its position to be the mean of its member points. This process is repeated iteratively till convergence. If cj ; j = 0; 1; :::K 1 are the centroids where cjm is the m'th dimension of the centroid element cj , and xi; i = 0; 1; :::N 1 be points in the sample space where xim being the mth dimension of the point xi then the euclidean distance dij between these two points is where xkm is the m'th dimension of the k'th element associated to the centroid cj and Nj being the number of elements associated to centroid cj.√This processes is continued for all points in the sample space, till no point changes its association (Convergence) or a maximum number of iterations are reached.

Given a set of observations ($\mathbf{x}_1$, $\mathbf{x}_2$, …, $\mathbf{x}_n$), where each observation is a *d*-dimensional real vector, *k*-means clustering aims to partition the *n* observation into *k* sets ($k \le n$)$\mathbf{S} = \{S_1, S_2, \ldots, S_k\}$ so as to minimize the within-cluster sum of squares (WCSS):where $\boldsymbol{\mu}_i$ is the mean of points in S_i. Standard algorithm for K clustering. The most common algorithm uses an iterative refinement technique. Due to its ubiquity it is often called the ***k*-means algorithm**.

The algorithm is given for initial set of K means.

Assignment step: Assign each observation to the cluster with the closest mean.

Si^(t)={Xj:||Xj-mi^(t)||≤||Xi-mi*^(t)|| for all i*=1,...k}

Update step: Calculate the new means to be the centroid of the observations in the cluster.

$$m_i^{(t+1)} = \frac{1}{S_i^{(t)}} \sum_{x_j \in S_i^{(t)}} x_j \qquad (3)$$

The algorithm is deemed to have converged when the assignments no longer change.

3. SOFTWARE IMPLEMENTATION

The software implementation for the proposed system is done by MATLAB.As per the block diagram data acquisition is done by generating ECG signals. The signals are then compressed by means of runlength encoding. The compressed ECG is transmitted via public network. These are the part of simulation works done in patients side. In the other part of simulation codes are done for extract the data from compressed ECG by attribute generation and selection. The extracted data's from the ECG are clustered. K-means clustering is used here in order to avoid the use of reduntand data's.After clustering the disease are classified from the obtained set of clustered data. In this paper three CVD diseases are identified and its compression and clustering methods are demonstrated by MATLAB.

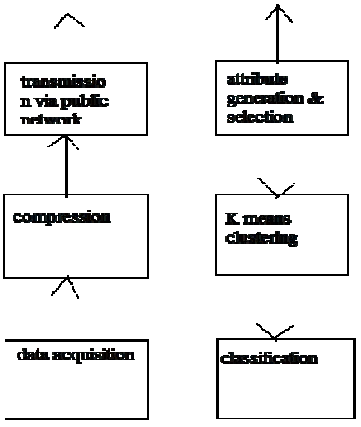


Figure 2 Block diagram for MATLAB simulation

4. HARDWARE DESCRIPTION

The hardware included in this paper demonstrates the working of the automatic health monitoring system. It consists of transmitter and receiver section(ie.patient and hospital server). It consists of PC with matlab interface for generation of three abnormalities and compression on transmitter side. Microcontroller used is 8051 used for transmit the data from PC to particular receiver. The GPS is used to identify the location of the patient. The Compressed ECG along with the position of the position is transmitted through GSM modem.RS232 is used to connect PC, microcontroller with GSM and GPS.

The receiver side has a GSM modem to receive the data and a PC to perform clustering from compressed ECG. The microcontroller here is to identify the mobile number whether it matches with the database. The clustered data is send to the transmitter (ie,mobile phone) in which the disease is classified by rule based algorithm and alert is given in case of emergengy to ambulance and cardiologist. The LCD here displays the type of disease.

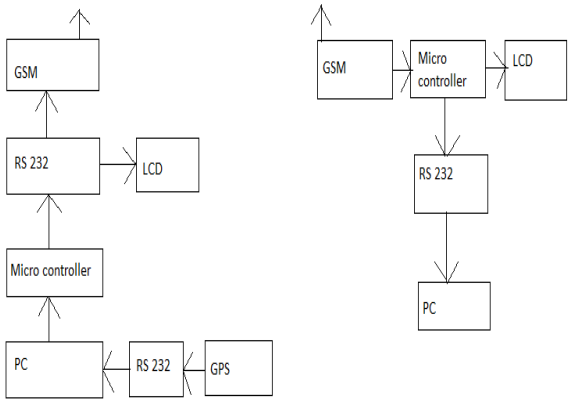
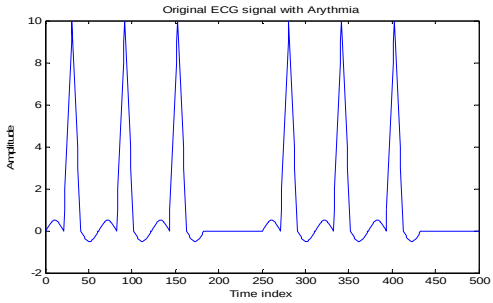


Figure 3 Hardware Block diagram

5. RESULTS

For three diseases its compression and data mining technique have been simulated and results have been obtained by MATLAB.



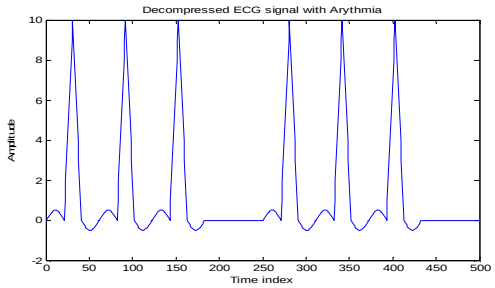


Figure 4 Compressed and decompressed Arythmia signal

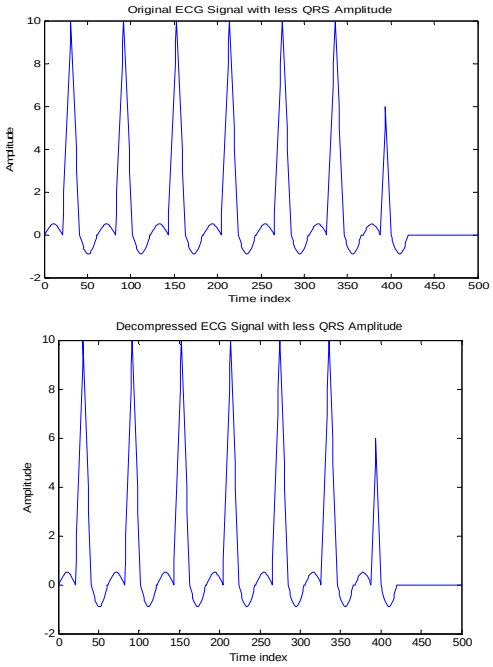


Figure 5 compressed and decompressed R-amplitude less.

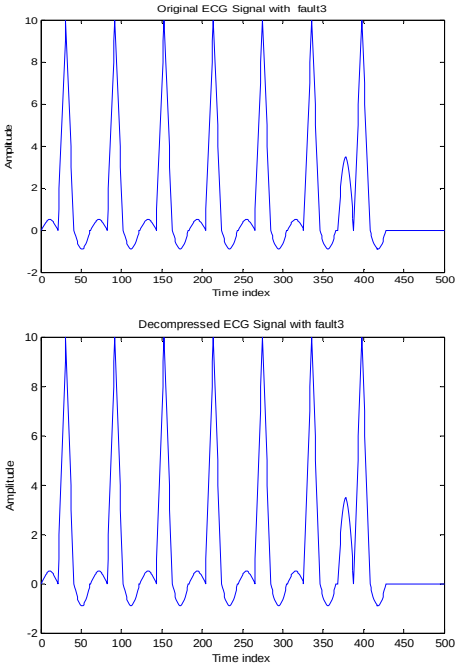


Figure 6 Compressed and decompressed T-U abnormality

6. CONCLUSION

Since the amount of ECG data are large in size, the usage of data compression technology is often used for transmission and storage .The diagnosis can be performed only after decompressing it which causes delay in diagnosis. However, faster diagnosis is absolutely crucial for patient's survival. Therefore, in this paper, we have demonstrated the simulation method and hardware design of an innovative method of faster cardiac abnormality detection mechanism directly from compressed ECG using data mining techniques in hospital server and patient mobile phone.

REFERENCES

[1] F. Sufi, A. Mahmood, and I. Khalil, "Diagnosis of cardiovascular abnormalities from compressed ECG: A data mining based approach," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 1, pp. 33–39, Jan. 2011.

[2] F. Sufi, I. Khalil, and J. Hu, "ECG-based biometric: The next generation in human identification," in *Handbook of Information and Communication Security*, P. Stavroulakis, Ed. New York: Springer-Verlag, 2010, pp. 309–331.

[3] F. Sufi, Q. Fang, I. Khalil, and S. S. Mahmoud, "Novel methods of faster cardiovascular diagnosis in wireless telecardiology," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 4, pp. 537–552, May 2009.

[4] F. Sufi and I. Khalil, "Enforcing secured ECG transmission for real time telemonitoring: A joint encoding, compression, encryption mechanism," *Security Commun. Netw.*, vol. 1, no. 5, pp. 389–405, Aug. 2008.

[5] R.-G. Lee, K.-C. Chen, C.-C. Hsiao, and C.-L. Tseng, "A mobile care system with alertmechanism," *IEEE Trans. Inf. Technol. Biomed.*, vol. 11, no. 5, pp. 507–517, Sep. 2007.

[6] F. Sufi, Q. Fang, and I. Cosic, "ECG R–R peak detection on mobile phones," in *Proc.29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBS)*,Aug. 2007, pp. 3697–3700.

[7] M. Blount, V. M. Batra, A. N. Capella, M. R. Ebling, W. F. Jerome, S. M. Martin, M. Nidd, M. R. Niemi, and S. P. Wright, "Remote healthcare monitoring using personal care connect," *IBM Syst. J.*, vol. 46, no. 1, pp. 95–113, Mar. 2007.

[8] G. D. Clifford, F. Azuaje, and P. E. McSharry, *Advanced Methods and Tools for ECG Data Analysis*. Norwood, MA: Artech House, 2006.

[9] B. Kim, S. Yoo, andM. Lee, "Wavelet-based low-delay ECG compression algorithm for continuous ecg transmission," *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 1, pp. 77–83, Jan. 2006.

[10] K. Hung and Y.-T. Zhang, "Implementation of aWAP-based telemedicine system for patient monitoring," *IEEE Trans. Inf. Technol. Biomed.*, vol. 7, no. 2, pp. 101–107, Jun. 2003.

[11] R. Istepanian and A. Petrosian, "Optimal zonal wavelet-based ECG data compression for a mobile telecardiology system," *IEEE Trans. Inf. Technol. Biomed.*, vol. 4, no. 3, pp. 200–211, Sep. 2000.