International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Special Issue 7, April 2014)
International Conference on Industrial Engineering Science and Application (IESA 2014)

Patient Information Implantation and Reclamation From Compressed ECG Signal By LSB Watermarking Technique

B. Halder[1], S. Mitra[2]
[1]Department of Information Technology, Neotia Institute of Technology, Management & Science, (Affiliated to WBUT), 24-pgs(s) - 743368, W.B (India)
[2]Department of Electronics, Netaji Nagar Day College, (Affiliated to University of Calcutta), Kolkata 700 092, W.B (India)

***Abstract*— In the application of telemedicine, ECG signal is transmitted to the Doctor end without any patient details. As a result, confusion is arisen between signal and patient's identity. To avoid this confusion, ECG signals need to be combined with patient confidential information when sent. A typical ECG monitoring device generates massive volumes of digital data. Huge amount of bandwidth is required for the transmission of the ECG signal for telemedicine purposes. This huge amount of bandwidth for the transmission of the ECG signal can be avoided if the signal is compressed after embedding the patient's personal information within the ECG signal. Since the ECG signal and the patient details integrated into one, bandwidth for the transmission can be reduced in telemedicine applications.**

In this paper ECG signals are watermarked with patient information using LSB watermark technique and compressed the huge amount of ECG data using ASCII character encoding in order to confirm patient or ECG linkage integrity and reduced the bandwidth in telemedicine applications. The whole module has been applied to various ECG data of all the 12 leads taken from PTB diagnostic database (PTB-DB) of physioNet and gives a highly compressed result that can be stored using far less digital space without distorting important ECG characteristics and at the same time, embedded information can be completely retrieved.

***Keywords*— telemedicine, watermarked, ASCII character, PRD, SD, CR**

I. INTRODUCTION

Telemedicine is defined as the "use of computers and telecommunications equipment to provide health care over long distances" (Larkin, 1997). It can be beneficial to patients living in isolated communities and remote regions, who can receive care from doctors or specialists far away without the patient having to travel to visit them [**1**].

Recent developments in mobile collaboration technology can allow healthcare professionals in multiple locations to share information and discuss patient issues as if they were in the same place [**2**]. Remote patient monitoring through mobile technology can reduce the need for outpatient visits and enable remote prescription verification and drug administration oversight, potentially significantly reducing the overall cost of medical care **[3]**. Telemedicine can also facilitate medical education by allowing workers to observe experts in their fields and share best practices more easily **[4].**

Although telemedicine clearly has a wide range of potential benefits, it also has some disadvantages. The main disadvantages are: a breakdown in the relationship between health professional and patient; a breakdown in the relationship between health professionals; issues concerning the quality of health information; and organizational and bureaucratic difficulties. On balance, the benefits of telemedicine are substantial, assuming that more research will reduce or eliminate the obvious drawbacks **[5].**

Digital watermarking is a method to use for authentication in telemedicine application. The purpose of watermarking is hiding a message, called watermark, related to a digital signal (i.e. an image, text, song, and video) within the signal itself. It is a concept closely related to steganography, in that they both hide a message inside a digital signal. However, what separates them is their goal. Watermarking tries to hide a message related to the actual content of the digital signal, while in steganography the digital signal has no relation to the message, and it is merely used as a cover to hide its existence. Watermarking is also called data embedding and information hiding.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Special Issue 7, April 2014)
International Conference on Industrial Engineering Science and Application (IESA 2014)

Throughout these techniques the embedded data will be 'invisible' to maintain the quality of the host data **[6]**. Although there is a large volume of works on watermarking and steganography, not many researchers have addressed issues related to ECG data. In **[7],** the ECG signal is used as a secret data, and embedded inside medical images like CT and MRI. Other researchers **[8]** implemented a wavelet based watermarking technique for ECG signal. In **[9]**, authors proposed to insert an encrypted version of the electronic patient record (EPR) in the **LSB** (Least Significant Bit) of the gray scale levels of a medical image. Similarly, in **[10]**, propose a LSB technique where the host image authenticates the transmission origin with an embedded message composed of various patient data (e.g ECG record), the diagnosis report and the doctor's seal.

The "Electrocardiogram" (ECG) as shown in **fig 1** is an invaluable tool used for the diagnosis and treatment of various cardiac and other related diseases. The volume of ECG data produced by monitoring systems can be quite large over a long period of time. To deal with the huge amount of electrocardiogram (ECG) data for analysis, storage and transmission; an efficient ECG compression technique is needed to reduce the amount of data as much as possible while pre-serving the clinical significant signal for cardiac diagnosis **[11].** An effective data compression scheme for ECG signal is used in many practical applications such as ECG data storage, ambulatory recording systems and ECG data transmission over telephone line or digital telecommunication network for telemedicine. The main goal of any compression technique is to achieve maximum data volume reduction while preserving the significant features and also detecting and eliminating redundancies in a given data set. Due to the diagnostic uses of medical data and since a small detail may be very important, ECG compression techniques have primarily focused on lossless methods.

Fig I: Normal ECG signal

ECG data compression algorithms have been mainly classified into three major categories: 1) Direct time-domain techniques, e.g., turning point (TP**)** **[12**], amplitude-zone-time epoch coding (AZTEC) **[13]**, coordinate reduction time encoding system (CORTES) [**14**] and Fan algorithm **[15]**. 2) Transformational approaches, e.g., discrete cosines transformation (DCT), fast Fourier transform (FFT), discrete sine transform (DST), wavelet transform (WT**)** **[16]** etc. 3) Parameter extraction methods **[17]** are mainly based on linear prediction and long-term prediction methods. In recent years, few mathematical model based methods like Jacobi Polynomials [18], Discrete Sins Interpolation (DSI) **[19]**, Cut N Align Beats (CAB) **[20]** approach, etc. are also reported.

There are extensive work on ECG data compression and hiding patient information in ECG signals, but individually. We propose a novel approach of doing both of the required functions together. In this paper, we focus on how to implant the patient personal information into the ECG signal by LSB watermarking technique **[9]** and how to compressed embedded ECG signal using ASCII character encoding [**11**]. The proposed algorithm allows not only the embedding but also the efficient retrieval of the embedded data. After embedding the patient's personal information, we compress the ECG data by characters encoding. The whole compression scheme is such that the compressed file contains only ASCII characters. In the reconstruction module of the proposed scheme, using the reverse logic of compression which is proposed in **[11],** original ECG signal is brought back and using the reverse logic of 'LSB watermark technique', embedded data i.e. patient's personal information is reconstructed. It is observed that the proposed scheme achieves much better result as compared to in **[11].**

II. METHODOLOGY

The proposed scheme is divided into following five main sections: (A) ECG data selection and Normalization (B) Embedding patient's personal Information (C) Data Compression (D) Data reconstruction (E) Patient personal information retrieval. All these steps are explained sequentially in the rest of the sections. The logic of ECG data compression with embedding patient personal information is written in the MATLAB 7.10a and the following steps are involved:

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Special Issue 7, April 2014)
International Conference on Industrial Engineering Science and Application (IESA 2014)

A. ECG data selection and Normalization:

The ECG signal which is taken from PTB diagnostic ECG database from www.physionet.org as a 'txt'file is copied into notepad and used as ECG data file. From this file, only the 'Voltage' values are taken which we intend to compress. Since the sampling frequency of the input ECG data is 1000 sample/sec, the "Time" axis can be easily generated during data reconstruction. Hence concentration is given only to compress the "Voltage" values discarding the 'Time' value. Suppose we have ECG data with N voltage values in each of the 12 leads. To implement our algorithm, we make a 2D array (ecg[i,j]) of each lead of the ECG data of length (N/8) x 8. Say for example N is 4000, and then the 2D array becomes 500 x 8 viz. 500 rows and 8 columns. Let as an instance Example 1, the content ecg[i,j] array is as below, where i=1 and j=1 to 8.

ecg[1,1]	ecg[1,2]	ecg[1,3]	ecg[1,4]	ecg[1,5]	ecg[1,6]	ecg[1,7]	ecg[1,8]
-0.0350	-0.0370	-0.0320	-0.0180	-0.0180	-0.0220	-0.0150	0

Example 1:

For getting smaller numbers and consequently better compression a difference array is constructed which contains the difference of every ECG sample and it's preceding except the first ''Voltage'' value which is kept unchanged. At the time of reconstruction, the first 'voltage' value is essential otherwise data retrievals will not be possible. Now, the sign of the every element of difference array is checked. For every positive number, a binary zero (0) and for every negative difference, a binary (1) is taken as sign bit for that corresponding 'Voltage' value. Decimal equivalent of this binary string will be used as the **sign-bit** of these corresponding eight 'Voltage' values. Now the sign bit is printed in the output file in its corresponding ASCII character. After getting the sign-bit, all negative numbers in the difference are made positive by multiplying -1. Now each and every number in the difference array is multiplied by 1000 because in standard ECG database and also in PTB-DB, voltages are recorded up-to three decimal points. During creation of the difference array and amplification (multiplying by 1000), the first value of every row is becoming high (possibly greater than 255) as it remain unchanged during creation of difference array.

To bring this number in the range of 0-255, we divide the number by 255 and keep the quotient at this (1st Column) position. The remainder is kept in the variable '***rem***'. A sample array after normalization is shown below:

c[1,1]	c[1,2]	c[1,3]	c[1,4]	c[1,5]	c[1,6]	c[1,7]	c[1,8]
0	2	5	14	0	4	7	15

Example 2

B. Embedding patient's personal Information process:

The second phase of the proposed scheme is to embed the **patient's personal information such as patient ID, age and sex** etc in to the ECG signal. The basic idea of this process is to watermark the patient Id, age and sex into the ECG signal by using the LSB watermarking technique **[9].** For this we create a 1D array from normalized 2D array of each lead of the ECG data. In such an array, we will embed patient's personal information by ***watermarking*** techniques.

The length of the patient Id is calculated firstly, and then converted to its corresponding binary data form. This binary data is embedded in the voltage values of the first ECG lead using LSB embedding technique. As we are using LSB technique, the possible change in ECG voltage values will only be by 1 viz. increased or decreased by 1 or will remain unchanged. Data bits will be scattered throughout the remaining 11 leads. This algorithm involves the following steps:

1) Embedding the length of the patient Id in first lead: Let the length of the patient Id or the number of characters is 12. The length of the characters is to be embedded in the first lead i.e.2nd column of actual input ECG file. First, the length of the embedded data (i.e.12) is converted to its corresponding binary value i.e. '1100'. These bits will be embedded with original ECG signal in the interval of 8, i.e. in 8th, 16th, 24th and 32nd position in first lead. Suppose the four values of position 8,16,24 and 32 in lead 1 are 15, 10, 11, 4 respectively, then after embedding '1100' the result will be shown below in example 3

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Special Issue 7, April 2014)
International Conference on Industrial Engineering Science and Application (IESA 2014)

	A	B / C	D	
8th	15	0000111**1**	0000111**0**	14
16th	10	0000101**0**	0000101**0**	10
24th	11	0000101**1**	0000101**1**	11
32nd	4	0000010**0**	0000010**1**	5

A= actual data B= binary value of actual data
C= actual with embedded data. D= embedded data:

Example 3

2) Calculation of the length of patient age and sex: Suppose the patient age is 65.Corresponding ASCII character of age is 'A', i.e number of character is one. Similarly the number of character of patient sex is one. It is either 'M' or 'F'. So the total number of character for embedding in the first lead i.e.2nd column of actual input ECG file is length of patient Id plus two. Suppose the length of the patient Id is 12 then total number of character is 14 which are to be embedded in the first lead i.e.2nd column of actual input ECG file. At the time of reconstruction, the length of the patient Id is essential otherwise data retrievals will not be possible.

3) Embedded data generation: Let us describe now how the embedded data is generated. Here, the patient Id, used in PTB data base for each ECG file, patient age and sex are considered as data for embedding in each ECG signal. Here patient Id and sex both are in text format and patient age has a numeric value. All these data's are converted into its corresponding ASCII value. The ASCII value of each character is converted to its corresponding binary form with seven bits. Suppose the file name is 's0015lre.txt', and used as patient ID which is of 12 characters; the total number of bits that has to be embedded is 12*7 =84 bits. Similarly, suppose the patient age is 65 and its corresponding ASCII value is 'A' and patient sex is either 'F' or 'M'. These two characters (i.e age and sex) again converted to its corresponding binary form with seven bits each. So, the total number of bits that has to be embedded in ECG file is 98(ie, 84+2*7).

4) Calculation of the number of bits and interval for embedding data: Let us describe now how to calculate the number of bits of a particular lead for embedding data. For this we scatter the embedded data bits in all the 12leads of ECG data because embedding all bits in a particular lead might increase the PRD of that lead. First lead is already reserved for the length of embedding patient Id.

The number of bits to be embedded in remaining 11 leads will be calculated by 'Bitload1' and 'Bitload2'.The equations are given below:

$$Bitload1=floor\ (TB/(TL-1))\ \ldots\ldots \quad (1)$$

$$Bitload2=mod(TB,(TL-1))\ \ldots\ldots \quad (2)$$

Where ***TB*** is the total data bits and ***TL*** is the number of leads for embedding. Suppose the total number of bits to be embedded is 98 i.e TB=98 and number of leads for embedding is 11 i.e TL = 11.Hence the corresponding value of Bitload1 and Bitload2 will be 9 and 8 respectively. So number of bits to be embedded in lead 2 to lead 11 is 9 and number of bits to be embedded in lead number 12 is 8.

Let us calculate now the interval at which these bits will be embedded in the respective lead will depend upon the first sign-bit of that respective lead. The interval for every bit will be calculated by using following equation:

$$Interval = sign\ bit + 50 + (8 - mod\ (sign\ bit, 8))\ldots \quad (3)$$

Here 50 is added as an offset because a situation might arise where the sign byte for a lead is zero. By making this offset we can be sure that the interval is always 50 and above. **Example 4** shown below, suppose the sign-bit of a lead is 128 then Interval is 184 (ie, [128+50+(8-mod(128,8))]) . So, the bits will be embedded at 184th, 368th, 552nd…..etc positions. If the first sign-bit of a lead is 128, then the positions at which 8 bits will be embedded are (1*184), (2*184), (3*184), (4*184), (5*184), (6*184), (7*184), (8*184).

	A	B	C	D
184	*46*	*0010111**0***	*1*	*00101111*
368	*20*	*0001010**0***	*1*	*00010101*
552	*3*	*0000001**1***	*1*	*00000011*
736	*26*	*0001101**0***	*0*	*00011010*
920	*110*	*0110111**0***	*0*	*01101110*
1104	*5*	*0000010**1***	*1*	*00000101*
1288	*21*	*0001010**1***	*1*	*00010101*
1472	*49*	*0011000**1***	*0*	*00110000*

A: Actual data, B: binary form of actual data, C: Embedded data, D: Actual with embedded data

Example 4

After the Text bits have been embedded in the leads, again embedded array is converted to 2D array.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Special Issue 7, April 2014)
International Conference on Industrial Engineering Science and Application (IESA 2014)

C. ECG data compression:

To compress the data, embedded integers are grouped maintaining some essential logical criteria. Three types of grouping are considered here as in [**11**] namely (a) forward grouping, (b) reverse grouping, (c) no-grouping. Three variables have been taken (f : for forward, r : for reverse and n : for no grouping position respectively, initialized with zero (0)) for denoting the positions of these groupings. Suppose the first pair of c[ ] array, i.e. c[1,1] and c[1,2], we take the value "1" (which is the first position of the pair) for denoting the grouping position. For all the other pairs we take the value "1" (which is the first position of the pair) for denoting the grouping position. For all the other pairs we take the second position for denoting the grouping position. For example in the second pair, i.e. c[1,3] and c[1,4], we would take value "4", for pairs c[1,5] and c[1,6] we would take value "6" and so on. After all the pairs are examined, we sum the values and store it in the respective variable, for all three grouping procedure, initially 'i' is set to 'one' and incremented by a factor of two.

For forward grouping, each number in c[i,j] array position is multiplied by 100 and added with the number in c[i,j+1]array position and checked whether the value is less than 255, then this grouped integer value is printed in character form. If forward grouping is not possible then it is to be checked that whether any '(c[i,j + 1] *100) + c[i,j]' is less than 255 or not. If so, then this reversed grouped value is printed in character form. If both forward and reverse groupings are not possible then, value at those positions are left unchanged and both integers are printed separately in character form.

Each set of grouped or not-grouped 'Voltage' values along with other necessary information (sign byte, f, r, n, etc.) will be printed in the output file in ASCII character form maintaining the following format.

Sign Bit	Grouped / ungrouped voltage values	Forward grouping position(k)	Reverse grouping position(z)	No grouping position(u)	Data Position	Remainder

Here, Data position holds the position where forward, reverse and no grouping position ends. In the ***example 5***, mentioned below, the value of f=7, r=4, n=8 and data position=9.sign bit=196 and rem=35.

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Decimal	196	200	104	106	4	5	7	4	8	0	0	0	9	35
ASCII char	A	E	h	j	EOT	ENQ	BEL	EOT	BS	NUL	NUL	NUL	HT	#

Example 5

The above compression algorithm is executed iteratively until all the ECG data contained in the input file are become compressed.

D. ECG Data reconstruction:

The *fourth phase* of the proposed technique is the ECG data reconstruction. After compression with embedded data, it is necessary to decompress the data for retrieving the original signal and patient's information at the receiving end. For this purpose, a modified version of the algorithm proposed in [**11**] is used to decompress the ECG data in lossless manner. From the compressed data file at a time one set of ASCII character is taken. There will be at most 14 characters in a set. Equivalent ASCII values of those characters are saved in an array and then reconstruction algorithm is applied on those to restore original eight 'Voltage' values. The first ASCII value among those contains the 'Sign byte' of those eight voltage values. Following ASCII values are the grouped / no grouped integers, grouping positions, Data position and 'rem' respectively. All these are necessary to get back the original eight voltage values. From the ***Example 5*** we know that, data position is 9, so, the 9^{th},8^{th} and 7^{th} are the grouping position. The value of f=7, r=4, n=8 and sign bit=196 and rem=35. Using the reverse logic of grouping technique, the original eight voltage values are generated and stored in an array , To obtain the actual value at the first position, we need to multiply the value at first value of the array with 255 and add the value of 'rem' to it. In the next step, the sign-bit will be converted into its corresponding 8-bit binary equivalent. In the binary string if any bit is '1' then the corresponding positional element of array will be multiplied by (-1). This array contains the reconstructed ECG voltage values.

E. Patient personal information retrieval:

The fifth phase of the proposed technique is to retrieve the patient's identification from the reconstructed ECG Data. This algorithm is the reverse logic of embedding patient's identification.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Special Issue 7, April 2014)
International Conference on Industrial Engineering Science and Application (IESA 2014)

1) Calculation of the length of Patient's Identification: As during embedding, we have embedded the size of data in the LSB of four values in 8^{th}, 16^{th}, 24^{th} and 32^{th} position in first lead. We will calculate the length of the embedded data (patient's ID) from the LSB identification of these values. Suppose the four values of position 8, 16, 24 and 32 in lead 1 are 20, 2, 11, 3 respectively. They will be interpreted providing the size of data embedded as:

Actual data	*binary of actual data*	*reconstructed data*
20	0001010**0**	**0**
2	0000001**0**	**0**
11	0000101**1**	**1**
3	0000001**1**	**1**

Example 6

From the above ***example 6***, the length of embedded data is '1100';its corresponding decimal value is 12, which means number of characters (NC) is twelve characters. Therefore the number of bits embedded will be NC*7, which for the above example will be 12*7=84.

2) Calculation of the length of Patient's age and sex: The number of character of patient age is one and number of character of sex is one which means that number of character is two. Therefore the number of bits embedded will be 2*7=14. So the total number of character for embedding in the first lead i.e.2^{nd} column of actual input ECG file is length of patient Id plus two. Suppose the length of the patient Id is 12 then total number of character is 14 which are to be embedded in the first lead i.e.2^{nd} column of actual input ECG file. Therefore the total number of bits which are to be embedded in the first lead of ECG file will be (84+2*7) =98.

3) Calculation of the number of bits and interval for embedding data: By reserve procedure as followed during the embedding, we will calculate the value of Bitload1 and Bitload2 by the equation:

Bitload1=floor (TB/ (TL-1)) ………… (4)

Bitload2= mod (TB, (TL-1)) ………… (5)

Suppose, the total number of bits to be embedded is 98 i.e

TB=98 and number of leads for embedding is11 i.e TL = 11.Hence the corresponding value of Bitload1 and Bitload2 will be 9 and 8 respectively.

So number of bits to be embedded in lead 2 to lead 11 is 9 and number of bits to be embedded in lead number 12 is 8.

Now the interval at which these bits will be embedded in the respective lead will depend upon the first sign bit of that respective lead. The interval for every bit will be calculated as:

Interval = sign bit +50 + (8 – mod (sign bit, 8))… (6)

If the first sign bit of the lead is 128, we add 50 to it making it 178. The actual interval in this case will be next multiple of 8 from 178, i.e. 184. For ***example7***, mentioned below, the intervals at which the bits are embedded in a lead with sign byte 128 are (1*184), (2*184), (3*184), (4*184), (5*184), (6*184), (7*184), (8*184).

Interval	*184*	*368*	*552*	*736*	*920*	*1104*	*1288*	*1472*
Actual data	47	21	3	26	110	5	21	49
binary	00101111	00010101	00000011	00011010	01101110	00000101	00010101	00110000
Retrieve data	1	1	1	0	0	1	1	0

Example 7

4) Recovering the Embedded Data bits and Text reconstruction: In the previous step; for a particular lead we calculated the interval at which the text bits have been embedded. Following the reverse procedure of what we applied during embedding, we will convert the value at the intervals obtained into binary format and will store its LSB for construction of the characters. The binary value will again be converted into decimal for application of further procedure. By iterating the procedure of LSB obtaining method mentioned above for each lead as per Bitload1 and Bitload2 values, we will obtain a 2D array of X*7, where X will be the number of characters that were embedded. Converting each row with seven binary values to corresponding decimal value, we can obtain the character which was embedded. (Every decimal value has corresponding character is ASCII character set).The character thus obtained is the data which was embedded.

Now, every number is divided by 1000. To get the original, each number is added with the previous value. The new values are stored in another array. This array is identical to the original ECG voltage values.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Special Issue 7, April 2014)
International Conference on Industrial Engineering Science and Application (IESA 2014)

III. MEASURES OF PERFORMANCE

The following measures are computed to evaluate the performances of the proposed algorithm.

- The first parameter that expresses the effectiveness of a data compression technique is the ***Compression Ratio (CR).***It is defined as below. The higher the value of CR, the better of the performance of the compression algorithm.

$$CR = \frac{OriginalECG\,data\ filesize}{Compressed\,ECG\ file\,size}$$

- The bit rate equals the inverse of the compression ratio. It is also known as ***bit per bit (bpb).***The main objective of a data compression method is to achieve the lowest bit rate.

$$bpb = \frac{CompressedFileSize}{OriginalECGDataFileSize}$$

- It is also possible to express the compression ratio as a percentage of the size of the original data. This measure is called ***Compression Percentage (CP)*** and it is defined as:

$$CP = (1 - bpb)x100\ (\%)$$

Taking the ratio of compression 1:4 as an example, that is to say that the original file is 4 times bigger than the compressed one, the compression percentage would be 75%.

- ***Percent Root Mean Square Difference*** (PRD) is a measure of error loss. This measure evaluates the distortion between the original and the reconstructed signal. PRD calculation is as follows:

$$PRD = 100X\sqrt{\frac{\sum_{i=1}^{n}(y_i - \overline{y_i})}{\sum_{i=1}^{n} y_i^2}}$$

Where y_i represents original signal samples and $\bar{y}_i$ represents reconstructed signal samples. The lower the PRD, the closer the reconstructed signal is to the original ECG data.

- Another numerical measure ***Quality Score*** (QS) was pro-posed in **[27]** to quantify the overall performance of compression algorithm. A high score of QS represents a good compression performance.

$$QS = \frac{CR}{PRD}$$

- ***The Standard Deviation*** (SD) of a data set is a measure of how spread out the data is. The English definition of the SD is: "The average distance from the mean of the data set to a point". The way to calculate it is to compute the squares of the distance from each data point to the mean of the set, add them all up, divide by (***n***-1) and take the positive square root. It is calculated as below:

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}}$$

Here, the symbol X refer to the entire set of numbers, the symbol ***n*** will be used to refer to the number of elements in the set X and the symbol $\bar{X}$(said "X bar") to indicate the ***mean*** of a sample is, and will only give the formula:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

It is a powerful statistics for comparing the degree of deviation from one data series to another, even if the means are drastically different from each other. The lower the value of SD, the better of the performance of the compression algorithm.

PRD, CR, BPB, CP, QS and SD obtained for Myocardial Infarction and Normal patients are given in Table 1.

International Journal of Emerging Technology and Advanced Engineering

Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Special Issue 7, April 2014)

International Conference on Industrial Engineering Science and Application (IESA 2014)

Table 1

FILE NAME	AGE	SEX	LEAD	PRD %	CR	QS	BPB	CP	SD
s0499_re.txt	58	M	12Lead	0.00274	7.295404	3623.982	12.88362	12.88365	0.0000160000
s0464_re.txt	67	F	12Lead	0.003799	7.401489	2544.351	9.745821	9.745851	0.0000000000
s0462_re.txt	77	f	12Lead	0.004349	7.453826	2649.046	10.14304	10.14306	-0.0000180000
s0463_re.txt	50	M	12Lead	0.003365	7.35102	2963.907	9.827201	9.827165	0.0000000000
s0015lre.txt	75	F	12Lead	0.005413	7.241806	1791.987	10.58152	10.58152	-0.0000050000
s0465_re.txt	100	M	12Lead	0.004687	7.419356	1927.233	10.23826	10.23827	0.0000090000
s0466_re.txt	30	M	12Lead	0.005393	7.454503	2470.757	13.27175	13.2718	0.0000010000
s0467_re.txt	45	F	12Lead	0.004884	7.424361	3824.478	12.62679	12.6268	0.0000000000
s0474_re.txt	24	F	12Lead	0.00627	7.358899	1294.135	10.30377	10.3038	-0.0000040000
s0475_re.txt	56	M	12Lead	0.004234	7.528023	2279.893	9.067906	9.067934	-0.0000390000
s0480_re.txt	51	F	12Lead	0.002265	7.313354	6407.6	9.039927	9.039942	0.0000000000
S0025lre.txt	34	M	12Lead	0.003604	7.290111	2201.596	8.522471	8.522483	0.0000000000
S0022lre.txt	41	F	12Lead	0.007101	7.192171	1379.333	10.27368	10.27373	0.0000000000
S0026lre.txt	55	M	12Lead	0.007313	7.203689	1180.047	9.632758	9.632756	0.0000270000
s0027lre.txt	77	F	12Lead	0.004635	7.187249	1729.03	8.383026	8.383054	0.0000270000
s0028lre.txt	44	F	12Lead	0.004631	7.187249	1736.707	8.383026	8.383038	0.0000000000
s0029lre.txt	55	M	12Lead	0.003433	7.238283	4006.055	11.40791	11.4079	0.0000000000
s0031lre.txt	25	F	12Lead	0.003913	7.429676	3133.902	9.523156	9.523146	0.0000000000
AVERAGE				**0.004557**	**7.331693**	**1608.88**	**10.2142**	**10.21422**	0.000007778

CR: compression ratio, BPB: bit per bit, CP: compression percentage, QS:Quality Score ,
PRD: percent root-meansquare difference, SD:Standard Deviation

From the above table1, we can see that using the proposed algorithm we can get average PRD of about 0.0045, average CR of about 7.33:1, average CP of about 86%, average QS of about 1608 and standard deviation of about 0.00000077. Due to this lower standard deviation, an excellent performance of the compression algorithm gets raised.

Following figures (Figs. 2–5) shows the original (blue) and reconstructed (green) ECG signals of 4's sample. Fig. 2 and Fig 3 are Myocardial Infarction (**MI**) ECG and Fig 4 and Fig 5 are Normal(N) ECG signal.

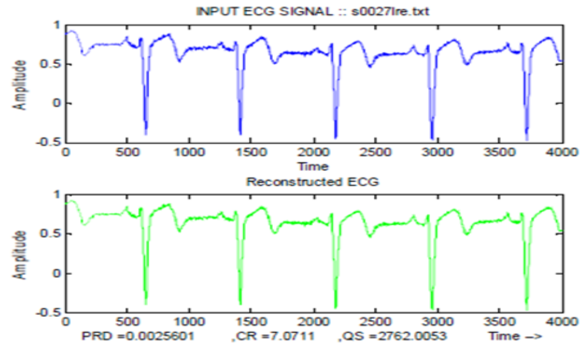

Fig 2: File name: S0027lre.txt (MI)

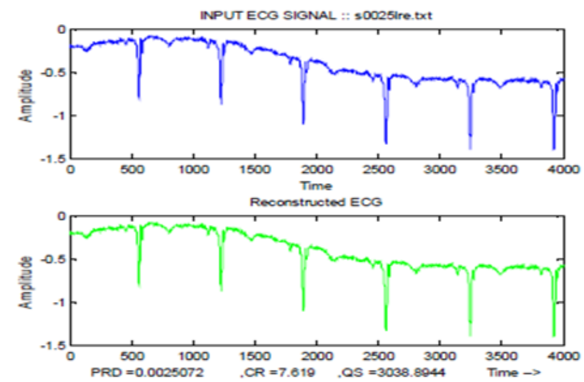

Fig 3: File name: S0025lre.txt (MI)

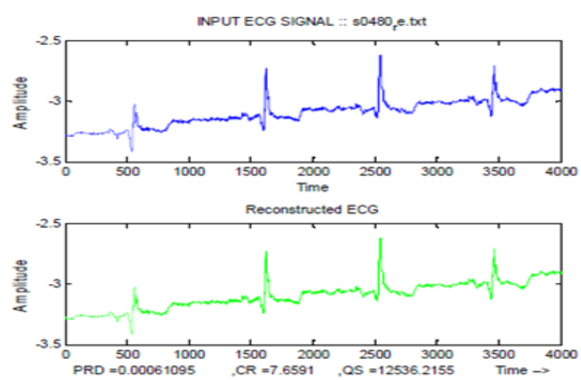

Fig 4: File name: S0480_re.txt (N)

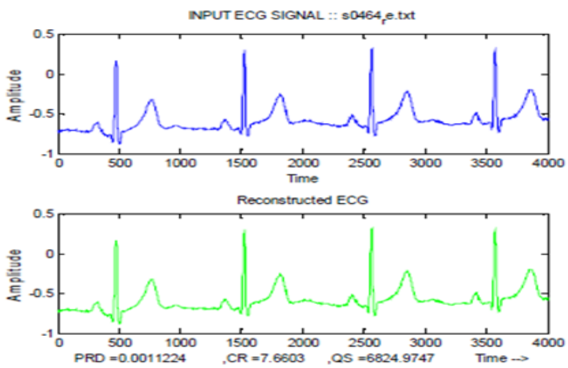

Fig 5: File name: S0464_re.txt (N)

IV. Performance Comparison

Comparison of CR, PRD and QS of various lossless and near lossless methods with the proposed method is given in Table 2.

Table 2

Algorithm Name	*PRD*	*CR*
AZTEC [13]	5.1	10
Hilton [16]	2.6	8
Johan [21]	3.9	8
SPIHT[23]	1.18	8
Perceptual Masks[24]	1.24	3.5
Al-Shrouf [25]	5.3	11.6
Fira and Goras [26]	0.61	12.74
TP [12]	28	2
USZZQ and Huffman coding of DSM [27]	2.73	11.06
Wang and Meng [22]	1.6	12
S.K. Mukhopadhyay et al. LLEDCCE [11]	0.023	7.18
Proposed Method	**0.0045**	**7.33**

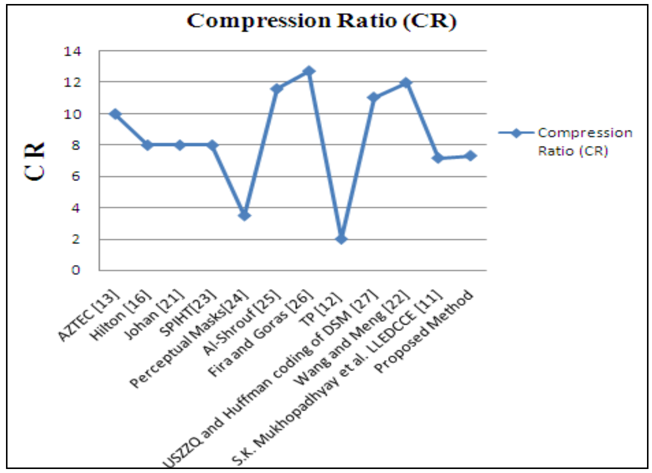

Fig 6: Comparison of CR with other algorithms

It is observed that the proposed method achieves a relatively high CR **(7.33)** as compared to different methods reported in Table 2. Though few methods achieve higher CR than the proposed method but it has the lowest PRD **(0.0045).** Due to this low PRD, an excellent QS value often gets raised. So, CR and PRD of the proposed method is comparable with the methods as listed in Table 2.

The figure 6 and 7 give the graphical comparison of PRD and CR values of various algorithms with our proposed method:

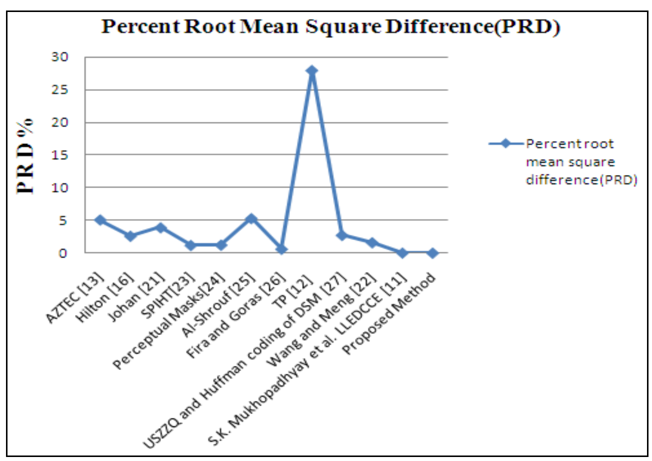

Fig 7: Comparison of PRD values with other algorithms

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Special Issue 7, April 2014)
International Conference on Industrial Engineering Science and Application (IESA 2014)

V. CONCLUSION

We proposed a new approach for implantation and reclamation of patient information from compressed ECG signals by using LSB watermarking technique. We proved statistically that the implanting of patient information does not affect the significant features of the ECG signals and it also does not induce any changes in the diagnosis. Proposed algorithm shows better performance for both normal and abnormal ECG signals. The performance of the proposed algorithm was compared with some other proposed ECG signal compression methods, specially in [11**]** and it is observed that the proposed algorithm gives a high compression ratio with low distortion and an excellent Quality Score than in [**11**].The results of this research will likely provide an improvement on existing techniques.

In Telemedicine applications, ECG signal is widely used to monitor cardiac activities of patients. The compressed ECG file with embedding patient's information may be transmitted to the doctor end over GSM network as text message for real time ECG analysis. Doctors can then use the ECG data by using the construction and reconstruction algorithms and also hide the patients information inside the ECG signal using this algorithm. As a result, the people who are living in remote areas and unable to meet with doctor face to face due to geographical distances, they are receive expert diagnosis and treatment from distant medical centers. In India, 70% population lives in rural areas, which are found struggling to access timely medical treatment, whereas 75% of qualified specialists doctors practice in urban centers. In future, the system will be enabled to provide some sort of medical assistance to those people, who can receive care from doctors or specialists far away without the patient having to travel to visit them [**1**]. In near future those features will be incorporated in the present system.

Acknowledgements

The authors acknowledge their deepest gratitude to Dr. M. Mitra and S.K. Mukhopadhyay of Department of Applied Physics, University of Calcutta and Prof. Subrata Bose Department of Information Technology, Neotia Institute of Technology, Management & Science, Jhinga, West Bengal, India, for their help.

REFERENCES

[1] Matthew.B and Andrea.F (June 2005).'Technology and managed care: patient benefits of telemedicine in a rural health care network.' Health Economics". Health Economics. P.559–573, June 2005.

[2] Haaff.C.V (March–April 2009)."Virtually On-sight". Just for Canadian Doctors. p. 22.

[3] Michael.S (2012).'The Mobile Wave: How Mobile Intelligence Will Change Everything'. Perseus Books/Vanguard Press. p. 153.

[4] Conde,J.G,De.S,Hall.R.W,Johansen.E, Meglan.D, Peng.G.C.Y,(Jan–Feb 2010). 'Telehealth Innovations in Health Education and Training.' Telemedicine and e-Health. p. 103–106.

[5] NM. Hjelm,'Benefits and drawbacks of telemedicine.', Telemed Telecare. 2005; p-60-70.

[6] Kozat.S.S,Vlachos.M, Lucchese. H, Herle.V, Yu. P. S," Embedding and Retrieving Private Metadata in Electrocardiograms", August 2009, Volume 33, pp 241-259

[7] Nambakhsh.M, Ahmadian.A, Ghavami.M, Dilmaghani.R, and Karimi-Fard.S. 'A novel blind watermarking of ECG signals on medical images using EZW algorithm'. In Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Volume 1, 2006

[8] Engin.M, Cidam.O, and Engin.E. 'Wavelet Transformation Based Watermarking Technique forHuman Electrocardiogram (ECG)'. Journalof Medical Systems, 29(6):589–594, 2005.

[9] Anand.D and Niranjan.U.C. "Watermarking Medical Images with Patient Information". In Proceedings IEEE/EMBS Conference, Hong Kong, China, October 1998, pp. 703-706.

[10] Miaou.S,Hsu.C,Tsai.Y,Chao.H " A Secure Data Hiding Technique with Heterogeneous Data-Combining Capability for Electronic Patient Records," in Proceedings of the World Congress on Medical Physics and Biomedical Engineering, Session Electronic Healthcare Records, IEEE-EMB, ed.,Chicago, USA, July 2000.

[11] Mukhopadhyay.S.K, Mitra.S, Mitra.M, "A lossless ECG data compression technique using ASCII Character encoding", Computers and Electrical Engineering 37 (2011) 486–497.

[12] Mueller.W.C, "Arrhythmia detection software for an ambulatory ECG monitor," Biomed. Sci. Inst., vol. 14, pp. 81–85, 1978.

[13] Cox.J.R, Nolle.F.M, Fozzard.H.A, and Oliver.G.C, AZTEC,"A preprocessing program for real time rhythm analysis," IEEE rans.Biomed. Eng., vol. BME-15, pp. 128–129, 1968.

[14] Abenstein.J.P and Tompkins.W.J, "A *new* data-reduction algorithm for real-time ECG rhythm analysis", IEEE Trans. Biomed. Eng. 1982. vol. 29, pp. 43–48,

[15] Gardenhire.L.W, 'Redundancy reduction the key to adaptive telemetry'. In: Proc. 1964 nat. telemetry conf, p. 1–16,

[16] Hilton.M.L "Wavelet and wavelet packet compression of Electrocardiograms" IEEE Trans Biomed Eng 1997;44(5):394–402.

[17] Nave.G, Cohen.A,"ECG compression using long-term prediction". IEEE Trans Biomed Eng 1993; 40:877–85.

[18] D.Tchiotsop, D.Wolf,V.L. Dorr,R. Husson . 'ECG data compression using Jacobi polynomials', In: Proc. IEEE EMBS, 22–26 August 2007. p. 1863–7.

[19] Manikandan.S,Dandapat.M.S. 'ECG signal compression using Discrete Sins Interpolation'..In: Third international conference on intelligent sensing and information processing (ICISIP 2005), Bangalore, India, 14–17 December 2005. p. 14–9

[20] Lee.H, Buckley.K.M. 'ECG data compression using cut and aligns beats approach and 2-D transforms'. IEEE Trans Biomed Eng 1999; P.556-64

[21] Johan.A.D, Nguyen.T.Q, Tompkins.W.J."ECG compressions using discrete symmetric wavelet transform". IEEE international Conference on engineering in medicine and biology society, Montreal, Que, Canada, vol. 1. p. 167–8.

[22] Wang.X, Meng.J. 'A 2-D ECG compression algorithm based on wavelet transforms and vector quantization'. Digital Signal Process 2008; 18:179–88.

[23] Lu.Z, Kim.D.Y, Pearlman.W.A. 'Wavelet compression of ECG Signals by the set partitioning in hierarchical trees (SPIHT) algorithm'. IEEE Trans Biomed Eng 2000; 47(7):849–56.

[24] Rodrigo, C.M.D., Fabrizia, M.M. and Leonardo, V.B., Near-lossless compression ofECG signals using perceptual masks in the DCT domain, CLAIB 2007. In: IFMBE Proceedings 18, pp. 229-231.

[25] Shrouf.A.A, Zahhad.M.A, Ahmed.S.M.'A novel compression algorithm for electrocardiogram signals based on the linear ediction of the wavelet coefficients'. Digital Signal Process 2003; 13:604–22.

[26] Fira.C.M, Goras. L. 'An ECG signals compression method and its validation using NNs'. IEEE Trans Biomed Eng 2008; VOL. 55, NO. 4, APRIL 2008, p.1319–26.

[27] Manikandan, M.S., Dandapat, S.: 'Wavelet threshold based ECG Compression using USZZQ and Huffman coding of DSM', Biomed. Sig. Proc. Control., 2006, 1, (4), pp. 261– 270