# A Novel Lossless ECG Compression Technique for Transmission in GSM Networks

**Diana Moses and C. Deisy**

**Abstract** This paper presents a novel Lossless ECG Compression using Symbol substitution (LECS) deployable on low computational devices (LCD) like mobile phones for effective use in telecardiology. Of the few LCD deployable compression algorithms, even losslessly compressed ECG suffers transmission loss in Global System for Mobile (GSM) networks due to the reduced character set supported by the SMS protocols. The LECS encodes using the Standard GSM Character ETSI GSM 03.38 set for un-trimmed ECG transmission. The evaluation using MIT-BIH Arrhythmia database showed an average compression-ratio (CR) of 7.03, Percentage-Root-mean-square-Distortion (PRD) as low as 0.0211 proving superior performance in both compression and quality for real-time mobile based telecardiology applications.

**Keywords** Lossless ECG compression · ECG in SMS · ECG transmission · Telecardiology · ECG transmission in GSM network · Symbol substitution based compression

## 1 Introduction

Telecardiology is the electronic transmission of cardiac data wiz the Electrocardiograph (ECG) acquired from the patient to a health care service provider for diagnostic purposes. The current availability of doctors is only 14 per 10,000 patients as indicated by the WHO [1]. Increase in aged population living alone and

D. Moses (✉) · C. Deisy
Thiagarajar College of Engineering, Madurai, India
e-mail: itsdianamoses@gmail.com

C. Deisy
e-mail: cdcse@tce.edu

M. Pant et al. (eds.), *Proceedings of the Third International Conference on Soft Computing for Problem Solving*, Advances in Intelligent Systems and Computing 259,
DOI: 10.1007/978-81-322-1768-8_78, 

lifestyle modifications are few factors that account for the inevitability of tele-cardiology based applications.

The Electrocardiogram is a noninvasive, transthoracic device used to record the electrical activity of the heart. Figure 1 shows a sample ECG waveform depicting the key features that aid in diagnosis of a heart disease. The key features used for diagnosis are HRV (Heart Rate Variability), RR Interval and Width of the QRS Complex [2]. The reflection of Cardiovascular Diseases (CVDs) occurs in the shape and size of the P-QRS-T waves randomly on the ECG timescale. Thus ECG is recorded for extended periods of time (24–48 h). This calls for compression of the acquired data both for remote monitoring and archival purposes.

The existing ECG compression algorithms can be classified into lossless and lossy based on the reconstruction potential of the algorithm. Lossless compression methods ensure no loss of information and exact reconstruction of original signal from the compressed signal. The basic approaches for lossless compression are based on Entropy, Dictionaries and Sliding windows. Entropy based methods include Huff-man, Shannon-Fano Coding and Arithmetic coding. Lempel–Ziv Welsch (LZW) and some of its variations employ dictionaries for compression. LZ77 and DEFLATE apply sliding window based compression [3, 4]. Antidictionary and ASCII based me-thods are also reported for lossless ECG compression. Lossless methods generally provide low compression ratios of 2:1 to 4:1 [5].

Lossy compression algorithms provide compression ratios of up to 10:1 to 20:1 where the high compression ratio is achieved by retaining only the important part of the data to be encoded while discarding others [5]. Lossy compression algorithms can be classified as: spatial, frequency, and time–frequency domain methods. Methods like direct data compression, model based and parameter extraction methods exploit redundancy in the spatial domain [5]. Model based methods such as Neural networks build mathematical models that are losslessy encoded by other coding schemes [6, 7]. Fourier transforms and Discrete Cosine Transforms exploit the redundancy at different energy levels of the transformed ECG signal [8, 9]. Multilevel resolution in both time and frequency domains is provided by wavelet transforms [9]. Wavelets are the most commonly used methods for lossy compression because of the efficient compression and the ability of the transform coefficients to intricately quantify the signal in both time and frequency domains for further analysis [9]. Generally transforms are not recommended for low computational devices such as mobile phones due to higher power consumption, requirement of buffering caused by data dependencies and increased latency.

Lossless methods are encouraged to avoid the possibility of losing biomedical signal artifacts of potential diagnostic significance. Lossless compression methods employ predictive preprocessing methods followed by an encoding scheme. Chua and Fang [10] employed a predictive discrete pulse code modulation (DPCM) and used Go-lumb Rice coding for encoding them along with a highly integrated VLSI design to losslessly compress ECG with very low power consumption. Koski adopted a structural method for recognition of ECG complexes and encoded them using complex extraction Huffman coding for improved compression performance

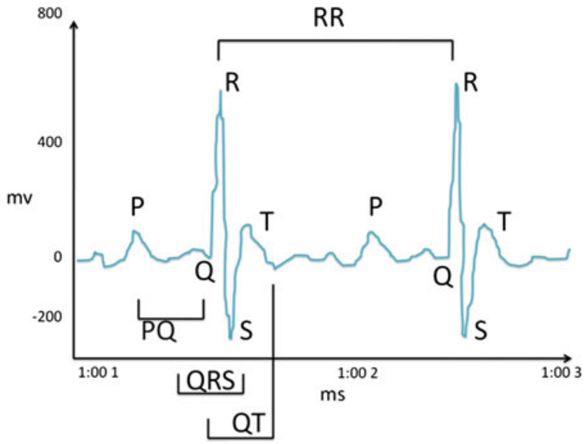


**Fig. 1** An ideal ECG waveform and its features

[11]. Sriraam [12], proposed a correlation dimension based predictor followed by arithmetic coding for EEG signal. Srinivasan et al. [13], employed the correlation dimension based predictor on wavelet transform coefficients and encoded them using arithmetic coding. Takiharo and Morita [14], proposed a novel arithmetic coding scheme based on antidictionaries which bypassed the predictive preprocessing phase and also yielded higher compression ratio. Boucheham [15], on the other hand extended the predictive phase into two stages. Line simplification methods for short-term prediction and Curve simplification methods for long term prediction were used.

Muhopadhyay et al. [16], applied a simple but efficient ASCII based encoding technique after grouping based preprocessing. He also proposed a lossy counter part with a comparatively high quality score [17]. The method adopted the standard ASCII character set with 128 characters of which 33 are non-printable control characters. To avoid loss of data due to use of non-printable characters when encoding with the ASCII table forward and reverse grouping methods are employed to encode the characters using the available 95 printable ASCII characters. Nevertheless compressed ECG suffers data loss when transmitted in SMS due to restricted character set supported by the SMS protocol. Sufi et al. [18] proposed a novel lossy compression scheme using the extended ETSI GSM 03.38 character set defined for the SMS protocol. A computationally light, lossless compression using this character set is required to transmit ECG in real-time telemonitoring systems. Towards achieving the aforementioned goals, a simple and efficient ECG compression algorithm called Lossless ECG Compression using Symbol substitution (LECS) for lossless storage and retrieval, untrimmed transmission via messaging protocols and applicable on low computational devices is proposed. Moreover, the high compression ratio coupled with lossless compression renders the proposed algorithm for ECG archival purposes as well. The remaining portion of the paper is organized as follows: Sect. 2 explains the LECS

Compression algorithm; Sect. 3 details the LECS Decompression procedure. The experimental setup is explained in Sect. 4. The results and discussions presented in Sects. 5 and 6 concludes the work and provides directions for further research.

## 2 Proposed Lossless ECG Compression Using Symbol Substitution (LECS) Algorithm

Lossless compression of data ensures no data loss and exact reconstruction of original ECG from the compressed ECG. Unlike media files where data loss can be acceptable for the amount of compression offered, ECG and other health data cannot be risked with the loss of data. The proposed technique is simple and efficient. The compression is based on symbol substitution using a symbol table constructed from the 124 printable characters in ETSI GSM 03.38 character set. The use of this reduced character set ensures that data transmitted is not trimmed because of reduced character set supported by SMS protocols. The LECS algorithm involves a series of preprocessing phases followed by ECG Compaction and symbol substitution based encoding. The block representation of the proposed algorithm is given in Fig. 2. The essential steps include Normalization, Differencing, Sign epoch marking, ECG Compaction, Symbol substitution based coding. Normalization and differencing have been the basic preprocessing steps in compression in many ECG Compression methods. In the proposed approach at every phase a compression parameter is appended to the compressed ECG header to aid in exact reconstruction. The compression parameters include normalization coefficient, first value of the normalized ECG signal and Sign epoch marker symbol. These parameters are explained in the following sections.

Each sample of ECG signal is represented by floating values and the plot is shown in Fig. 3a. Normalization is the process of converting these floating point values to integers thus reducing the computational cost required for further processing. This is achieved by multiplying each of the samples by an integer called normalization coefficient (such as 200) as in Eq. 1. The Normalized ECG is shown in Fig. 3b. The normalization coefficient and the first value in the normalized ECG are appended to the normalized ECG signal.

$$ECG_n = ECG_{org} \times NC \quad (1)$$

where $ECG_n$ is the Normalized ECG, $ECG_{org}$ is the Original ECG, NC is the Normalization coefficient.

The normalization process reduces the number of bytes required to store the data from 4 bytes (floating point value) to 2 bytes (integer value). This also reduces the amount of computation needed, by cutting down the floating point operations. This allows the compression algorithm to be successfully implemented on a low computational device such as a mobile phone.


**Fig. 2** Block diagram for LECS compression

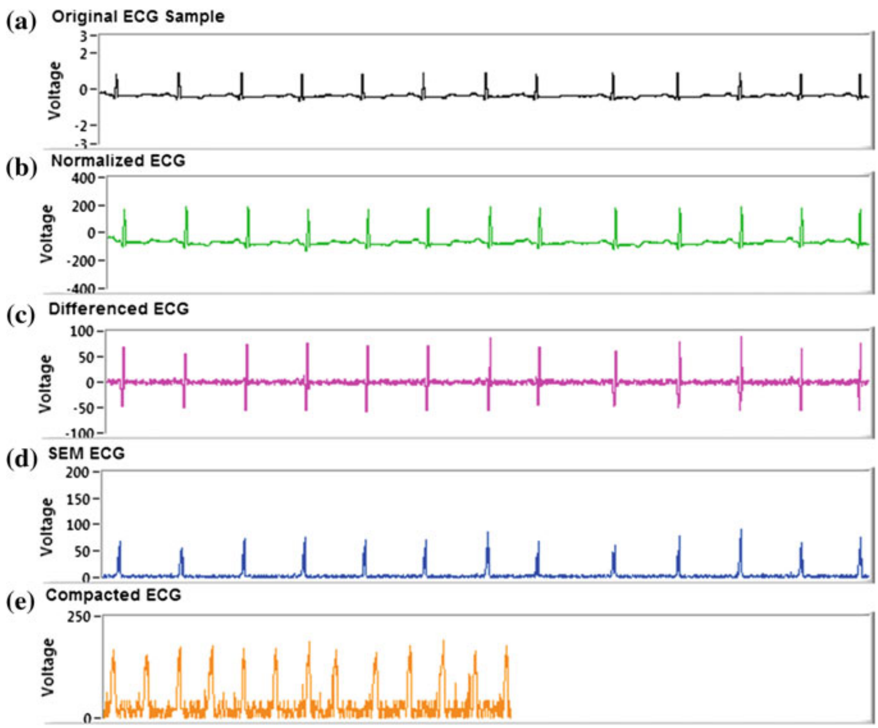


**Fig. 3** Compression steps for record 100

The normalized ECG signal differenced as given in Eq. 2. The Differenced Norma-lized ECG, $ECG_d$, is shown in Fig. 3c.

$$ECG_d = ECG_n(i) - ECG_n(i+1) \tag{2}$$

The key process before encoding is the Sign Epoch Marking (SEM). Sign Epoch Marking is done by inserting the Marker symbol at the start of a series of values with the negative sign value. After sign marking the negative values are translated as to positive given by ECGst prior to encoding as shown in Fig. 3d.

The sign translated values represented by $ECG_{st}$ are then encoded using a symbol substitution table. The table may be commonly agreed upon by both sender and receiver to avoid the transmission of the symbol substitution table for every acquired ECG signal. While compacting consecutive pairs of values are taken from $ECG_{st}$. To amplify the compression, the ECG is compacted by employing the following steps.

Values are taken in pairs. If both the values in the pair are single digit values, it is then encoded using a single symbol indexed by $SS_{ref}$ (given in Eq. 3) in the symbol substitution table.

$$SS_{ref} = ECG_{st}(i) \times 10 + ECG_{st}(i+1) \tag{3}$$

When double digit values are encountered in $ECG_{st}$, each value in the pair (either single digit or double digit) value is encoded using a separate character indexed by $SS_{ref1}$, $SS_{ref2}$ (Given in Eqs. 4, 5). The Compacted ECG is shown in Fig. 3e.

$$SS_{ref1} = 100 + ECG_{st}(i) \tag{4}$$

$$SS_{ref2} = 100 + ECG_{st}(i+1) \tag{5}$$

Occasionally occurring three digit values are encoded using the symbol indexed using the whole three digit value prefixed by a fixed symbol used to denote that the symbol has been used to encode a three digit value. The output produced is a text file with symbols used in the symbol table that could be transmitted through SMS.

## 3 Proposed LECS Decompression Algorithm

Decompression is the exact reverse process of compression as given in Fig. 4. The compressed ECG is first decoded using the symbol table. The indices of the symbols constitute decompressed output. The values are then incorporated with sign using the marker symbol used in SEM step of LECS Compression algorithm. The compression parameters appended in the front of the compressed ECG consist of NC, first value of normalized signal. The first value of the normalized symbol is used for reverse differencing followed by division by NC called Inverse Normalization. The complete reverse of the compression method ensures exact reconstruction of the ECG signal.

## 4 Experimental Setup

### *4.1 Data Description*

Experimental evaluation was carried out for all ECG samples from the Massachusetts Institute of Technology Arrhythmia database (MIT BIH adb) and MIT Compression Test Database (MIT cdb). The MIT BIH Arrhythmia database contains 48 half hour excerpts of two channel (Modified Limb Leads MLII, V1)

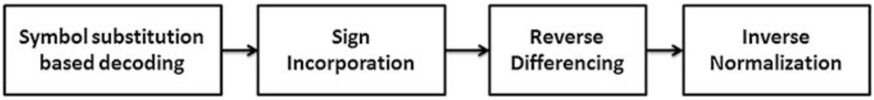


**Fig. 4** Block diagram for LECS decompression

ambulatory ECG recordings. The modified limb leads are used since Holter monitoring is carried for extended periods of time (over 24 h), physical activity is obstructed when ECG is acquired using standard limb leads. The modified limb leads are use to avoid the interference and positioned to closely match signals acquired from the standard limb leads. Normal QRS Complexes and ectopic beats are prominent in the modified limbs. MIT ADB includes 23 random recordings and 25 selected recordings to represent clinically significant but less common arrhythmias. Each recording is sam-pled at 360 Hz, each sample represented by 11 bits over a 10 mV range [19].

## *4.2 Symbol Table from ETSI Character Set*

SMS, MMS protocols support only limited character sets (like ETSI GSM 03.38) wherein even lossless compression methods are susceptible to truncation and data loss during transmission. Generally text in an SMS may be encoded using any of the three available standards: GSM 8-bit character set, GSM 7-bit character set or Unicode UCS-2. The maximum size of an SMS is limited to 1,120 bits in the GSM network. When using GSM 8-bit character each character is encoded using an 8-bit character code. Therefore when encoding using GSM 8-bit character set the maximum size of an SMS is 140 characters. Special characters in SMS text which are not present in the GSM 8-bit character set are encoded using Unicode UCS-2 where each character is represented using 16 bits reducing the maximum character limit of SMS to 70. But it is possible to squeeze in a few more characters into the SMS if the SMS contains only characters available in the GSM 7-bit character set. 160 characters encoded using GSM 7-bit code can be accommodated in the limited 140 8-bit codes. Thus using characters present in GSM 7-bit character set increases the 140 character limit and allows 160 characters to be transmitted in a single SMS. Both GSM 7-bit and GSM 8-bit character sets are defined in the ETSI GSM03.38 standard. Constructing a symbol table with only characters from the GSM 7-bit character set allows more characters to be accommodated in a single SMS and ensures untrimmed data transmission. The proposed method employs a symbol table constructed from 124 printable characters of GSM 7-bit character set defined by the ETSI GSM 03.38 standard to encode the sign translated ECG. Figure 5 shows all the 124 characters used for constructing the LECS symbol table.

**Fig. 5** GSM 03.38 based reduced character set for encoding

@ £ $ ¥ è é ù ì ò Ç Ø ø Å å Δ _ Φ Γ Λ Ω
Π Ψ Σ Θ Ξ Æ æ ß É ! “ # ¤ % & ‘ ( ) * +
, - . / 0 1 2 3 4 5 6 7 8 9: ; < + > ?
¡ A B C D E F G H I J K L M N O P Q R S
T U V W X Y Z Ä Ö Ñ Ü § ¿ a b c d e f g
h i j k l m n o p q r s t u v w x y z ä
ö ñ ü à

## 5 Performance Measures

Generally compression schemes are evaluated on two aspects: the compression and distortion performances. The compression performance is quantified as compression ratio, and is defined as the ratio between the number of bits in the compressed signal number of bits in the original signal (As in Eq. 6).

Compression Ratio (CR):

$$CR = n(ECG_{org})/n(ECG_{comp}) \tag{6}$$

where n() represents the number of bits in original ECG signal ($ECG_{org}$) and compressed ECG($ECG_{comp}$) respectively.

The distortion of the compression scheme is measured by different metrics such as Percent Root-mean-square Distortion (PRD). PRD quantifies the error percentage between the original signal and the reconstructed signal (Given in Eq. 7).

Percentage Root mean square Distortion (PRD):

$$PRD = \sqrt{[(ECG_{org} - ECG_{decomp})2/ECG_{org})2]} \tag{7}$$

where $ECG_{org}$ represents samples in the original signal and $ECG_{decomp}$ represents the decompressed signal.

However, in ECG Compression the clinical acceptability of the reconstructed signal plays an important role. The fuzzy and non-algorithmic nature of this factor propels the need for an efficient lossless compression algorithm.

## 6 Results and Discussions

The LECS compression, decompression algorithm was implemented in J2ME (Java Micro Edition) deployable on low computational devices such as mobile phones and the visualization of results was obtained using LabView8.5. The LECS algorithm for records The LECS was evaluated using all the records from both MIT ADB. The Compression phases and the compressed ECG data for record 100 from MIT ADB are shown in Fig. 6.

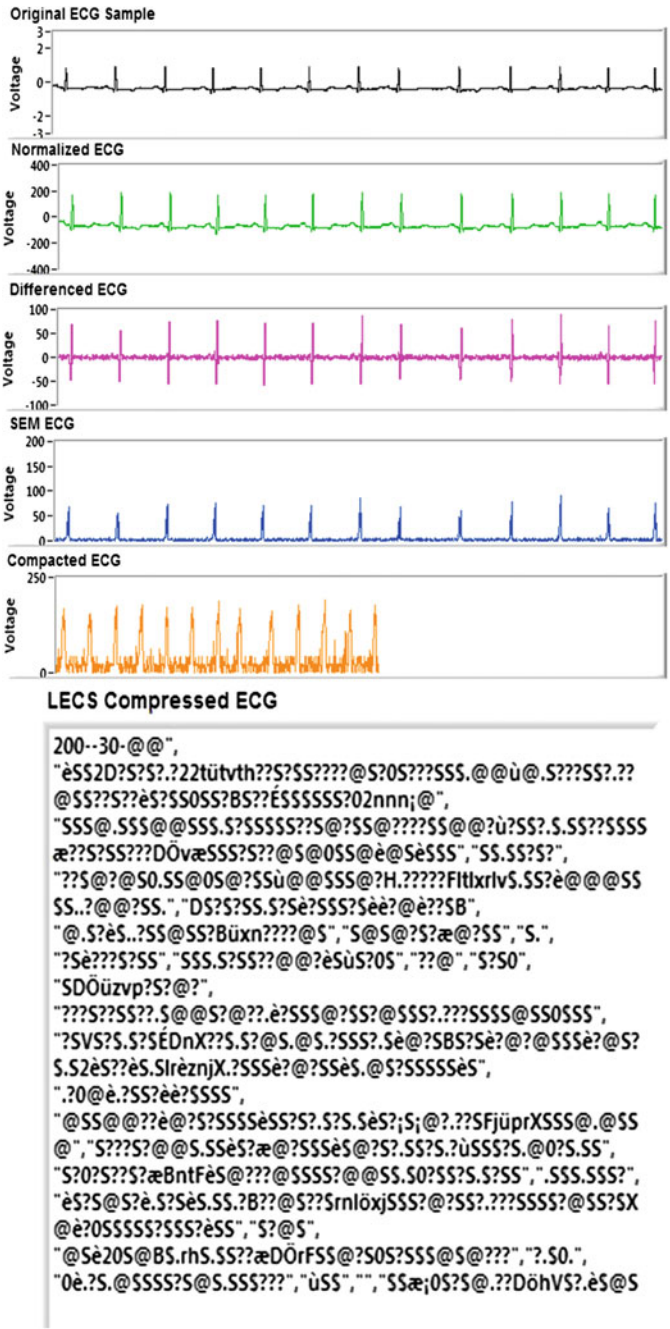


**Fig. 6** LECS compression phases for MIT BIH ADB 100

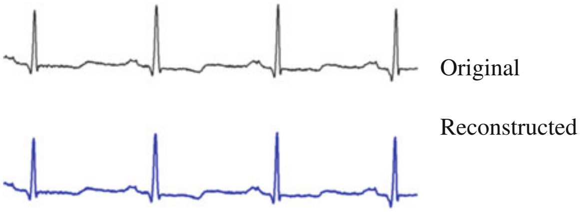


**Fig. 7** LECS decompression—MITBIH ADB 100

**Table 1** Results of proposed LECS for MIT BIH arrhythmia dataset

| Record no. | CR | PRD | Record no. | CR | PRD |
|---|---|---|---|---|---|
| 100 | 7.42 | 0.0200 | 201 | 7.29 | 0.0200 |
| 101 | 7.47 | 0.0224 | 202 | 7.45 | 0.0173 |
| 102 | 7.35 | 0.0224 | 203 | 4.52 | 0.0224 |
| 103 | 7.29 | 0.0224 | 205 | 7.34 | 0.0173 |
| 104 | 7.09 | 0.0265 | 207 | 7.35 | 0.0245 |
| 105 | 7.02 | 0.0245 | 208 | 6.75 | 0.0173 |
| 106 | 6.92 | 0.0200 | 209 | 7.14 | 0.0265 |
| 107 | 6.03 | 0.0332 | 210 | 7.21 | 0.0224 |
| 108 | 7.27 | 0.0224 | 212 | 6.93 | 0.0245 |
| 109 | 6.74 | 0.0245 | 213 | 6.37 | 0.0245 |
| 111 | 6.99 | 0.0200 | 214 | 6.91 | 0.0141 |
| 112 | 7.35 | 0.0200 | 215 | 6.83 | 0.0316 |
| 113 | 7.17 | 0.0265 | 217 | 6.38 | 0.0141 |
| 114 | 7.41 | 0.0173 | 219 | 7.15 | 0.0224 |
| 115 | 7.38 | 0.0173 | 220 | 7.34 | 0.0224 |
| 116 | 6.78 | 0.0141 | 221 | 7.17 | 0.0173 |
| 117 | 7.32 | 0.0224 | 222 | 7.45 | 0.0200 |
| 118 | 6.61 | 0.0141 | 223 | 7.18 | 0.0300 |
| 119 | 7.11 | 0.0173 | 228 | 7.17 | 0.0283 |
| 121 | 7.39 | 0.0173 | 230 | 7.09 | 0.0173 |
| 122 | 6.90 | 0.0200 | 231 | 7.19 | 0.0200 |
| 123 | 7.41 | 0.0173 | 232 | 7.31 | 0.0173 |
| 124 | 7.25 | 0.0245 | 233 | 6.76 | 0.0141 |
| 200 | 6.86 | 0.0200 | 234 | 7.27 | 0.0200 |
| Avg CR | 7.04 | | | Avg PRD | 0.0211 |

The Comparison of original ECG signal with the reconstructed signal for the records 100 is shown in Fig. 7.

The different performance measures for records in MIT ADB (48 records) are given in Table 1. The maximum compression ratio of 7.47 was achieved for record 101. The Minimum compression ratio of 4.52 was obtained for record MIT ADB 203.

The comparison of existing near lossless, lossless with proposed algorithm is presented in Table 2. Sriraam [12], achieved an average CR of 3.23 with PRD of 0.02. Mukhopadhyay et al. [16] achieved a CR of 7.18 and PRD of 0.023 and

**Table 2** Comparison of proposed LECS with existing algorithms

| Author | Signal | Dataset | Algorithm | | CR | PRD |
|---|---|---|---|---|---|---|
| | | | Prediction | Coding | | |
| Sriraam [12] | EEG (12 bit) | University of Bonn & Acquired | Correlation dimension based Prediction | Arithmetic coding | 3.23 | 0.02 |
| Mukhopadhyay et al. [16] | ECG | PTB | Differencing, normalization grouping | ASCII based encoding | 7.18 | 0.023 |
| Boucheham [15] | ECG (11 bit) | MITBIH ADB | Line, curve simplification based prediction | – | 7.71 | 0.025 |
| Proposed | ECG (11 bit) | MITBIH ADB | Differencing, normalization & sign epoch marking | Symbol table based encoding | 7.04 | 0.0211 |

Boucheham [15] achieved a CR of 7.71 and PRD of 0.025. Although the compression ratio of Mukhopadhyay et al. [16] and Boucheham [15] are higher than the proposed LECS algorithm the Percentage Root mean square Distortion (PRD) of the proposed algorithm is lower yielding higher quality in the compressed ECG.

## 7 Conclusions

This paper presents a simple yet efficient lossless compression algorithm using symbol substitution based encoding. The performance of the proposed scheme is evaluated using all records from MIT BIH Arrhythmia database and MIT BIH Compression Test database. The method performs well with both normal and abnormal ECG data, and achieves good CR with low distortion, preserving the morphological features in the reconstructed signal. The simplicity if the technique renders it compatible in low-computational devices such as mobile phones. The symbol table constructed using GSM 7-bit character set ensures lossless transmission through SMS in GSM Networks. Compared with some of the existing methods, the higher compression ratio coupled with higher quality and simplicity can be efficiently put to use in telecardiology and other remote monitoring systems.

**Acknowledgments** The authors acknowledge their cordial gratitude to Dr. Moses. K. Daniel, Professor and head of Medicine Department, Government Rajaji Hospital, Madurai, India for the valuable clinical advice and the effort and time spent in evaluating the quality of numerous reconstructed ECG samples.

## References


1. World Health Organization Information.: http://www.who.int/gho/publications/world_health_statistics/2012/en/index.html
2. Sufi, F., Khalil, I.: Enforcing secured ECG transmission for realtime telemonitoring: a joint encoding, compression, encryption mechanism. Sec. Comm. Netw. **1**, 389–405 (2008)
3. Huffman, D.A.: A method for the construction of minimum-redundancy codes. Proc. IRE **40**, 1098–1101 (1952)
4. Welch, T.A.: A technique for high-performance data compression. Computer **17**, 8–19 (1984)
5. Jalaleddine, S.M., Hutchens, C.G., Strattan, R.D., Coberly, W.A.: ECG data compression techniques-A unified approach. IEEE Trans. Biomed. Eng. **37**(4), 329–343 (1990)
6. Zigel, Y., Cohen, A., Katz, A.: ECG signal compression using analysis by synthesis coding. IEEE Trans. Biomed. Eng. **47**(10), 1308–1316 (2000)
7. Fira, C.M., Goras, L.: An ECG signals compression method and its validation using NNs. IEEE Trans. Biomed. Eng. **55**(4), 1319–1326 (2008)
8. Reddy, B.R.S., Murthy, I.S.N.: ECG data compression using Fourier descriptors, IEEE Trans. Biomed. Eng. BME-**33**(4), 428–434 (1986)
9. Batista, L.V., Kurt Melcher, E.U., Carvalho, L.C.: Compression of ECG signals by optimized quantization of discrete cosine transform coefficients. Med. Eng. Phys. **23**, 127–134 (2001)
10. Chua, E., Fang, WC.: Mixed bio-signal lossless data compressor for portable brain-heart monitoring systems. IEEE Trans. Consum. Electron. **57**(1), 267–273 (2011)
11. Koski, A.: Lossless ECG encoding. Comput. Methods Programs Biomed. **52**(1), 23–33 (1997)
12. Sriraam, N.: Correlation dimension based lossless compression of EEG signals. Biomed. Signal Process. Control **7**(4), 379–388 (2012)
13. Srinivasan, K., Dauwels, J., Reddy, M.R.: A two-dimensional approach for lossless EEG compression. Biomed. Signal Process. Control. **6**(4), 387–394 (2011)
14. Takahiro, O.T.A, Morita, H.: On-line electrocardiogram lossless compression using antidictionary codes for a finite alphabet. IEICE Trans. **93-D**(12), 3384–3391 (2010)
15. Boucheham, B.: ShaLTeRR: a contribution to short and long-term redundancy reduction in digital signals. Sig. Process. **87**(10), 2336–2347 (2007)
16. Mukhopadhyay, S.K., Mitra, S., Mitra, M.: A lossless ECG data compression technique using ascii character encoding. Comput. Electric. Eng. **37**(4), 486–497 (2011)
17. Mukhopadhyay, S.K., Mitra, S., Mitra, M.: An ECG signal compression technique using ASCII character encoding. Measurement **45**(6), 1651–1660 (2012)
18. Sufi, F., Fang, Q., Khalil, I., Mahmoud, S.S.: Novel methods of faster cardiovascular diagnosis in wireless telecardiology. IEEE J. Sel. Areas Commun. **27**(4), 537–552 (2009)
19. Goldberger, A.L., et al.: PhysioBank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. Circulation **101**(23), e215–e220 (2000)