Int. J. Nonlinear Anal. Appl. 13 (2022) No. 1, 603-613
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2022.5542

An innovative and robust technique for human identification and authentication based on a secure clinical signals transmission

Baqer A Hakim[a,*], Ahmed Dheyaa Radhi[b], Fuqdan AL-Ibraheemi[c]

[a]*College of Dentistry, University of Al-Ameed, Karbala PO Box 198, Iraq*
[b]*College of Pharmacy, University of Al-Ameed, Karbala PO Box 198, Iraq*
[c]*Department of Computer Engineering and Information Technology, Faculty of Engineering, Razi University, Kermanshah, Iran*

(Communicated by Madjid Eshaghi Gordji)

Abstract

EEG (Electroencephalogram) is brain waves measure. It is available test allowed to discover the brain functions over time. The brain troubles are evaluated by EEG. It is used to locate the activity in the brain during a seizure and to consider the patients who suffer from brain functionality problems. These troubles include tumors, coma, confusion and long-term difficulties (such as weakness associated with a stroke). The acquisition of EEG signals requires contact and liveliness and these signals are changes under stress that make so potentially unnecessary if it is acquired under menace. In this paper, an innovative and robust solution for this problem is introduced. To this end, the manner depends on models of various data compression models of information-theoretic plus the metrics symmetry related to Kolmogorov complexity. The proposed procedure compares two EEG segments and clusters the data into three groups: a corresponding record for each participant, a distinct person for each group, and self-participant. The technique was used to determine the database participant based on EEG signals. Using a distance measuring approach suggested in this scheme, a 1-NN classifier was constructed. Nearly every person in the underlying database could be accurately identified by the classifier with 96% accuracy.

Keywords: Electroencephalogram (EEG), Pearson Correlation (PrCo), Euclidean Distance (EucDis), Signal Encryption.
2010 MSC: Please write mathematics subject classification of your paper here.

*Corresponding author

Email addresses: h.bagher@alameed.edu.iq (Baqer A Hakim), ahmosawi@alameed.edu.iq (Ahmed Dheyaa Radhi), fuqdan@razi.ac.ir (Fuqdan AL-Ibraheemi)

Received: June 2021 *Accepted:* September 2021

1. Introduction

With technology development, the concern of person identification is increased based on biometric data (such as fingerprint, iris, face, etc.) or systems based on password and username. However, these means are vulnerable to intrusion by professional hackers or being susceptible to be falsified [3].

According to recent studies, EEG signals have a lot of potential in biometric systems. To identify pathological characteristics of these signals, it is common to strive to reduce the variability that distinguishes the EEG signal. EEG signals are fascinating signals in many biometrics applications due to their variety, which is a source of richness [4].

The biometric identification of EEG signals can be classified into fiducial and non-fiducial. Fiducial used to build the internal data model while non-fiducial is alternative methods. When the EEG wave is influenced by noise, it has a significant error. Several non-fiducial methods need R-peak delineation to align windows, but other methods such as autocorrelation function analysis are entirely non-fiducial and do not require window alignment [5]. The use of EEG biometric identification is required. Because of the circadian cycle or certain conditions, the EEG shows changes in rhythm or amplitude. The methods used to comprehend EEG waves may be susceptible to variations. The system was created to deal with a variety of noise sources, including movement, muscle, and so on. The application's classification strength and efficiency have been demonstrated. This is an example of a parameter-free application, such as a compression method.

This methodology makes advantage of free data mining parameters to overcome the challenges that standard EEG methods based on biometrics face [6].

KChSoC was utilized in this study to overcome the majority of the issues and constraints. In this case, $K(x)$ of The KChSoC of x may be defined as the smallest amount of software that creates the variable x before stopping (x is a string of binary digits). To solve this difficulty, because KChSoC is a non-mathematical technique, it is usually approximated by a computable measure, such as stress-based language complexity metrics. Limble-Ziv [11].

2. Related Works

Recently, EEG signals compression field has been very popular in the area of information technology transfer. The following are some of the researchers who presented their research in that field:

In 2011, Srinivasan et al. [14] presented a technique for EEG signals various lossless compression. The technique discussed a simple pre-processing technique where EEG signal before compression is organized in 2D matrix form, and discuss 2 stages for 2D matrix signals compression with SPIHT and arithmetic coding [1].

Sriraam described a compression of EEG signals quality on demand utilizing ANN to achieve a higher compression at a lower bit rate while retaining clinical information content in 2011 [15].

In 2012, Sriraam developed a unique and efficient high-performance EEG lossless compression method based on WT predictors and artificial neural networks. The technique includes pre-processing using IWT, prediction with predictors, and changing the offset of the prediction residual through improved context-based bias cancellation [16].

DWPT decomposition may be used to any form of 1-D biological EEG signal and provides an adaptive strategy for optimum EEG signal encoding in order to compression [2].

In 2015, Thi and Ng published a study of current lossy compression algorithms published in the previous two decades, with the goal of analyzing and comparing them qualitatively [8].

3. Biometric Characteristic of EEG

In order to define the use of EEG as a biometric, it is important to determine how EEG meets biometric characteristics requirements. The ideal EEG biometric characteristic can be:

1. If EEG signals have the universal characteristic.
2. If EEG signals easily measured for the individual when technically easy to obtain the characteristic.
3. If EEG signals are unique when two individuals have not identical characteristics.
4. If EEG signals are permanent when the characteristic of EEG signals does not change over time. Table 1 shows the normal ranges of wave amplitudes and durations. A perfect EEG biometric characteristics can be extent satisfy the requirements depending on the objective and the application of the biometric system. The EEG signals is the sign of life. Their difficulty lies in their uniqueness and continuity. Fig 1 shows the variation of EEG of one individual, and fig 2 shows the normal EEG signal for a normal human being. Fig 1 shows EEG standard signals and fig. 2 shows the different normal EEG waves of the normal human. The approximations supply maximum borders of complexity that permit developing the similarity measurements and dissimilarity measurements for the most current implementations of this concept. The idea in this proposed work is to the determination of EEG Biometrics scheme based on theoretical data models for data compression.

4. Kolmogorov-Chaitin-Solomonoff Complexity (KChSoC)

KChSoC in the information theory is the length of the shortest computer program (such as text length) that produces the object as output. KChSoC considers computational resources need to appoint the object and also known as KChSoC -Chaitin, program-size complexity, descriptive complexity, or algorithmic entropy.
If you use a string like 111111... it will continue to count % the same manner. Although the string is 100 characters long, you can simply produce it using a small application. Consider the string "232046622087638...," and so on for a thousand numbers. It's meant to be a random string, and writing a software to print something shorter than that would be quite tough. In other words, other from the name, there is no way to describe this seemingly random text. The concept for KChSoC developed from this observation of the differences between these two chains. Below is a description of KChSoC's Eq. (4.1) [12]:

$$C_j(x) = \begin{cases} \min(|p| : f(p) = x), & \text{if } x \in \text{ran f;} \\ \infty, & \text{otherwise.} \end{cases} \tag{4.1}$$

5. Symbolic Aggregate Approximation Algorithm (SAXA)

SAXA is a programming language that converts time series transformations into strings. The technique is built on inheriting the original algorithm's simplicity and cheap computing complexity while offering sufficient range query processing selectivity. SAXA is divided into two stages:

1. The original time-series is transforms Piecewise Aggregate Approximation of time series (PAA representation)
2. Transforms Aggregate Approximation of time series (PAA) representation to the string.

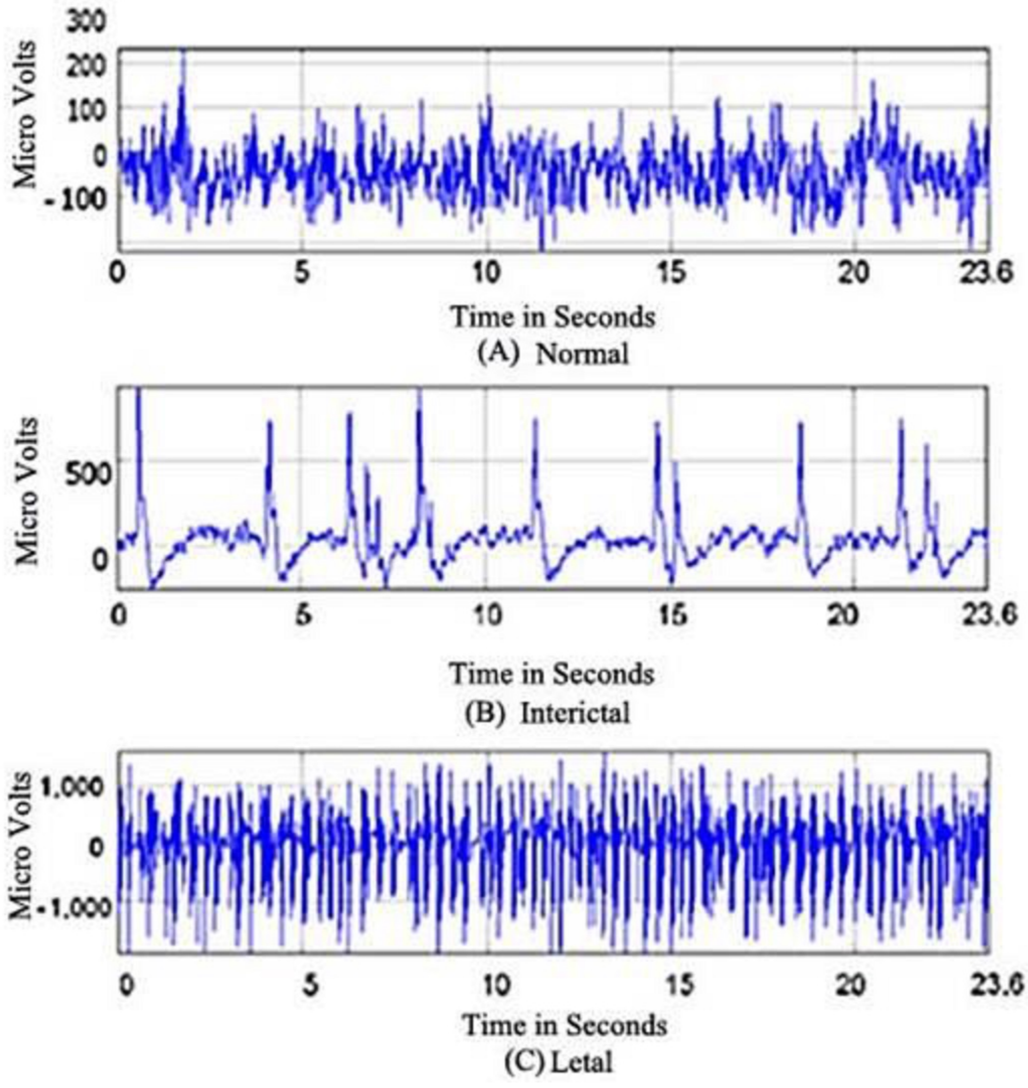

Figure 1: Sample of EEG Standard Signals

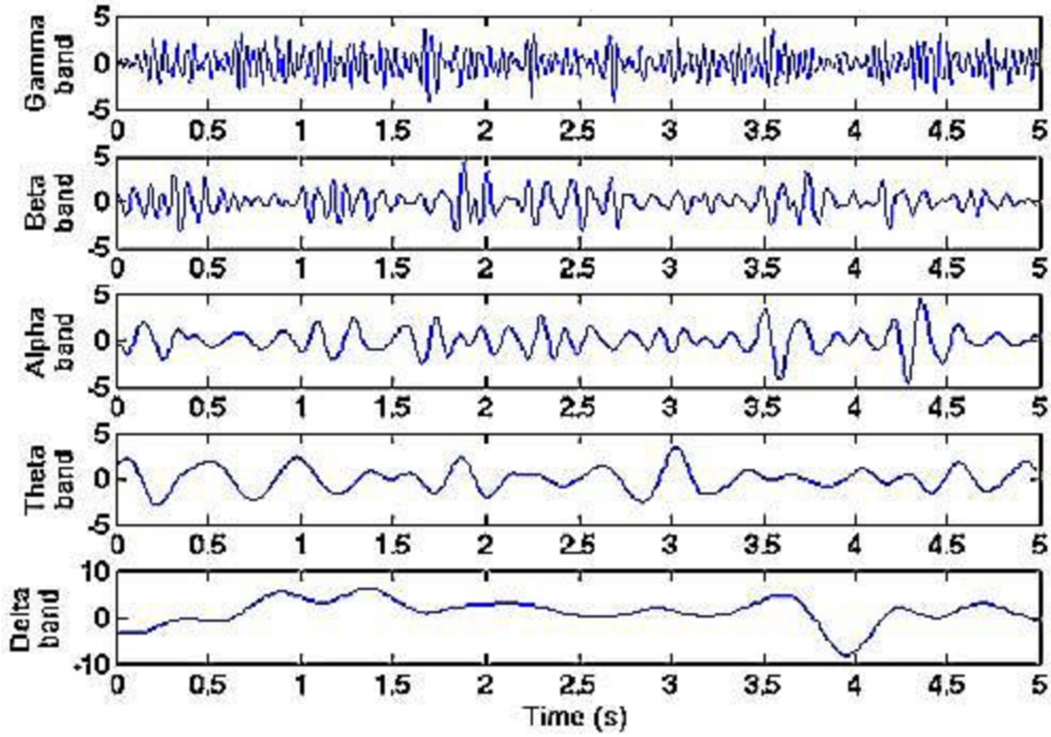

Figure 2: Normal EEG waves

Using an alphabet A of size $a > 2$, SAXA converts a time series Y of length n into a string of arbitrary length xx, where $x \leq nx \leq n$ s usually. PAA is a time-based approximation. series X of length n into sector $X' = (x'1, ..., x'M)X' = (x'1, ..., x'M)$ of any arbitrary length $M \leq nM \leq n$ where each of $x\bar{i}x\bar{i}$ is calculated as follows q. (5.1) and (5.2)[10]:

$$c^{*}i = alpha * j,\ iif, C' * i \in [\beta j - 1 - \beta j) \tag{5.1}$$

$$MINDIST(\hat{Q}, \hat{C}) \equiv \frac{\sqrt{n}}{\sqrt{w}} \sqrt{\sum_{i=1}^{w} (dist(q * i, c * i))^2} \tag{5.2}$$

6. Procedure Description

6.1. EEG Dataset

This method was developed utilizing EEG data from the CHB-MIT Scale database. Despite the fact that the dataset included both pathologies and controls, only the control group was chosen. Only the EEG sample frequency of 1000 Hz was used. The band-stop filter (a basic filter that passes unmodified frequencies but attenuates those frequencies to extremely low levels in a specified range) was used to remove EEG signal interference. The MFMF (Maximally Flat Magnitude Filter) was used to filter the high-frequency interference. It's a form of DSP filter that was created with the goal of having a flat frequency response in the passband.
The goal of this study was to reduce external variability. To select the same database, make the following decisions [13]:

1. Build a real database consists of 75 participants and 300 medical records.

2. Each participant record must be split into five segments of 20 seconds.

Due to all the participants have not more than one medical record, then considered only one medical record from each participant.

6.2. Quantization

The technique is centered on preserving the original algorithm's simplicity and cheap computing complexity while offering adequate sensitivity and selectivity while processing range queries. Because patient medical records (PMR) are kept on a computer, they are digital and symbolic records. As a consequence, compression algorithms work well with symbolic data; nevertheless, the EEG is a numerical signal with a real value. The conversion of actual EEG signals to symbolic signals is the most important step in this situation. The SAX algorithm must be used to finish the converting phase. Time series is divided into N segments and the mean value of each symbol is calculated. These are the most essential elements when it comes to this approach: font size and w dimension [9]. The quantization is the lossy method, so must be a guarantee to preserved the information of EEG to guaranty the correct EEG analysis. In this method, in order to gain the optimal parameters, it performed a set of quantization. The dimension of alphabet size between 2 - 20 symbols, but the width of the windows is 10x20 sample points.
For each quantization phase, must apply EucDis and PrCo or called Moment Correlation (MoCo) between original and the quantized signal were calculated [14].

6.3. Pearson Correlation (Moment Correlation) (PrCo) or (MoCo)

PrCo coefficient is the variance between two-variable and the output is divided by its standard deviations. The above definition format includes the product time, i.e. the product average of random variables adjusted according to the mean; Hence the product time-modifier in the name. The degree to which two data sets are connected can be evaluated by their correlation. In statistics, the most widely discussed measure of correlation is the (PrCo). The linear connection between two data sets is provided by the PrCo. The PrCo or MoCo is usually represented by one of two letters [7]:

- Greek letter rho (ρ) for a population.
- The letter "r" for a sample.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x2 - (\sum x)][n \sum y2 - (\sum y)2]}} \tag{6.1}$$

The principle of algorithm action is the size of the alphabet and the vector w of the string the series segments.

6.4. Compression Method

The measure of similarity between binary objects is the main problems dealt with using the KChSoC theory. A method was suggested, based-on the distance information depends on the length of the shorter binary program. This distance relies on conditional complexities. An algorithm able to collect the data knowledge to be compressed was proposed in this paper in order to understand the basic of KChSoC.
It is necessary to collect statistics in order to create an internal representation of the received data. To calculate similarity between binary items, uses the idea of relative compression between two

objects by utilizing the knowledge of another object. The NRC (normalized relative compression) of x specified y is described by Eq. (6.2):

$$NRC(x,y) = \frac{C(x\|y)}{|x|} \tag{6.2}$$

Where $|x|$ is object size.
Data amount of x that cannot describe by y is information gave by this measure. To build the internal model of y, the combination $C(x\|y)$ of several orders (k) of finite-context models (FiCoMd)can be used, and utilizing the model built from y, x can be encoded. There are three parameters can be considered: κ, Υ and α. By depending on If two EEG signals are compared, the values will get for the same participant are lower than values compared to another participant.

6.5. The Statistical Analysis

Training dataset for each participant consists of four segments, whereas the residual segment was utilized for testing. To assess the suggested method, there are two steps:
First step: The findings were shown as the mean and standard deviation of the computed measure NRC between records. The outcomes were categorized as self-evaluation (SelfEval) NRC value of the record with itself, participant measure NRC value computed across recordings of the same person, eliminating self-similarity, and out measure (NRC value calculated between records of different participants).
Second step: Performance assessing of the identification method is the basic theory in this research paper that is evaluated by sensitivity means, accuracy and specificity (Eq.(6.3),(6.4) and (6.5)).

$$sensitivity = \frac{TP}{TP + FN'} \tag{6.3}$$

$$specificity = \frac{TN}{FP + TN'} \tag{6.4}$$

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN'} \tag{6.5}$$

TP, TN, FP, and FN denote the number of true positives, true negatives, false positives, and false negatives, respectively. Below fig 3 show the original EEG signals and reconstructed signals.

Sensitivity is the ability of a classifier to correctly identify a person (participant). Accuracy gives the ratio of actual results (true positives and negatives) in all possible cases. Specificity measures the classifier's ability to correctly reject someone.

7. Results and Discussion

In this investigation, the CHB-MIT scalp EEG database was employed. Only the healthy entrants are taken into consideration. In this database, there are 40 entrants. Data was gathered on different days for some individuals. The testing is based on each participant training one record and dividing it into five 20-second parts. The finitial 15 participants are used to optimize model parameters, while the remaining individuals are utilized to evaluate parameters. Reciprocal information, PrCo, and the EucDis between quantized EEG signals were utilized to assess quantization performance.
The EEG waves was transformed into the symbolic sequences allowing the application of the NRC measure. This measure need three parameters: κ, Υ and α. Using the first parameters in the proposed EEG method, a mixture model was used for the specific FiCoMd with $\Upsilon = 0.2, \kappa =$

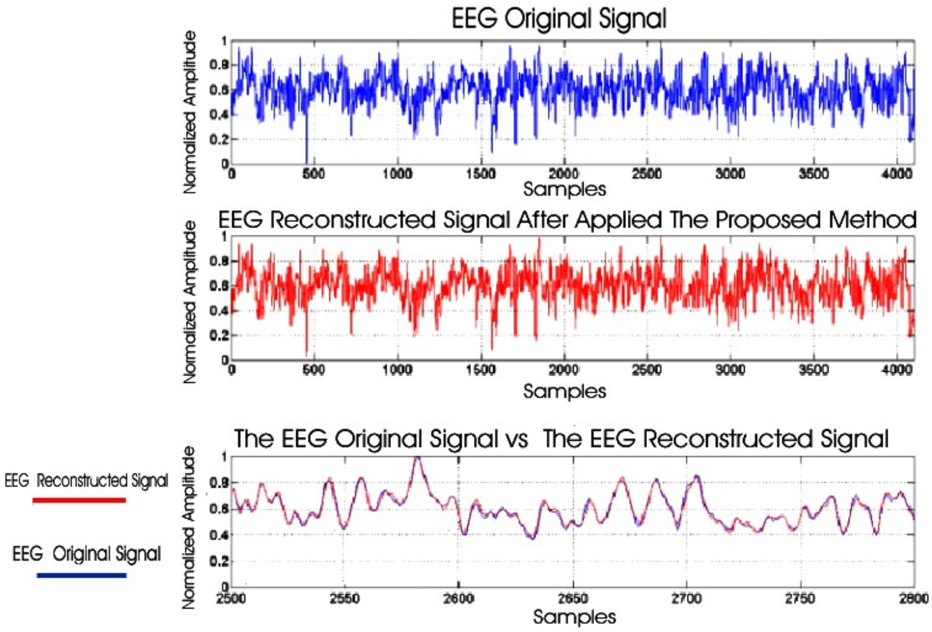

Figure 3: The EEG Original and Reconstructed Signals

Table 1: The Ratio of Percent of PRD (Root mean square Difference) for proposed method and other average method

Compression Ratio (CR)	PRD (%)	
	Proposed Method	Related Literature (Average methods)
20	9	7
30	13	11
40	19	16
50	23	20

set of $\{1;2;3;4;5;6;7;8;9\}$, and $a = \{1;1=50;1=100;1=200;1=300;1=400;1=500;1=1000\}$. In order to evaluate the performance of the proposed method, it has been used to minimize the SelfEval of the record with itself.

The procedure of biometric identification is detailed in Table 1. The NCR can tell the difference between the three data sets, the EEG-training-segments, and if the participant is the same or different: (a). (a) EEG-training-segments with a mean value of 0:140; (b) EEG-testing-segments with a mean value of 0:300 from the same participant; and (c) EEG-testing-segments from separate individuals (with a mean value of 0:410).

The amplitude distribution and spectral distribution of EEG segment is shown in Figure 4.

Table 2: The performance analysis of EEG depending on PRD,PSNR and MSE

EEG Dataset	PRD	Time Duration (sec)	PSNR	MSE
Dataset(1)	2.88	1.44	12.989	2.04E+06
Dataset(2)	2.96	1.42	17.46	8.94E+05
Dataset(3)	3.23	1.42	22.794	7.99E-16
Dataset(4)	3.41	1.43	25.336	9.31E+12

Table 3: Quantization levels of different samples of dataset

EEG Dataset	Threshold Level (%)	Level Quantizing (1)	Level Quantizing (2)	Level Quantizing (3)
Dataset(1)	2	10	9	8
Dataset(2)	5	9	8	7
Dataset(3)	9	6	5	4

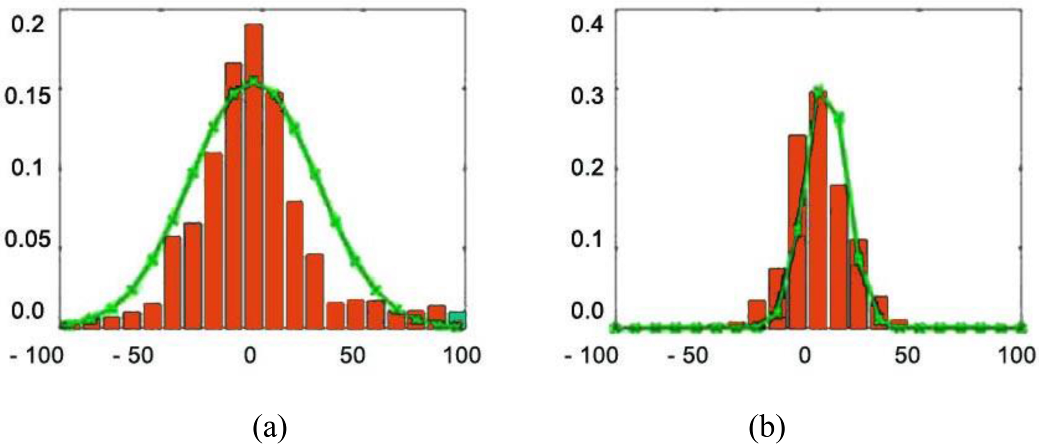

Figure 4: Amplitude distribution histogram of two EEG Signals

Table 4: A gaining results from a tested sample

Patient ID	P(1)	P(2)	P(3)	P(4)
Primary CR	93.14	93.2	93.1	93.22
Optimized CR	92	91	91	92
Primary PRD	77.41	56.77	56.01	69.86
Optimized PRD	8.8	7.25	8.32	8.46
Stopping Iteration	2	3	2	3

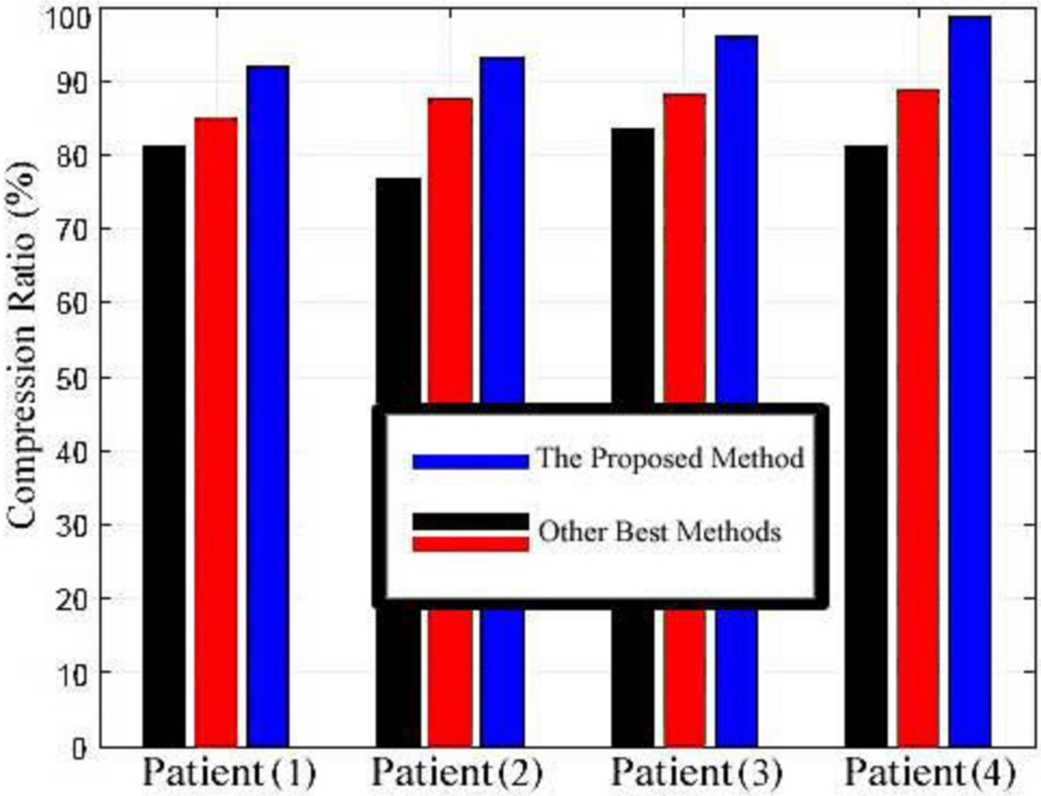

Figure 5: Comparison between the proposed algorithm using KChSoC function with other Algorithms

8. Conclusions

The proposed model allowed to comparing two EEG segments without applying the traditional approaches that need to brainwaves segmentation. The model was described to be feasible approach to EEG biometric analysis, due to this method presents 99% for accuracy on the database used. The model has certain limitations; thus it needs to improve its acceptance rate. The EEG waves technique was validated by collecting data on several days and taking into account environmental variability circumstances. In addition to identifying, the approach may also be used to authenticate a person's identity. New records are compared to existing records and only if the NCR value falls within an acceptable range for an individual are they accepted.

References

[1] A.S. Abdulbaqi, A.J. Obaid and S.A. Hmeed Alazawi, *A smart system for health caregiver based on IoMT: toward tele-health caregiving*, Int. J. Online Biomed. Engin. 17(7) (2021) 70–87.

[2] A.Sh. Abdulbaqi, A.J. Obaid and A.H. Mohammed, *ECG signals recruitment to implement a new technique for medical image encryption*, J. Discrete Math. Sci. Cryptog. (2021) 1–11.

[3] A.S. Abdulbaqi and I.Y. Panessai, *Designing and implementation of a biomedical module for vital signals measurements based on embedded system*, Int. J. Adv. Sci. Technol. 29(3) (2020) 3866–3877.

[4] M. Abo-Zahhad, S.M. Ahmed and S.N. Abbas, *State-of-the-art methods and future perspectives for personal recognition based on electroencephalogram signals*, IET Biom. 4(3) (2015) 179–90.

[5] B.C. Armstrong, M.V. Ruiz-Blondet, N. Khalifian, K.J. Kurtz, Z. Jin and S. Laszlo, *Brain print: assessing the uniqueness, collectability, and permanence of a novel method for ERP biometrics*, Neurocomp. 166 (2015) 59–67.

[6] K. Brigham and B.V.K.V. Kumar, *Subject identification from electroencephalogram (EEG) signals during imagined speech*, 2010 Fourth IEEE Int. Conf. Biomet. Theory, Appl. Syst. (2010) 1–8.

[7] A. Buda and A. Jarynowski, *Life-Time of Correlations and its Applications*, Andrzej Buda and Wydawnictwo Niezależne, 2010.

[8] P.T. Dao, X.J. Li and H.N. Do, *Lossy compression techniques for EEG signals*, 2015 Int. Conf. Adv. Tech. Communic. (2015) 154–159.

[9] D. Del Testa and M. Rossi, *Lightweight Lossy compression of biometric patterns via denoising auto encoders*, IEEE Signal Proc. Lett. 22(12) (2015) 2304–2308.

[10] S. Fauvel and R.K. Ward, *An energy efficient compressed sensing framework for the compression of electroencephalogram signals*, Sensors 14(1) (2014) 1474–1496.

[11] H.U. Jian-feng, *Multi feature biometric system based on EEG signals*, Proc. 2nd Int. Conf. Interaction Sci. (2009) 1341–1345.

[12] M. Li and P. Vitanyi, *An Introduction to Kolmogorov Complexity and Its Applications*, Springer Science & Business Media(SSBM), 2009.

[13] A. Shukla and A. Majumdar, *Row-sparse blind compressed sensing for reconstructing multichannel EEG signals*, Biomed. Signal Proc. Cont. 18 (2015) 174–178.

[14] K. Srinivasan, J. Dauwels and M.R. Reddy, *A two-dimensional approach for lossless EEG compression*, Biomed. Signal Proc. Cont. 6(4) (2011) 387–394.

[15] N. Sriraam, *Quality-on-demand compression of EEG signals for telemedicine applications using neural network predictors*, Int. J. Telemed. Appl. (2011) 860549.

[16] N. Sriraam, *A high-performance lossless compression scheme for EEG signals using wavelet transform and neural network predictors*, Int. J. Telemed. Appl. (2012).