1. Introduction

Data compression is critical in digitally loading information on computer discs and conveying it to communication links. There are two basic compression techniques; lossy (Inexact) or Lossless (Exact) can be used for data compression. To produce the original data, lossless compression may be reversed, whereas lossy compression removes some information or presents minor errors in reversal. For text, lossless compression is appropriate, where each digit is significant, while lossy compression might be reasonable for pictures or speech (The limitation of the frequency spectrum is an example of lossy compression in telephony) [1]. In lossless compression, bits have been reduced by recognizing and removing redundancy in statistics. One of the most lossless compression techniques is “Run Length Encoding (RLE)” which is considered a simple framework that provides a strong loss-less compression for the data values who has the same value [2]. In this algorithm, the repeated run of characters has been coded by using two elements: the first one represents the counter which memorizes how long the data is; the second one recorded the repetitive element that constitutes the data. If a data item (a) occurs at (b) consecutive times in the input data stream, RLE replaced the (b) occurrences with the single pair (ab) as shown in Figure 1.

![Figure 1: Run Length Encoding algorithm.](image-url)

Digital revolution communication has some drawbacks like accidentally the secret message could be potentially intercepted and read on its way to the destination. Cryptography here could be a good solution to this problem which involves encrypting such a secret message, hence only the destination person has a key to decrypt it, then privacy will be achieved [3]. Therefore, digital communication should be based on a design that is articulated in two basic operations which are heterogeneous or in some cases conflicting, but it is needed to be applied to the secret file to ensure encrypting efficiency and security. This study handled data encryption using a lossless data compression application.

Different studies handled data security issues as in [4] where hybridization of chaotic maps with DNA cryptography has been used to hide and encrypt both medical images and medical reports. Ponnambalam et al. [5] have proposed a system to convert Electrocardiograph (ECG) signals into QR codes. Mustafa et al. [6] have proposed an encryption algorithm based on the Advanced Encryption Standard algorithm (AES) to
improve the security of ECG signals through transmission using 
three different keys. Triple Data Encryption Standard (Triple-
DES) has been implemented in [7] with the aid of the hash 
function and DNA cryptography base to encrypt different iosignals (ECG, EEG, and EMG) into DNA format to add a 
security layer for the proposed system.

On the other hand, some studies focused on image 
compression as in [8] in which a hybrid compression technique is 
applied on ninety-hand vein images to combine both advantages of 
lossy and lossless techniques, their goal is to maximize the 
Compression ratio by preserving the images' details. Their 
applied technique resulted in a compression ratio reaching 
89.56%. in [9] the authors aimed to maximize the compression 
ratio by preserving images' information, they proved that the 
maximum accepted ratio for each image is chosen by three 
experts. Their last enhancement stage is done by isolating the 
region of interest (ROI) in the image then applying the 
compression procedure by the experts gave promising results.

II. Methodology

In this study, a crypto compression scheme has been 
proposed using the hybridization of a novel updated version of 
RLE called zero-RLE is applied for performing lossless data 
compression on the medical data (biosignals, images, and videos) 
with DNA encoding rules. the compressed data seemed like an 
encrypted form which is required in data security then finally 
the compressed data is encoded into DNA format as shown in Figure 2.

![Figure 2: Block diagram of the proposed compression-encryption scheme.](image)

A. Dataset

In this study, three types of data are utilized in Appling the proposed 
compression procedure by the experts gave promising results. 

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Data description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical Images</td>
<td>Chest X-ray images have been tested for covid-19 [10]</td>
</tr>
</tbody>
</table>
|                 | Electroencephalograph (EEG) signals down were loaded from the DEAP dataset for EEG signals recorded at different emotions [12].
|                 | Electromyograph (EMG) downloaded from EMG data for gestures Data Set These are files of raw EMG data recorded by MYO Thalamic bracelet In case hand clenched in a fist [13]. |
| Medical Videos  | Twenty eyes videos of 24 subjects were included in the study about the Field of View (FOV) of eyes. |

B. Zero Run-length Encoding (zRLE)

Zero-Run Length Encoding is an updated version of RLE which originated in [15] for lossless compressing for 
implantable optogenetic visual prostheses. zRLE involves 
checking the pixel value and comparing it with specific ecif 
threshold; if the pixel value is higher than the threshold value 
hence keeping the pixel value as it is, if the pixel value is less 
than the threshold value hence rep replace with zero, the 
threshold value for each type of data is calculated from 
minimum peaks in the histogram. Finally, check all zeros: (For 
non-repeated zero replaced with one- For repeated zero, 
replaced with (count, zero)). The flowchart of zRLE is shown in 
Figure 3.
C. DNA Encoding

It is a novel technology that involves utilizing DNA molecules in saving data it. In digital applications converting binary data into a DNA sequence that’s adding a new security level for data encryption [16]. Eight probability encoding rules may be used for converting binary data into DNA sequence as shown in Table 2. In this study, Rule-3 is used.

![Figure 3: Flowchart for zRLE.](image)

### Table 2: DNA Encoding rules

<table>
<thead>
<tr>
<th>Rule</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>00</td>
<td>00</td>
<td>01</td>
<td>01</td>
<td>10</td>
<td>10</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>C</td>
<td>10</td>
<td>01</td>
<td>11</td>
<td>00</td>
<td>11</td>
<td>00</td>
<td>10</td>
<td>01</td>
</tr>
<tr>
<td>G</td>
<td>01</td>
<td>10</td>
<td>00</td>
<td>11</td>
<td>00</td>
<td>11</td>
<td>01</td>
<td>10</td>
</tr>
<tr>
<td>T</td>
<td>11</td>
<td>11</td>
<td>10</td>
<td>10</td>
<td>01</td>
<td>01</td>
<td>00</td>
<td>00</td>
</tr>
</tbody>
</table>

### D. System Evaluation Metrics

Several evaluation metrics can be used to evaluate the privacy, robustness, and efficiency of the proposed algorithm. In this study, seven evaluation metrics [17] are used:

- a. Encryption/ Decryption time: represents the time of encryption or decryption.

- b. Histogram: represents the frequency distribution of the pixel intensity values. The histogram of the encrypted image should be completely differenced from the histogram of the original image.

- Compression Ratio = \( \frac{\text{un compressed Size}}{\text{Compressed Size}} \)

- c. Correlation coefficient: measures the strength of the linear relationship between variables, it can be calculated from:

  \[
  r_{\alpha\beta} = \frac{\text{cov}(\alpha, \beta)}{\left(\sqrt{\text{VAR}(\alpha)}\right)\left(\sqrt{\text{VAR}(\beta)}\right)}
  \]  
  \[
  \text{VAR}(\alpha) = \frac{1}{N} \sum_{i=1}^{N} (\alpha_i - E(\alpha))^2
  \]  
  \[
  \text{cov}(\alpha, \beta) = \frac{1}{N} \sum_{i=1}^{N} (\alpha_i - E(\alpha)) \cdot (\beta_i - E(\beta))
  \]

  Where: \((\alpha, \beta)\) are the original and reconstructed data. \((i):\) pixel which calculation is done. \(E(x):\) Data Expectation. \(\text{cov}(\alpha, \beta):\) is the covariance between \(\alpha\) and \(\beta\). \(\text{VAR}(\alpha)\) gives the value of variance at pixel value \(\alpha\), \(N:\) is the total number of pixels.

- d. Mean Squared Error (MSE) analysis: Mean Squared Error (MSE) has been used to check the diffusion characteristic of an image of the cryptosystem. It indicates that the relationship is too complex to be predicted easily.

  \[
  MSE = \left(\frac{1}{W \cdot H}\right) \cdot \sum_{i=1}^{W} \sum_{j=1}^{H} (I(i,j) - K(i,j))^2
  \]

  Where I and K refer to the pixel values of both original and encrypted images. \((i,j)\) refer to the pixel location. \((W \text{ and } H)\) refer to the image dimensions.

- e. Unified Average Changing Intensity (UACI):

  UACI is the difference between the average intensity between the plain and encrypted images. The theoretical value for UACI is about 33%. UACI can be calculated from:
\[ UACI = \sum_{ij} \left| \frac{C1(i,j) - C2(i,j)}{T} \right| \times 100\% \] (5)

\((C1, C2)\): are the original and reconstructed data. \((T)\): represents the whole number of the pixels in the encrypted image, \((F)\): represents the largest supported pixel value (it will be 255).

f. Structure Similarity Index Measurement (SSIM):
SSIM measures the similarity between two images. SSIM can be calculated from:
\[
SSIM(x, y) = \left( \frac{2\mu_x\mu_y + C1}{\mu_x^2 + \mu_y^2 + C1} \right) \left( \frac{2\sigma_{xy} + C2}{\sigma_x^2 + \sigma_y^2 + C2} \right)
\] (6)

Where: \((x, y)\): is the original and reconstructed data. \((\mu_x, \mu_y)\) is the average of \(x, y\). \((\sigma_x, \sigma_y)\) is the variance of \(x, y\). \((\sigma_{xy})\) is the covariance of \(x, y\). \((C1, C2)\) are two constants to stabilize the division with a weak denominator.

i. Compression Ratio: it represents the ratio between compressed data size to original data size.

2. Results

The goal of this study is to propose a crypto-compression system to perform both encryptions for data for security and compression for more transmission speed in a small time to enable medical teamwork in handling medical data safely when transmitting through public channels. Zero Run-length encodings, a lossless compression algorithm, are used in this study to reduce data size with the aid of DNA encoding rules for data encryption. This experiment is implemented on Windows 10 using MATLAB (R2018a) on a personal laptop with the following specifications: CPU 2.7 GHz Core i7, and 8 GB RAM.

Different evaluation metrics have been calculated starting with the histogram of original data and decompressed one which is shown in Fig. 4 where the histogram of the original data was like the histogram of the decompressed one. This proposed security scheme is achieved a satisfactory result and a good performance with values reached to the ideal values (100% similarity and zero MSE) which have appeared after applying these evaluated metrics. The correlation between the original data and the restored one indicated its correlated relation as they were identical.

![Figure 4: The histogram of the original data (a) for biosignal (1D data), b) for image (2D data), c) for Videoframe data) and the decompressed one using zRLE](image)

Table 3 showed the statistical evaluation metrics achieved by the proposed cryptosystem with each type of data. In Table 3, Encryption/ Decryption time for biosignal and images were very small which indicated its ability to real-time applicability. SSIM was 100% which means that the input historical with the output that proves lossless compression, MSE is also equal to zero due to lossless compression with zero error. The compression ratio was differentiated between (10% to 25%).

22
Table 3: Experimental result for the proposed crypto-compression system

<table>
<thead>
<tr>
<th></th>
<th>Medical Images</th>
<th>Biosignals</th>
<th>Medical Videos</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ECG</td>
<td>EEG</td>
</tr>
<tr>
<td>Encryption time (s)</td>
<td>11.6 s</td>
<td>10 s</td>
<td>12.123 s</td>
</tr>
<tr>
<td>Decryption time (s)</td>
<td>7.5 s</td>
<td>8.78 s</td>
<td>11.04 s</td>
</tr>
<tr>
<td>SSIM</td>
<td>100 %</td>
<td>100 %</td>
<td>100 %</td>
</tr>
<tr>
<td>MSE</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UACI</td>
<td>33.3</td>
<td>33.34</td>
<td>33.5</td>
</tr>
<tr>
<td>Correlation Coefficient between original and reconstructed data</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Compression ratio</td>
<td>16 %</td>
<td>25 %</td>
<td>13 %</td>
</tr>
</tbody>
</table>

3. Discussion

Proposing a crypto-compression system requires the essential specification to perform its goal. In this study, zRLE proved its efficiency in performing lossless compression for data without a loss (zero MSE), so we must compare our results with other studies which focused on using RLE in the same application or in using it in another application. Firstly, we have compared our result in medical image compression and encryption with other studies that have the same goal on images as in Table 4:

Table 4: The Comparing of the proposed crypto-compression system of image encryption with other studies

<table>
<thead>
<tr>
<th>Author</th>
<th>Methodology</th>
<th>Data specs</th>
<th>Compression ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sh. Khaleel [18]</td>
<td>Fuzzy Neural Network Clustering with RLE</td>
<td>Tree gray img. 128*128</td>
<td>Tree time.: 13.86 %</td>
</tr>
<tr>
<td>T. Agung [19]</td>
<td>Watermarking scheme with RLE</td>
<td>fetal-lens-1h.bmp 640x480</td>
<td>1.32 %</td>
</tr>
<tr>
<td>Our study of image</td>
<td>zRLE with DNA encoding</td>
<td>Medical in. 784*1024</td>
<td>16 %</td>
</tr>
</tbody>
</table>

Other studies have focused on 1D data encryption like audio signal, and biosignals, here we have compared these studies with other system applications on biosignals. The result of the proposed system proved that our proposed methodology could perform both encryption and compression in a small time with zero MSE which indicated the data has been reconstructed exactly through the process. Also, the compression ratio concerning others is accepted and achieved in a good time as shown in Table 5.
Methodology | Data specs | Compression ration |
---|---|---|
S. Akhter, 2010 [20] (RLE) to compress (DCT) coefficients of ECG. | MIT-BIH Arrhythmia | 14.87 % |
B. Carpentieri, 2018 [21] RLE + 3DES | 1D data: 3229KB | 21 % |
Our study in ECG biosignals RLE + DNA Encoding | ECG biosignals from MIT-BIH Arrhythmia | 25 % |

4. Conclusions

Medical data is very sensitive data as it contains patients’ data besides medical reports, biosignals, medical images, and medical videos. These data can’t be sent online without encryption as it can be hacked, alternated, modified, or deleted. Although there are different methodologies for data compression and encryption, Zero-Run length encoding performs both fast and good compression ratios for data then it is supported with DNA encoding rules to enhance security levels through data transmission. The proposed scheme is very fast and accurate with zero error and 100% similarity percent that can be applied to real-time data compression and encryption. In the future, we will support such a proposed system with one of the cryptographic algorithms like AES to enlarge keyspace and hence increase the security level for the proposed system.

Conflict of Interest

The authors declare no conflict of interest.

Abbreviation and symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>AES</td>
<td>Advanced Encryption Standard</td>
</tr>
<tr>
<td>DCT</td>
<td>Discrete Cosine Transform</td>
</tr>
<tr>
<td>DES</td>
<td>Digital Encryption Standard</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>ECG</td>
<td>Electrocardiograph</td>
</tr>
<tr>
<td>EEG</td>
<td>Electroencephalograph</td>
</tr>
<tr>
<td>EMG</td>
<td>Electromyograph</td>
</tr>
<tr>
<td>FOV</td>
<td>Field of View</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean Squared Error</td>
</tr>
<tr>
<td>RLE</td>
<td>Run Length Encoding</td>
</tr>
<tr>
<td>SSIM</td>
<td>Structure Similarity Index Measurement</td>
</tr>
<tr>
<td>UACI</td>
<td>Unified Average Changing Intensity</td>
</tr>
<tr>
<td>zRLE</td>
<td>Zero Run-Length Encoding</td>
</tr>
</tbody>
</table>

References


