NCBI Medical Data Encryption with Lossless DNA Compression

Anfal Emad Lafta*, Sahar Adil Kadhum
College of Sciences for Women, University of Babylon, Hilla, Babel, Iraq
Emails: scw869.anfal.emad@student.uobabylon.edu.iq; dr.sahar.Adill@gmail.com

Abstract

The health information data includes reports on the patient's condition, including addresses, names, tests, treatments, diagnoses, and medical history. It is sensitive information for patients, and all means of protection must be provided to prevent third parties from manipulation or fraudulent use. It has been discovered that DNA is now a reliable and efficient biological media for securing data. Data encryption is made possible by DNA's bimolecular computing powers. In this paper proposed a new strategy of safeguard the transfer of sensitive data over an unsecured network using cryptography with non-liner function, and DNA lossless compression to enhance security. The work gains best results in compression processes, as percentages range 75%. for character compression, the different rate ranges between 91% to 94%, and the compression rate ranges from 35% to 37%. the retrieving data with an accuracy rate up to 100% without any data loss, as well as excellent percentages within the, Compression Ratio, Compression Factor, Error Rate, Accuracy measures.

Keywords: NCBI; DNA; Cryptography; Clinical data privacy; DNA compression

1. Introduction

Electronic health record is a type of records digital health that is regular be created or updated as well as disseminated through internet for accurate data collection and facilitate efficient thanks to the proliferation of an intelligent and smart technologies[1] . A patient's EHR is a combination of information around patient maintained by the providers of relevant healthcare, containing of all patients demographics[2], Medical and HER history with other relevant information its sent in online record is vulnerable to unauthorized duplication, tampering or eavesdropping, and other forms of theft because its send in real text[3]. Through todays heavy reliance on internet and digital communication networks, protecting private data are more important than ever [4].

Through, the increasing focus on protecting information has continued, especially in recent years, in more effective ways. It has focused on the main concept of securing data using methods that prevent all adversary attempts from decrypting or seizing any clear data. Here the concept of encryption explains the practice of encrypting data or studying how to encrypt [5]. This data is transferred before it is transferred online to public networks by converting clear data into encrypted data that is incomprehensible and difficult to break, so that authorized and authorized persons can view its contents. This will provide protection for the data and prevent unknown persons from viewing or understanding that data[6]. In light of this, different encryption methods are frequently used because of the many benefits they provide to protect data [7].

Here we must point out that one of the main challenges of encryption operations is the clarity of the encryption on the message[8]. In other words, anyone can notice the presence of encryption on the message because the text will appear in an incomprehensible and suspicious way [9], which will attract attention and greatly increase the suspicions of attackers who are trying to access real data in different ways. illegal. Therefore, there has emerged a need for a new method to address the main challenge in encryption processes [10]. The compression could be lossy or irreversible if the cover in the medium can't be reconstructed to its real form [11]. A reversible and lossless

Doi : https://doi.org/10.54216/JCIM.140201
Received: January 17, 2024 Revised: March 08, 2024 Accepted: June 17, 2024

approach successfully reconstructs the cover in medium by errorless [12]. By development the latest methods, cyberattack strategies also are evolving, making it critical to create a robust encryption scheme that is tough to break [13].

2. Related Works

The literature on lossless compression has a wide range of data compression methods and algorithms; different approaches may be available for a given type of data. Before implementing DNA compression using a few technical techniques, First, we will go over a few different approaches to data compression.

Xiangjun Wu, et al in 2017: proposed in his work by the new encryption scheme as a lossless color image to means of DNA sequence operations, two simple improved chaotic systems, and the ciphertext diffusion technology in crisscross mode. Compared with the current DNA-based image encryption schemes, our proposed scheme involves not only DNA-level confusion and different fusion but pixel-level diffusion, which will enhance the security, complexity, and sensitivity of the cryptosystem. some biological and algebraic operations are constructed for the DNA sequences to promote the application in cryptography. Also, they use exclusive (XOR) operation for DNA sequences is adopted to encrypt and decrypt the digital images. The XOR operation for DNA sequences is performed according to the traditional binary XOR[14].

Monika Bartwal, and Dr. Rajendra Bharti in 2017, proposed a lossless method for sensitive data hiding and other method for image cryptography by ChoasBlock for encrypted image. Lossless means when the marked image is considered reliable, in this case the embedding distortion can be all of removed from the marked image afterward embedded method has been extracting. This steps uses features of the pixel difference to embed more data than other random partitions using blockbased a sharpness index filtering with refined of a one level wavelet decomposition shifting method to prevent image distortion problems[15].

Seo-Joon Lee, et al in 2018 proposed Block-chain used FASTA and FASTQ used a lossless compression (BAQALC) method, the lossless compression method that allows for the perfect storage with transmission of an immense amounts of sequence data of DNA that was being generate through NGS. Even, reliability issues and security exist in public DNA sequence data. The proposed solution is envisioned to contribute to providing an efficient and secure transmission and storage platform for next-generation medical informatics based on smart devices for both researchers and healthcare users [16].

Diogo Pratas, Morteza Hosseini, Jorge M. Silva, and Armando J. Pinho in 2019, proposed Jarvis algorithm as a new method that applied the competitive prediction depend on 2 various classes. Weighted stochastic repeat models (WSR models) with Weighted context models (WC models). WC models can be using a soft blending method, with a decaying forgetting factor of models for substitutional tolerant context. The WSR models also use the soft blending method, with a decaying forgetting factor, between multiple repeat models of specific word length. Both applied half-programs to handle inverted repeats. The competitive prediction is dependent on maximum probability on each one of class at a precise moment. The method is trained over a prediction by using a WC. A last probability, for every base is coded by an arithmetic encoder method[17].

Deloula Mansouri et al in 2020, proposed a description of a novel reference free sequence of DNA compressor call is DNACSBE. So, the hybrid lossless compressor within 3 stages is called DNAC-SBE. The other bases positions are having a frequency that are lower than Bi in the first replaced with zero's, from starting within the bigger base Bi and also positions of each one of a Bi be replaced with ones. Second, utilizing two separate methodologies, we suggest a novel single-block encoding strategy (SEB) to encode the generated streams. This SEB takes advantage of the nearby bits' positions within the block[18].

K. Punitha,Dr. A. Murugan in 2020, proposed the R-pattern algorithm using the dictionary-based compression of DNA sequence and attains a better compression ratio and is tested with the benchmarking datasets from NCBI(National Center for Biotechnology Information) [19].

Syed Mahamud Hossein in 2022, presents a work being carried out by minimize an executing-time with lossless compression rate as a DNA sequence length an increase of the large amount. Protection of DNA sequence of databases for hacker are the challenge question. This work method called lossless DNA sequence compression that was develop dependent of searching for an exact Palindrome with Repeat. Single hidden characteristics of a DNA

Doi : https://doi.org/10.54216/JCIM.140201
Received: January 17, 2024 Revised: March 08, 2024 Accepted: June 17, 2024

sequence is approximate repeats, the feature of Palindrome with Repeat was considered by his work. The proposed method applied for minimize capacity of storage with reduce a transmission cost. DNA sequence compression is optimize based on encoding exact palindromes with repeats in match positions. They should be not overlapping of a palindrome with repeat method for DNA sequence compression. By the proposed method, also compression 2 files are produced library and compressed files. Library file acts is the providers and signature security. Characters set of palindrome with repeat method also act as the provide strong data security with private key[20].

Sarah Elnady, et al in 2022, presents the new method for reference dependent of lossless compression, so they proposed DNA sequences stored in FASTA format that can be act in the layer above of gzip compression. several experiments are performed for evaluate his method and experimental results present that is be able for obtain promising ratios of compression to saving up of 99.9 % reaching and space of gain at 80% for some plant genomes. The proposed method in this work was succeeds a performing of compression at the acceptable time[21].

Yi Niu, Mingming Ma , Fu Li , Xianming Liu and Guangming Shi in 2022, propose a lossless quality score compression dependent of adaptive of coding order. The main idea of proposed work is traverse a quality score adaptively at a most correlative trajectory based on sequencing process. Based on cooperating and adaptive arithmetic coding with the improved of context strategy, adaptive of coding order was achieves state of the art quality score compression performances and moderate complexity for next generation sequencing data[22].

3. The Proposed System

The proposed system consists of two processes, the embedding and extraction. The embedding process includes five main stages: (Prepare Medical data patient information, doctors' diagnosis, healthcare service providers, drug manufacturers, etc.), Re-arrange data stage, encryption using non-linear function, DNA lossless compression. On the other hand, the extraction process, the recipient receives a file containing the locations array. Figure 1 illustrates the general block diagram of the proposed system.

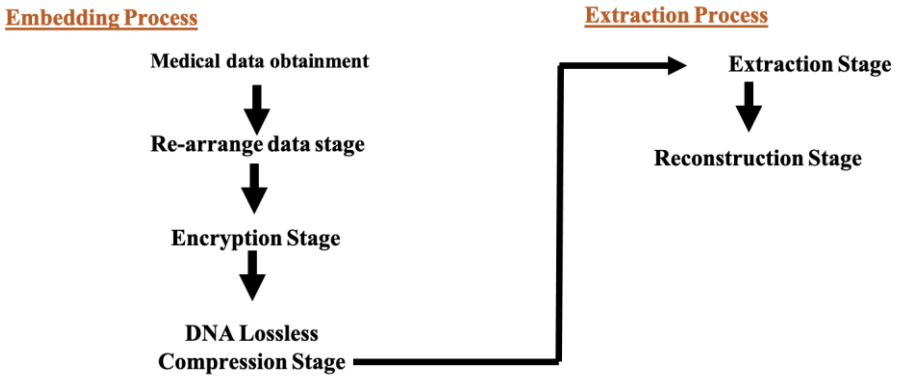

Figure 1: Proposed System Block Diagram

3.1 Embedding Process

In this process, the sender has to process several stages before sending the medical data to the receiver such as:

3.1.1 Medical Data Obtainment Stage

obtainment the data is the first step done within the proposed work. Medical Data could be a patient information, doctors' diagnosis, healthcare service providers, drug manufacturers, etc., to arrange these data in a form to be manipulated.

3.1.2 Re-arrange Data Stage

To complement the process is conducted for all characters involved in the integration process, through which all phases of the characters are reflected as a proactive step within the work stages. In this stage, focuses on arranging the bits of the medical report in the form of a 2D array in which all data are zeros and ones, arranged according to their sequence within the medical report, where it is prepared for the next steps.

Doi : https://doi.org/10.54216/JCIM.140201
Received: January 17, 2024 Revised: March 08, 2024 Accepted: June 17, 2024

3.1.3 Encryption Stage

This stage is considered a basic stage within the work during which a non-linear function is performed implicitly within the 2D-array. The nonlinear function used in our work can be considered as a basic function in the encryption stage, its operation is implicit without using an encryption key. An xor operation is performed between the rows of data in some manner takes place between the odd and even rows. The even rows are replaced by the xor value, leaving the odd rows at their initial values as shown in figure 2.

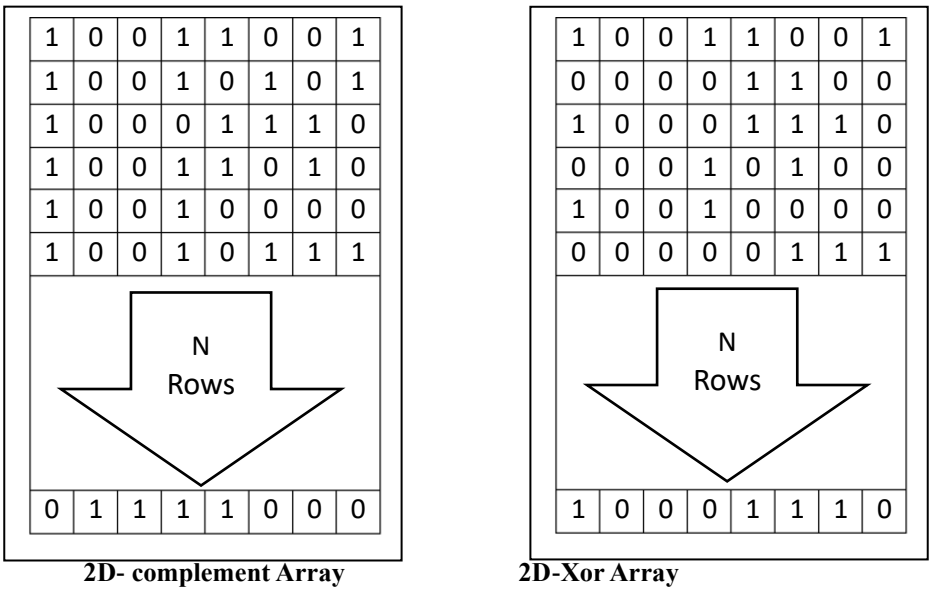

Figure 2: Encryption Process

3.1.4 DNA Coding Stage

Using DNA coding is also scheduling in this work, as the outputs of the non-linear array are encoded so that every two bits represent one letter of the DNA bases. Thus, each byte can be represented by four letters of the DNA bases that's illustrate in table 1. Scheduling DNA procedure apply additional security to the data medical data using lossless policy that ensures speed of implementation and difficulty breakable.

Table 1: DNA coding

bits	DNA letter
00	**A**
01	**C**
10	**T**
11	**G**

The DNA coding chain is rearranged in a way that ensures randomness without repetition, and this increases the strength and stability of the proposed coding method. Where a number is generated between (0 and 3), each number is represented by a letter from the DNA chain and is tested and added to the new chain if it was not mentioned previously, with the condition of testing when generating each time. This ensures randomness and non-repetition of letters.

3.1.5 DAN lossless compression stage

It is a procedure that used during work to reduce the amount of bloated data during coding operations in the DNA sequence. This procedure is based on taking the location of each letter in the random sequence against the letter in the DNA sequence.

The process of compressing data occurs at this stage, where of compressing the DNA chain takes place. Through our observation of the outputs of the previous procedures, we noticed that there was an increase in the amount of data, which necessitated adding compression to reduce the amount of data during sending or saving. We took each letter

Doi : https://doi.org/10.54216/JCIM.140201
Received: January 17, 2024 Revised: March 08, 2024 Accepted: June 17, 2024

from the DNA matrix by replacing the location of the corresponding letter in the random DNA matrix. Here, the position number is represented by the letter of the random array instead of the letter of the DNA array. The random DNA string is recoded into binary numbers instead of each letter being represented by two numbers. Here the data will be compressed without loss and with the compression of 75% percentage as shown in table 2.

Table 2: DNA lossless compression

DNA Coding	T	G	G	T	A	A	A	G	T	G	G	A	A	A	C	G	T	G	T	T
Random DNA	G	C	A	G	G	T	G	A	A	T	T	G	A	G	G	A	T	T	A	T
Array of location	5	0	3	9	2	7	8	4	10	6	11	12	15	18	1	13	16	14	17	19
Array of bits	**11010011**				**11101100**				**00101011**				**00111100**				**10100010**			
Compression array	**211**				**236**				**43**				**60**				**162**			
Lossless Compression	**1 letter**				**1 letter**				**1 letter**				**1 letter**				**1 letter**			

3.1.6 Random DAN Sequence

It is a procedure that we used in the proposed system; to generates a random DNA sequence with the same length as the original DNA sequence. We use it in the process of securing important medical data that needs to be transferred or saved.

3.2 Extraction Process

The recipient on the other hand receives an array of locations that he needs in the decompression process in a way that allows him not to lose the sent data. The recipient decompresses and encrypts using the steps used by the sender in reverse sequence.

3.2.1 Reconstruction Stage

This is the final stage of the work and takes place entirely at the second party, the recipient side, will calculate the length of the location array as a first step to calculating the length of the message or medical report. calculating the locations of the DNA sequence and encoding according to the array of locations, and then the steps are reversed in order to decompress and recode the DNA sequence into a series of bits. The string of bits is divided into a two-dimensional array, and the non-linear function is called between the columns and rows, and the values of the even columns are restored to their value before the non-linear function, and then it goes through the complement stage in order to return the bit values to what they were before encryption, and then the conversion function from binary to ASCII to retrieve the medical report data , so that the final medical report appears in its final form to the receiver site as in figure 3.

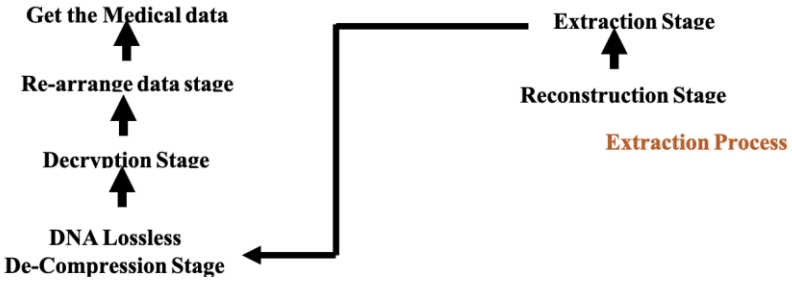

Figure 3: Extraction Process block diagram

4. Experimental Results

This section describes performance of the proposed system through practical implementation of the proposed system and their obtained results from the experiments used to assess the methods performance System. The proposed techniques have been implemented using several metrics measurement.

the system was tested on medical reports of various sizes. The proposed system demonstrated a compression rate of 75% for the data. The well-known metrics were also calculated the efficiency of the proposed system, and they all

Doi : https://doi.org/10.54216/JCIM.140201
Received: January 17, 2024 Revised: March 08, 2024 Accepted: June 17, 2024

demonstrated the strength and merit of the encryption method, the DNA chain coding method, and the DNA compression method becomes an integrated system in terms of encryption, coding, and compression to keep pace with modernity and continuous development in the field of information security and preservation.

4.1 DNA lossless compression

Eight real genome sequences were taken from the National Center for Biotechnology Information (NCBI) databases for eight known viruses (Influenza A Virus, Enterovirus, Papillomaviruses, Hepatitis C Virus, Flavi virus, herpes viruses, Corona Virus, Ortho poxvirus) to test the efficiency of the compression strategy used in our proposed system. Different sizes of virus files were downloaded, and each file was identified with a reference number according to the National Center for Biotechnology Information encoding. Also, during the work test, the size of each file was considered before compression and after compression, and the percentage of difference between the two sizes was calculated and the percentage of compression was calculated, as shown in the following table, with a difference between the percentages as a result. The length of each report string varies, as well as the number of ASCII codes that represent white spaces and empty values, as they all occupy space and size within each file as shown in table 3,4,5,6, and figures 4, 5, 6, 7

Table 3: Percentage of Eight types of Viruses

NCBI ID	Virus name	Original size	Compressed DNA size	Compression percentage	Compression diff percentage
NC_002023.1	Influenza A Virus	2407 Byte	892 Byte	37.0585791%	91.84601394%
MW373957.1	Enterovirus	3950 Byte	1420 Byte	35.9493671%	94.22718808%
OL704825.1	Papillomaviruses	9177 Byte	3373 Byte	36.7549308%	92.4940239%
NC_004102.1	Hepatitis C Virus	9920 Byte	3655 Byte	36.8447581%	92.30202578%
NC_021069.1	Flavi virus	11175 Byte	4171 Byte	37.3243848%	91.2811156%
NC_028891.1	Herpes viruses	23601 Byte	8599 Byte	36.4348968%	93.18012422%
NC_045512.2	Corona Virus	30757 Byte	11467 Byte	37.2825698%	91.36983706%
NC_048657.1	Ortho poxvirus	44758 Byte	16412 Byte	36.6683051%	92.67941802%

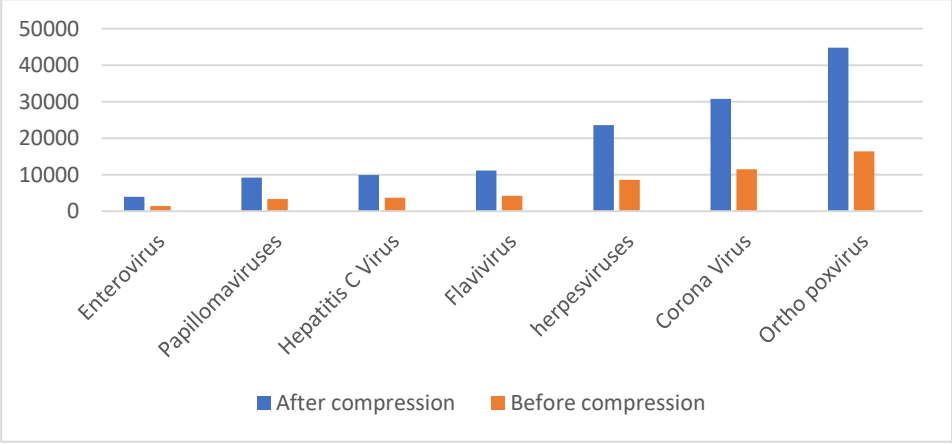

Figure 4: DNA virus file size before / after compression

The figure shows different sizes for eight different files. Each file represents DNA sequence of a real virus whose data was taken from the NCBI website, and each virus has a different file size measured in bytes.
the compression ratio is defined in terms of uncompressed and compressed data rates instead of data sizes[23]:

Doi : https://doi.org/10.54216/JCIM.140201
Received: January 17, 2024 Revised: March 08, 2024 Accepted: June 17, 2024

Table 4: Compression Percentage of Eight types of Viruses

Virus name	Original size	Compressed DNA size	Compression percentage
Influenza A Virus	2407 Byte	892 Byte	37.0585791%
Enterovirus	3950 Byte	1420 Byte	35.9493671%
Papillomaviruses	9177 Byte	3373 Byte	36.7549308%
Hepatitis C Virus	9920 Byte	3655 Byte	36.8447581%
Flavi virus	11175 Byte	4171 Byte	37.3243848%
Herpes viruses	23601 Byte	8599 Byte	36.4348968%
Corona Virus	30757 Byte	11467 Byte	37.2825698%
Ortho poxvirus	44758 Byte	16412 Byte	36.6683051%

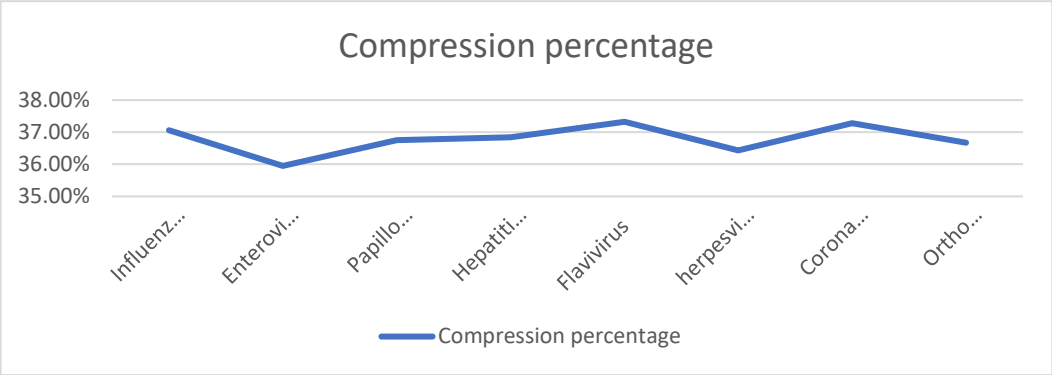

Figure 5: compression percentage of eight DNA viruses before / after compression

From the figure above, we can clarify that the compression ratios for the DNA files of the eight selected viruses range from 35% to 37% for different sizes represented in bytes, and this indicates the stability of the data compression process used in the work.

Compression Factor (CF) It is the inverse of the compression ratio, which is the ratio [24] between the size of the original virus file and the size of the compressed virus file.

Compression factor = $\frac{size\ of\ original\ text\ file}{size\ of\ compression\ text\ file}$ ……(2)

Table 5: Compression Factor of Eight types of Viruses

Virus name	Original size	Compressed DNA size	Compression Factor
Influenza A Virus	2407 Byte	892 Byte	2.69843049%
Enterovirus	3950 Byte	1420 Byte	2.78169014%
Papillomaviruses	9177 Byte	3373 Byte	2.72072339%
Hepatitis C Virus	9920 Byte	3655 Byte	2.71409029%
Flavi virus	11175 Byte	4171 Byte	2.67921362%
Herpes viruses	23601 Byte	8599 Byte	2.74462147%
Corona Virus	30757 Byte	11467 Byte	2.68221854%
Ortho poxvirus	44758 Byte	16412 Byte	2.72715087%

Doi : https://doi.org/10.54216/JCIM.140201
Received: January 17, 2024 Revised: March 08, 2024 Accepted: June 17, 2024

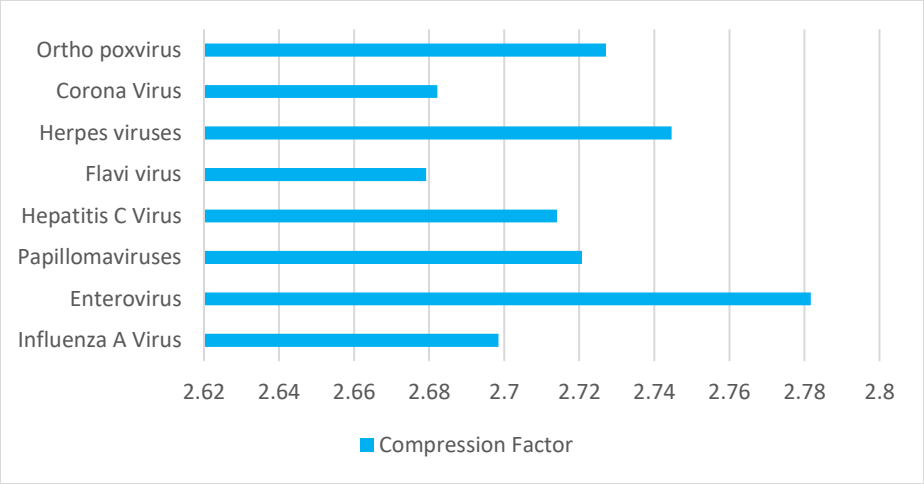

Figure 6: Compression Factor of eight DNA viruses before / after compression

Different percentage is defined in terms of different between uncompressed and compressed data rates instead of data sizes [25]:

$$\text{Different percentage} = \frac{|a-b|}{(a+b) \div 2} \times 100 \quad \text{.....(3)}$$

Let a is original file and b is compression file

Table 6: Compression Different Percentage of Eight types of Viruses

Virus name	Original size	Compressed DNA size	Compression diff percentage
Influenza A Virus	2407 Byte	892 Byte	91.84601394%
Enterovirus	3950 Byte	1420 Byte	94.22718808%
Papillomaviruses	9177 Byte	3373 Byte	92.4940239%
Hepatitis C Virus	9920 Byte	3655 Byte	92.30202578%
Flavi virus	11175 Byte	4171 Byte	91.2811156%
Herpes viruses	23601 Byte	8599 Byte	93.18012422%
Corona Virus	30757 Byte	11467 Byte	91.36983706%
Ortho poxvirus	44758 Byte	16412 Byte	92.67941802%

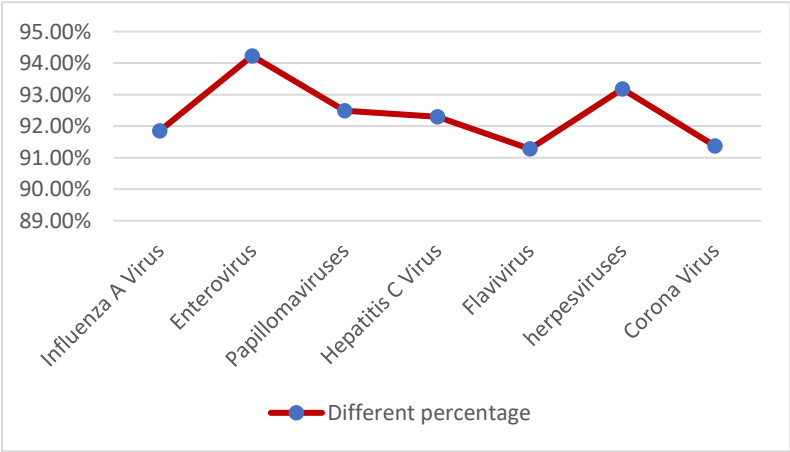

Figure 7: Different percentage of eight DNA viruses before / after compression

Doi : https://doi.org/10.54216/JCIM.140201
Received: January 17, 2024 Revised: March 08, 2024 Accepted: June 17, 2024

Accuracy and error rate are easy to calculate and interpret [26], but they may not capture the nuances and challenges of different file processing tasks [27]. For example, accuracy may not reflect the severity of different types of errors, or the variability of file quality and complexity.
Word error rate can then be computed as:

$$WER = \frac{S+D+I}{N} = \frac{S+D+I}{S+D+C} \quad \text{……..(4)}$$

Were S is the number of substitutions, D is the number of deletions, I is the number of insertions, C is the number of correct words, N is the number of words in the reference (N=S+D+C)
The ER was calculated for a file containing a medical report that was encrypted, compressed in five types of selected encryptions. The results showed that the error rate was 0.0 for all types of files that we used in our work were identical, and there was no error happen in our work.

$$\text{WER} = \frac{S+D+I}{N} \times 100$$
$$\text{Accuracy} = 100 - \text{WER} \quad \text{……..(5)}$$

Table 7: WER and Accuracy of Eight types of Viruses

Virus name	Original size	Compressed DNA size	WER	Accuracy
Influenza A Virus	2407 Byte	892 Byte	0.0	100.0
Enterovirus	3950 Byte	1420 Byte	0.0	100.0
Papillomaviruses	9177 Byte	3373 Byte	0.0	100.0
Hepatitis C Virus	9920 Byte	3655 Byte	0.0	100.0
Flavi virus	11175 Byte	4171 Byte	0.0	100.0
Herpes viruses	23601 Byte	8599 Byte	0.0	100.0
Corona Virus	30757 Byte	11467 Byte	0.0	100.0
Ortho poxvirus	44758 Byte	16412 Byte	0.0	100.0

By following up the compression results that obtained from implementing proposed work, indicate that the work achieves better results in compression processes, as the percentages range to 75%. for character compression, and different rate ranges between 91% to 94%, and the compression rate ranges from 35% to 37%, compression factor between 2.6% to 2.7%. These are all excellent percentages compared to retrieving data with an accuracy rate of up to 100% and without data loss.

5. Conclusion

For this paper, a proposed new strategy to safeguard the transfer of sensitive medical data over an unsecured network using cryptography, and DNA lossless compression. The proposed system was tested according to existing measures was achieved a stable Compression Ratio, Compression Factor, Error Rate, Accuracy and other measures. The proposed system is considered a safe system as it does not allow changes in the data of the original medical report or the DNA sequence, the compression process make it smaller and unreadable.

References

[1] Priyanka, & Singh, A. K. (2023). A survey of image encryption for healthcare applications. *Evolutionary Intelligence*, *16*(3), 801-818.
[2] Almeida, B. D. A., Doneda, D., Ichihara, M. Y., Barral-Netto, M., Matta, G. C., Rabello, E. T., ... & Barreto, M. (2020). Personal data usage and privacy considerations in the COVID-19 global pandemic. *Ciencia & saude coletiva*, *25*, 2487-2492.
[3] Noor, N. S., Hammood, D. A., Al-Naji, A., & Chahl, J. (2022). A fast text-to-image encryption-decryption algorithm for secure network communication. *Computers*, *11*(3), 39.
[4] Naji, M. A., Atee, H. A., Jebur, R. S., Hammood, D. A., Der, C. S., Abosinnee, A. S., ... & Ahmad, R. B. (2021, September). Breaking A Playfair Cipher Using Single and Multipoints Crossover Based on Heuristic Algorithms. In *2021 4th International Iraqi Conference on Engineering Technology and Their Applications (IICETA)* (pp. 47-53). IEEE.

Doi : https://doi.org/10.54216/JCIM.140201
Received: January 17, 2024 Revised: March 08, 2024 Accepted: June 17, 2024

[5] Dagadu, J. C., Li, J. P., & Aboagye, E. O. (2019). Medical image encryption based on hybrid chaotic DNA diffusion. *Wireless Personal Communications*, *108*, 591-612.
[6] Dey, S., & Ghosh, R. (2018). A review of cryptographic properties of S-boxes with Generation and Analysis of crypto secure S-boxes. *Cryptology ePrint Archive*.
[7] Kester, Q. A., Nana, L., Pascu, A. C., Gire, S., Eghan, J. M., & Quaynor, N. N. (2015). A cryptographic technique for security of medical images in health information systems. Procedia Computer Science, 58, 538-543.
[8] Nayak, Padmalaya, and G. Swapna. "Security issues in IoT applications using certificateless aggregate signcryption schemes: An overview." Internet of Things 21 (2023): 100641.
[9] Shawkat, Shihab A., Najiba Tagougui, and Monji Kherallah. "Information Security: A Review on Steganography with Cryptography for Genetic Algorithm (GA) Transaction." In 2023 7th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), pp. 1-7. IEEE, 2023.
[10] Zhang, Bowen, and Lingfeng Liu. "Chaos-based image encryption: Review, application, and challenges." Mathematics 11, no. 11 (2023): 2585.
[11] Nassra, Ihab, and Juan V. Capella. "Data compression techniques in IoT-enabled wireless body sensor networks: A systematic literature review and research trends for QoS improvement." Internet of Things (2023): 100806.
[12] Wang, Jiaqi, and Bo Ou. "Video reversible data hiding: A systematic review." Journal of Visual Communication and Image Representation (2023): 104029.
[13] Ghiasi, Mohammad, Taher Niknam, Zhanle Wang, Mehran Mehrandezh, Moslem Dehghani, and Noradin Ghadimi. "A comprehensive review of cyber-attacks and defense mechanisms for improving security in smart grid energy systems: Past, present and future." Electric Power Systems Research 215 (2023): 108975.
[14] Wu, X., Wang, K., Wang, X., & Kan, H. (2017). Lossless chaotic color image cryptosystem based on DNA encryption and entropy. Nonlinear Dynamics, 90, 855-875.
[15] Bartwal, M., & Bharti, R. (2017). Lossless and Reversible Data Hiding in Encrypted Images With Public Key Cryptography. In Rice (pp. 127-134).
[16] Lee, S. J., Cho, G. Y., Ikeno, F., & Lee, T. R. (2018). BAQALC: Blockchain applied lossless efficient transmission of DNA sequencing data for next-generation medical informatics. applied sciences, 8(9), 1471.
[17] Pratas, D., Hosseini, M., Silva, J. M., & Pinho, A. J. (2019). A reference-free lossless compression algorithm for DNA sequences using a competitive prediction of two classes of weighted models. Entropy, 21(11), 1074.
[18] Mansouri, D., Yuan, X., & Saidani, A. (2020). A new lossless DNA compression algorithm based on a single-block encoding scheme. Algorithms, 13(4), 99.
[19] Partee, J., Hazell, R., Solsi, A., & Santerre, J. (2020). Compressed DNA Representation for Efficient AMR Classification. SMU Data Science Review, 3(2), 5.
[20] Hossein, S. M. (2022). Substitution based DNA Sequences Compression-Encryption Method.
[21] Elnady, S., Sayed, S., & Salah, A. (2022). HADC: A Hybrid Compression Approach for DNA Sequences. IEEE Access, 10, 106841-106848.
[22] Niu, Y., Ma, M., Li, F., Liu, X., & Shi, G. (2022). ACO: lossless quality score compression based on adaptive coding order. BMC bioinformatics, 23(1), 219.
[23] Varshney, Ankit, K. Suneetha, and Dhananjay Kumar Yadav. "Analyzing the Performance of Different Compactor Techniques in Data Compression & Source Coding." In 2024 International Conference on Optimization Computing and Wireless Communication (ICOCWC), pp. 1-6. IEEE, 2024.
[24] Sebai, Dorsaf, Manel Zouaoui, and Faouzi Ghorbel. "Seismic data compression: an overview." Multimedia Systems 30, no. 1 (2024): 38.
[25] Tang, Yehui, Yunhe Wang, Jianyuan Guo, Zhijun Tu, Kai Han, Hailin Hu, and Dacheng Tao. "A Survey on Transformer Compression." arXiv preprint arXiv:2402.05964 (2024).
[26] Horsman, Graeme. "Sources of error in digital forensics." Forensic Science International: Digital Investigation 48 (2024): 301693.
[27] Sheikh, Md Rasel, and Paulin Coulibaly. "Review of Recent Developments in Hydrologic Forecast Merging Techniques." Water 16, no. 2 (2024): 301.

Doi : https://doi.org/10.54216/JCIM.140201
Received: January 17, 2024 Revised: March 08, 2024 Accepted: June 17, 2024