e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

DNA: THE FUTURE DATASTORE AN INTERDISCIPLINARY APPROACH FOR DATA RETRIEVAL AND SECURITY

Akanksha Jojy Jacob*1

*1Department Of Information Technology, B. K. Birla College, Mumbai, Maharashtra, India.

DOI : https://www.doi.org/10.56726/IRJMETS45246

ABSTRACT

In this era of increasing data, data storage is posing a great threat to all the domains. Effective retrieval and security is as important so as to store the data. This research presents a multidisciplinary approach combining the branch of genetics (DNA encoding), Cryptography (Hyperchaotic Encryption), Noise resilience and Wavelet based denoising techniques. Initially a test image is converted into DNA encoded data, then a Hyperchaotic Encryption algorithm is applied to defend the problems related to security. A noise attack is being introduced to demonstrate the noise which is normally present during the data retrieval stage. To overcome the problems due to noise during the retrieval, Wavelet based denoising technique and TV (Total Variation) denoising technique is used to remove the added noise. On applying Wavelet based denoising technique, the percentage match increased to high note. PSNR (Peak Signal to Noise Ratio) values, SSI (Structural Similarity Index) scores , percentage matches are also presented to check the relevance of the experiments conducted. The above experiments show that data can be stored in DNA in future and effectively retrieved irrespective of various security issues

Keywords: DNA, Data Storage, Hyperchaotic Encryption, Wavelet Denoising, Data Security, Total Variation (TV) Denoising.

I. INTRODUCTION

With the advancement in technology and multimedia, in every domain the amount of data is generated at very high volume. We need to look out for different mechanisms for storing this data similarly, the security of data has been a most integral factor. As far as image is concerned which are widely used in various sectors like health, tourism, military, social the security of data is considered of prime importance. Unlike text, digital images carry certain inherent characteristics like large data storage capacities, high resolution etc. There are several encryption algorithms designed which ensures the security of the data. Nowadays attempts have also been made to store this digital data inside DNA molecules which provides with high storage mechanism and ensures durability of the data. Since our molecular DNA has several distinguishing features like high storage density, high massive and ultra-low energy consumption, the DNA-based image encryption has attracted more researchers' attention. Current storage technologies, including optical and magnetic devices, are reaching their limits for storing this humungous data. One of the approach is to store digital data inside DNA molecules. According to statistics, 215 petabytes of data can be stored in a single gram of DNA. The two major limitations which comes into picture in this type of storage is the cost issues and the retrieval issues. The first issue can be reduced by generating the copies of DNA by cloning, etc. The second issue states that the image stored may not be retrieved in the same manner as it is stored, either it may contain some noise or any distortion would occur to the original data stored, even the security of our data is at risk while storing the data inside DNA molecules. This research paper addresses all these problems by combining cryptography, genetics and denoising techniques.

Nowadays attempts have also been made to store this digital data inside DNA molecules which provides with high storage mechanism and ensures durability of the data[3]. It is estimated that by 2025 the global data storage may rise up to 1.75 $*10^{14}$ whereas the current density is 103 GB/ mm3 [2]. Because of the DNA computing excellent properties, such as high storage density, high massive and ultra-low energy consumption, the DNA-based image encryption has attracted more researchers' attention [4]. Current storage technologies, including optical and magnetic devices, are reaching their information density limits and are thus not suitable for long term (>50 years) storage, which means that valuable information needs to regularly be transferred to newer storage media if it is to be preserved for future generations [2]. In [11] there is a discussion of chaos

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

based encryption and decryption algorithm which includes DNA encoding and generation of the encrypted image. Similarly in [13] Hyperchaotic encryption algorithm based on bit level permutation and DNA encoding is discussed. A noise attack is also being performed which is analyzed by the researchers.

Molecular DNA referred to as Deoxyribonucleic acid is considered as a storage molecule. It contains four nitrogenous bases : Adenine (A) , Guanine (G) , Cytosine (C), Thymine (T). These sequences of nitrogenous bases carries genetic information. Important feature of DNA to replicate allows to generate the own copies. DNA sequencing is important used in genetics and genomics.

Cryptography is a branch of science which converts the readable data into unreadable format generally known as the 'cipher text' using various mathematical algorithms and generating random keys. One of the type of cryptography is 'Symmetric Cryptography' wherein the same key is used for encryption and decryption. Hyperchaotic encryption algorithm is a kind of symmetric cryptographic technique where the same key is used for encryption as well as for decryption. This approach differs from the normal hyperchaotic cryptography since it used DNA encoding as well as bit level permutation. Bit level permutation is to arrange individual bits of the existing binary data to produce new binary sequence.

Images are highly prone to noise attacks during the retrieval of the data. Noise attack is simply any distortion or variations introduced into the image. There are various types of noise attacks like Gaussian, Salt and Pepper, Speckle. During the course of the research Gaussian Noise attack is demonstrated which occurs due to sensors, capturing image via camera, etc.

To remove the induced noise various denoising techniques can be used. One such technique is Wavelet based denoising, which uses wavelet transforms to disintegrate the image into various scales and levels of details and then removes the noise from these scales. This technique is widely used for multiscale analysis of the image which preserves it fine and minute details. Another important advantage of this technique is that , It preserves the edges of the image during the process which maintains the originality of the image to the mark. On the other hand TV (Total Variation) denoising technique works on the principle pf minimizing the total variation of an image. This technique has a major disadvantage of removing the finer details.

By combining all of these above mentioned processes an attempt has been made to store the image into DNA molecule digitally and applying hyperchaotic encryption algorithm. This research proves that data can be stored inside DNA molecules and effective techniques when applied will not pose much challenges for data retrieval and security.

II. METHODOLOGY

Data Preparation

The image used for the experiment is the Lena image. This image is considered as a standard test image because of the details present in the image, its shading and texture which makes it easier for image processing.

- Image Processing: The image is converted to gray scale at the initial phase and converted into 8 bit binary data and represented as a single string.
- DNA encoding: The given binary data is converted into DNA encoded data using the encoding rules.
- Performing bit level permutation: Tent map is used for performing bit level permutation. A tent map is defined as a mathematical function which is widely used in chaos theory. The tent map variable is initialized as 0.1.

Introducing Noise

Gaussian Noise Attack: Gaussian noise is a kind of statistical noise which follows normal distribution. For images that are hand captured. It is generated with a defined mean and standard deviation. Noisy encrypted image is generated by adding noise to the encrypted image.

Evaluation

- Structural Similarity Index (SSI) score: Structural Similarity Index (SSI) predicts the match of the original image to the noisy image. In other words, it shows at how much percent the noise has distorted the original image.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

- Peak Signal-to-Noise Ratio (PSNR) and Mean Squared Error (MSE): PSNR is calculated to to check the overall quality of the image whereas MSE value displays the error between the images.
- Percentage Match: Percentage match is used to find the match between the encrypted image and the noisy encrypted image after Gaussian noise attack.
- PM = SSI * 100

Denoising Technique

Total Variation Denoising: TV denoising is performed on the noisy encrypted image with a specified parameter. The PSNR and MSE value is being calculated for the denoised image generated using this process.

Wavelet based Denoising: The noisy encrypted image is then subjected to wavelet based denoising using a wavelet and threshold method. The denoised image is again reconstructed using this technique.

Figure 1: Original Lenna Image

III. RESULTS AND DISCUSSION

This section discusses various experimental conclusions to interpret the findings of the research. The main aim of the experiment was to prove the effectiveness of data storage within molecular DNA, Encryption done using Hyperchaotic algorithm which is known for its robustness and the final improvement achieved through wavelet based and TV based denoising techniques.

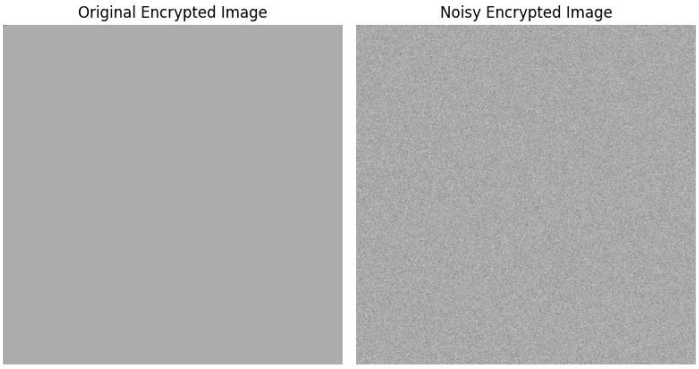

Figure 2: Encrypted Image

T.V. based denoising technique showed the following PSNR and MSE values when comparing with original to noisy and denoised image as shown below:

Table 1. T.V. Denoising

Evaluation metrics	Original Image vs Denoised Image
PSNR	3.45 dB
MSE	29354.23
Percentage Match	54.02

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

These values of PSNR and MSE indicates that noisy image contains noise .PSNR value of 3.45 dB and percentage match of 54.02 suggests that noise is still present and there can be chances of improvement.

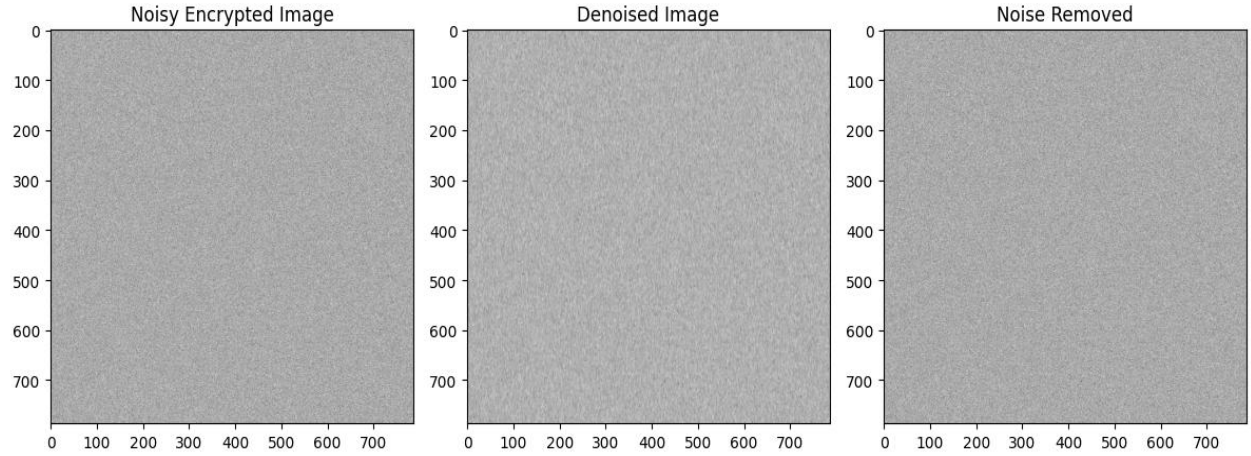

Figure 3: T.V. Denoising

Wavelet based denoising technique showed the following PSNR and MSE values when comparing with original to noisy and denoised image as shown below:

Table 2. Wavelet Denoising

Evaluation metrics	Original Image vs Denoised Image
PSNR	19.71 dB
MSE	695.25
Percentage Match	99.9

These values of PSNR and MSE indicates that noisy image contains noise . The high value of PSNR indicates that image has minimal noise and high similarity with original one. The percentage match of 99.9 suggests that the image is almost identical to the original image.

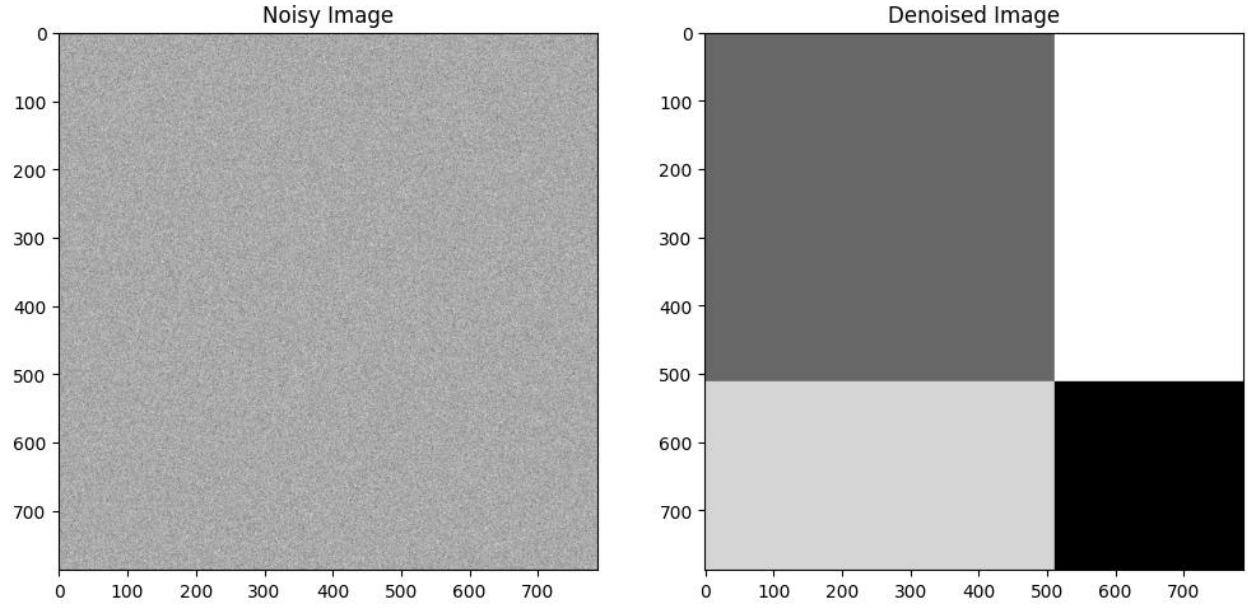

Figure 4: Wavelet based Denoising

The above results suggests that the denoising techniques applied has effectively removed noise without degrading the quality of the image which makes it suitable for various applications.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

IV. CONCLUSION

The overall research was an attempt to store an image into molecular DNA. Similar to this, various other forms of data can also be stored and easily retrieved. Hyperchaotic encryption algorithm proved to be one of the finest algorithms due to its speed, efficiency and also its robustness. The gaussian noise attack injected ample amount of noise to the image which proved to be a real time noise. Two types of denoising techniques were compared wherein Wavelet based denoising outstood the TV based on preserving the quality, sharpness and also removing the noise to a larger extent. All these results give a positive way for the digital data storage inside molecular DNA. The algorithms used also suggest the positive outcomes of this research conducted. The data stored inside molecular DNA would be safe and secure. If prone to any attack it can recover the document with high percentage of similarity to the original data. This research contributes to the growing field biological data storage and cryptography and also for future innovations and inventions to formulate new methods for storing data produced at a humungous rate.

V. REFERENCES

[1] Chao Pan S. Kasra Tabatabaei , S. M. Hossein Tabatabaei Yazdi , Alvaro G. Hernandez, Charles M. Schroeder Olgica Milenkovic , Rewritable 2D DNA based data storage with machine learning reconstruction, 2022

[2] Andrea Doricchi, Casey M Platnich, Andreas Gimpel and Denis Garoli ,A Emerging Approaches to DNA Data Storage: Challenges and Prospects, 2022.

[3] Ilan Shomorony, "DNA Based storage : Models and Fundamental Limits, 2020.

[4] Jingshuai Wang 1,2 , Fei Long 1, Weihua Ou3, 'CNN-based Color Image Encryption Algorithm using DNA Sequence Operations,2017.

[5] Chao Pan1, S. M. Hossein Tabatabaei Yazdi , S Kasra Tabatabaei1 Alvaro G. Hernandez , Charles Schroeder1 Olgica Milenkovic, "Image processing in DNA, 2022.

[6] Bingzhe Li , Nae Young Song† , Li Ou , and David H.C. Du, "Can we store the whole worlds Data in DNA storage ?,2020.

[7] Chisom Ezekannagha a,b,* , Marius Welzel a,b , Dominik Heider a,b , Georges Hattab a,b,DNA smart: Multiple attribute ranking tool for DNA data storage systems,2023.

[8] Ala Saleh Alluhaidan,DNA Sequence Analysis for Brain Disorder Using Deep Learning and Secure Storage, 2022.

[9] Hemalatha Gunasekaran, K. Ramalakshmi, A. Rex Macedo Arokiaraj1 S. Deepa Kanmani,3 handran Venkatesan , C. Suresh Gnana Dhas ,Analysis of DNA Sequence Classification Using CNN and Hybrid Models,2021.

[10] Yiming Dong, Fajia Sun, Zhi Ping , Qi Ouyng Long Qian ,DNA storage: research landscape and future prospect,2020.

[11] Yaghoub Pourasad, Ramin Ranjbarzadeh, Abbas Mardani,A New Algorithm for Digital Image Encryption Based on Chaos Theory,2021 [12] Ravinder Rao Peechar Sucharita V,A chaos theory inspired, asynchronous two-way encryption mechanism for cloud computing,2021

[12] Ravinder Rao Peechar Sucharita V,A chaos theory inspired, asynchronous two-way encryption mechanism for cloud computing,2021.

[13] Tao Wanga Ming-hui Wanga, Hyperchaotic image encryption algorithm based on bit-level permutation and DNA encoding, 2020.