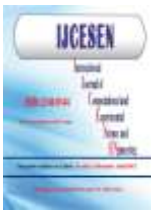
International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 10-No.4 (2024) pp. 1258-1270
http://www.ijcesen.com



ISSN: 2149-9144

Research Article

Efficient DNA Cryptography Using One-Time Pad and Run-Length Encoding for Optimized Ciphertext Storage

A. Rajeshkhanna[1], S. Kiran[2], A. Ranichitra[3*], S. Hemasri[4]

[1]Assistant Professor, Dept. of Computer Science, Sri. S.R.N.M College Sattur, TN, India,
Email: rajeshkanna@srnmcollege.ac.in- **ORCID:** 0000-0001-5161-4334

[2]Assistant Professor, Dept. of CSE, YSR Engg. College of YVU Proddatur, AP, India.
Email: rkirans125@gmail.com- **ORCID:** 0000-0002-0725-3356

[3]Assistant Professor, Dept. of Computer Science, Sri. S.R.N.M College Sattur, TN, India
* **Corresponding Author Email:** ranichitra@srnmcollege.ac.in- **ORCID:** 0000-0001-6071-0635

[4]Research Scholar, Dept. of Computer Science, Sri. S.R.N.M College Sattur, TN, India.
Email: hema0129@gmail.com- **ORCID:** 0009-000106103-0276

Article Info:

DOI: 10.22399/ijcesen.641
Received : 17 November 2024
Accepted : 20 November 2024

Keywords:

ASCII,
Run Length Encode,
Cryptography,
DNA OTP,
Encryption.

Abstract:

Cryptography ensures data privacy by transforming data into unreadable formats that only authorized individuals can decrypt. With the increase in electronically stored and transmitted data, enhanced methods for data protection are required. DNA cryptography, leveraging the genetic structure of DNA, provides a promising approach for secure communication and data storage. This paper introduces a novel DNA-based cryptographic method employing a DNA one-time pad (OTP) combined with modified run-length encoding to reduce ciphertext size. Unlike traditional cryptography, which often results in a larger ciphertext than plaintext, our proposed method demonstrates a significant reduction in ciphertext size. Experimental results reveal that for input text files of 1MB, 2MB, 3MB, 5MB, and 10MB, the ciphertext sizes were reduced by up to 20KB, 40KB, 60KB, 100KB, and 200KB, respectively. This reduction not only enhances storage efficiency but also minimizes transmission costs, marking a substantial advancement over existing DNA and classical cryptography methods. Future work will explore the application of this technique for encrypting biological data and incorporating DNA barcoding for improved data authentication and reliability.

1. Introduction

Cryptography is required in any condition that requires privacy or secrecy to protect the data and trade secrets. Cryptography preserves data by converting them into an unreadable format. Only authorized persons with a secret key can decipher the ciphertext. In some cases, encrypted messages are broken by cryptanalysis, which is also known as code breaking. Several algorithms, techniques, and methods have been proposed to provide security for the information to be transmitted. DNA cryptography has emerged as an advancement to all techniques [1]. In 1953, DNA was formed as a double helix of deoxyribonucleic acid, which consists of the genetic information and functions of all living organisms and viruses. A series of nucleotides (guanine, adenine, thymine, and cytosine) are encoded to form the genetic information noted with the help of the letters G, A, T, and C. The nucleotide pairs guanine-cytosine and adenine thymine, which are connected to a sugar and phosphate molecule, permit the DNA helix to maintain a regular helical structure, which is independent of its series. The identification of specific human beings is useful. The nucleotides are spaced every 03 nanometres giving a remarkable data density of 18Mbits per inch. One gram of DNA consists of approximately 10" DNA bases or 10 terra bytes of information. To store all the information in the world, a few grams of DNA is sufficient. Research has been conducted on implementing DNA using steganography, which preserves sensitive information such as biological data [2]. The aggregate of electronically stored and transmitted data is increasing daily. Therefore, enhanced

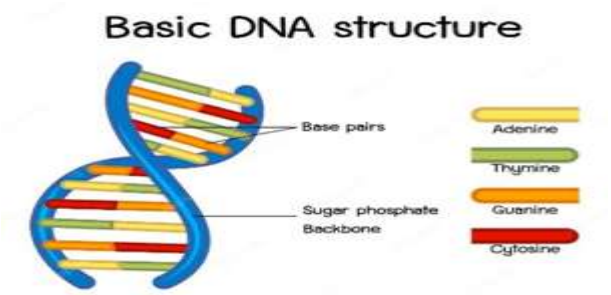


Figure 1. *Structure of DNA*

protection and control of the transmitted data are needed to protest various threats. Therefore, as an advancement to cryptography, both DNA computing and quantum computing have come into existence. Figure 1 is the structure of DNA. In DNA cryptography, a DNA molecule stores and transmits data in the form of DNA sequences. In contrast, quantum cryptography encryption techniques are based on quantum mechanics. In quantum cryptography, information is transmitted using quantum bits, also called qubits, which are single-photon particles. The DNA cryptography sector is a recent development that investigates the possibility of using DNA sequences for secure communication and information storage. DNA's unique features—which include its capability for error correction, dense facts storage, and large-scale records encoding in a small format, are the use forces at the back of DNA cryptography. A key thing of DNA cryptography is lossless compression, which makes it feasible to transmit and store genetic facts efficiently at the same time as keeping its authentic content. Some of the problems that need to be resolved earlier than in DNA cryptography depend on lossless compression and may be positioned in practice. The use of DNA sequences for each encoding and decoding statistic ensures that the compression method is lossless; that is, that no statistics are misplaced all through the compression and decompression processes is the basis for implementing DNA cryptography based on lossless compression [3]. The use of DNA cryptography based on lossless compression presents the following research challenges.

- Value and Scalability: Explore the price of DNA synthesis and sequencing at this scale. Recollect the scalability of the system for managing massive volumes of facts.
- Storage and retrieval: Manages the challenges related to garage and retrieval of DNA information. Explore green storage mechanisms and retrieval techniques for encoded DNA sequences.
- Bioinformatics Integration: Integrate bioinformatics equipment and techniques to enhance the accuracy of DNA series manipulation and analysis.

The main objective of the proposed work is

- To reduce the amount of space or storage required for ciphertext in the DNA molecule.
- To reduce the storage of cipher text, plain text is encrypted with the DNA one-time pad (OTP) method, and a lossless compression technique has been proposed.

The rest of the paper is organized as follows: In section2, related work is discussed. Section3, demonstrates previous work on cryptography. Section4, describes the proposed DNA OTP method and the implementation of the run length code. Section5, illustrates the analysis of the results. Section6, Describes the conclusions.

2. Literature Survey

Aieh, A., Sen, A., Dash, S. R., & Dehuri, S. et al.[4] proposes a new method of encryption using DNA as the key to demonstrates a shared secret key for communication. The authors suggest using the Diffie–Hellman key exchange algorithm with DNA sequences as the shared secret key. They explain the theoretical background of DNA cryptography and the working principle of the Diffie-Hellman key exchange. The authors also describe the proposed method in detail and present the results of their experiments, which demonstrate that their method is effective in terms of security and computational complexity.

Biswas, M. R., Alam, K. M. R., Tamura, S., & Morimoto, Y. et al.[5] a new method for DNA-based cryptography. The proposed method uses dynamic mechanisms, in which the encryption and decryption keys are generated dynamically using a combination of user-defined inputs and a randomly generated sequence. The results show that the proposed method provides the maximum degree of security against various types of threats, including brute-force and dictionary attacks. Dynamic key generation provides a higher level of security and makes the method more resistant to attacks.

Devi, K. R. & Prabakaran et al.[6] proposed an improved bilateral information security system based on DNA sequences that can be used as a supplement to conventional cryptographic systems. The authors proposed a four-stage approach for DNA-based encryption, which includes key generation, encryption, decryption, and key destruction stages, which increases the security and complexity of the system. The results demonstrate that the proposed system exhibits a higher level of security and efficiency than conventional cryptographic systems. The authors concluded that the DNA-based approach can be a potential solution

to enhance bilateral information security in various domains, including the military, government, and financial sectors.

Gehani et al.[7] proposed a new approach to cryptography using DNA. This explains how the DNA molecule can be used as a cryptographic key and how it can be manipulated to perform encryption and decryption operations. This paper describes a number of DNA-based encryption schemes, including one-time pads, block ciphers, and public-key cryptosystems. This paper also discusses the security issues related to DNA-based cryptography and provides possible solutions to overcome these challenges. Gupta, L. M., Garg, D. H., & Samad, D. A. et al. [8] proposes an enhanced DNA-based security model that employs a minimized ciphertext method to enhance security and efficiency in DNA cryptography. The authors first described the traditional DNA-based security model and its limitations. They then introduced their proposed model, which involves converting plaintext into a binary format and then using a random number generator to generate a key. The ciphertext was then encrypted using DNA encoding and decoding techniques. The results demonstrate that the proposed model yields better security and efficiency than traditional DNA-based security models. The outcomes of the defined model are effective for achieving DNA cryptography. Hammad, B. T., Sagheer, A. M., Ahmed, I. T., & Jamil, N. et al.[9] presented an overview and comparison of symmetric and asymmetric DNA-based cryptography techniques. The authors discussed the limitations of traditional cryptographic systems and the potential advantages of using DNA-based cryptography, such as higher security and increased data storage capacity. In conclusion, this study provides a useful comparison of symmetric and asymmetric DNA-based cryptography techniques and highlights the potential advantages and limitations of using DNA as a basis for cryptographic systems.Kaundal, A. K. & Verma, A. K. et al.[10] proposes an extension of the Feistel structure to DNA cryptography. The proposed DNA-based Feistel structure involves dividing the plaintext into two parts, each of which is encoded using DNA strands. The round function operates on two strands using a combination of XOR and AND operations, followed by a permutation. The authors used a 5-bit representation for each nucleotide base and a lookup table to generate the S-boxes used in the round function. The swap operation was performed by interchanging the two DNA strands. The results show that the proposed structure has good cryptographic properties and can be used to generate secure ciphertext. The proposed structure could serve as a starting point for the development of new DNA-based cryptosystems.

Majumder et al [11] proposed the use of DNA-based message encoding for secure data communication and cryptography. The author proposed a technique that uses DNA sequences as a basis for encoding messages, which are then transmitted over a communication channel. This technique involves breaking down the message into smaller units and then converting these units into DNA sequences using a mapping algorithm. The DNA sequences were then transmitted over a communication channel and decoded at the receiver end using another mapping algorithm. Overall, this paper presents an interesting approach to preserving data communication and cryptography using DNA-based message encoding.

Akkasaligar, P. T. & Biradar et al [12] proposed a novel approach for medical image encryption using DNA cryptography. The authors suggested a selective encryption approach that encrypts only the regions of interest (ROIs) in medical images using a DNA-inspired encryption algorithm. The proposed algorithm uses four DNA bases (A, T, C, and G) to represent plaintext pixels and employs a random key generation process using the DNA series. The key generation process uses the Diffie-Hellman key exchange technique to ensure secure key distribution between the sender and receiver. The proposed method was evaluated using two performance metrics, that is, the correlation coefficient and mean square error (MSE), to compare the original and decrypted images. The results prove that the proposed method is effective in achieving high levels of security and maintaining the image quality. Basu, S., Karuppiah, M., Nasipuri, M., Halder, A. K., & Radhakrishnan, N. et al.[13] presents a bio-inspired cryptosystem that combines DNA cryptography and neural networks for improved security and efficiency. The proposed cryptosystem utilizes a DNA-based encryption technique, in which plain text is transformed into a DNA series using a set of predefined rules. The evolved DNA sequence is then subjected to a chaos-based encryption process to obtain the cipher text. The neural network was trained on the DNA sequence of the sender and receiver, and the resulting weights of the trained network were used as a shared secret key for encryption and decryption. Experimental results demonstrate that the proposed cryptosystem outperforms existing DNA-based cryptography methods and traditional encryption techniques in terms of security and efficiency. Babaei et al.[14] proposed a novel method for text and image encryption that depends on the integration of the chaos theory and DNA computing. The encryption algorithm performs a two-stage process, where in the first stage, the plain text or image is converted into a binary form and shuffled using the chaotic logistic

map. In the second stage, the shuffled binary code is encrypted using DNA computing techniques, where DNA strands represent the binary code, and DNA operations represent encryption operations. The experimental results prove that the proposed method provides a high encryption speed and robustness against attacks.

The table 1 shows related to DNA computing revealed the following research gaps.

3. Existing System

The encryption method comprises of two major phases. The first phase translates plain text into DNA codons and the second phase converts DNA codons into variable-length bit patterns based on their frequency. The ASCII values of the characters are used to transform plaintext to binary, which is then stored in the variable M (for example, A the ASCII code is 65 and its equivalent binary code is 1000001and Z ASCII code is 90 and its binary number is 1011010, similar to a to z). The DNA key is generated dynamically and converted into binary using the 2-bit digit codes 00, 01, 10, and 11 for A, C, G, and T, respectively, as shown in Table 2. An XOR operation is performed between plaintext M and the DNA key. This intermediate result is then converted to DNA sequences[2]. The number of occurrences of A, T, G, and C were calculated from this DNA form of ciphertext. The second phase included the construction of an optimal prefix code for DNA nucleotide bases using prefix coding. It removes the two nucleotides with the lowest frequencies from the set and creates a subtree with these two characters as the leaves. The root of this subtree is assigned a frequencythat is the summation of the frequencies of these two characters. The root of the subtree, along with the frequency, was added to the nucleotide set. This procedure was repeated until only one symbol remained in the set. To obtain the prefix code, the left edge is marked 0 and the right edge is marked 1. The prefix code for the nucleotide base is "path from root to leaf." The final ciphertext is obtained by encoding A, T, G, and C according to their prefix code words[7].

3.1 Prefix Encoding

Prefix encoding is a type of data compression technique that assigns a unique binary code to each symbol in a set of symbols such that no code is a prefix of another code. This property makes it possible to decode a sequence of codes unambiguously by simply examining their binary representation. Prefix encoding is widely used in various applications such as text and image compression, data transmission, and cryptography [15].

Steps in prefix coding

- Calculate the frequency of occurrence of each symbol in the message.

Table 1. *Summary of the Literature Review*

Author	Process/Method	Research Gaps
[4]	DNA for a shared secret key cryptosystem with Diffie hellman key sharing technique	The manipulation of DNA sequences takes a lot of time to work with DNA sequences as compared to RSA, DES etc.,
[5]	DNA cryptography based on dynamic mechanisms i.e. ‘dynamic sequence table’ and ‘dynamic DNA encoding’.	The process is adaptable only to the plain text it should be extended to images, audio, video files etc.,
[6]	Improved bilateral information security system based on DNA sequences	uses classical cryptographic technique of substitution method in DNA Cryptography
[8]	Improved DNA-based security model that employs a reduced ciphertext technique to enhance security and efficiency.	Algorithm may be improved to obtain reduced cipher text with significant smaller encryption time for smaller as well as large file size data.
[9]	Overview and comparison of symmetric and asymmetric DNA-based cryptography techniques	The computational time of RSA algorithm combined with DNA coding was longer. less efficient than the compressed DNA asymmetric cryptography.
[10]	Extension of the Feistel structure to DNA cryptography	Requirement of maximum computation time, hi-tech bimolecular laboratory and high computational complexity.
[11]	Secure Data Communication and Cryptography Based on DNA Based Message Encoding	Requires more computation time.
[12]	Selective encryption approach that encrypts only the regions of interest (ROIs) in medical images using a DNA-inspired encryption algorithm	Due to the unique nature of DNA cryptography and very complex confusion property of chaotic map, they require more computational time.
[13]	Bio-inspired cryptosystem that combines DNA cryptography and neural networks for improved security and efficiency	Additional security and authentication measures to strengthen the work.

Table 2. *DNA Conversion table*

Nucleotides	Binary value
A	00
C	01
G	10
T	11

- The symbols are sorted according to their frequencies in ascending order
- Create a binary tree with each symbol at a leaf node, where each internal node represents the sum of the frequencies of its children, as shown in Figure 2.
- The two nodes with the lowest frequency are combined into a new node, whose frequency is the sum of the two nodes. This step is repeated until only one node remains, which is the root of the tree.
- Assign code word 0 to each left branch and code word 1 to each right branch, as shown in Figure 3.
- Traverse the tree from the root to each leaf node and record the code words associated with each symbol.

Suppose we have a message "ABBCCCDDDDDEEEE" and we want toencode it using prefix encoding. Here are the steps:

- The frequency of each symbol is calculated: A = 1, B = 2, C = 3, D = 5, and E = 4.
- Sort the symbols by ascending order frequency: A: 1, B: 2, C: 3, E: 4, D: 5
- Create the binary tree:

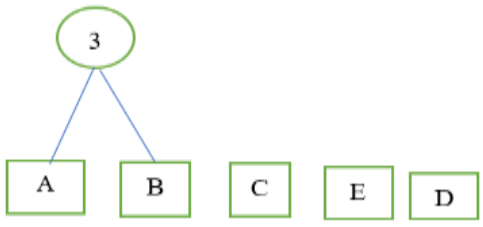


Figure 2. *Tree Structure Representing the Frequency*

- Assign code words:
- The prefix codes for each symbol areA, 000; B, 001; C, 01; D, 11; E, 10.
- The encoded message is:
- 000001001010101111111111110101010

At each step of the algorithm, a binary tree was constructed.

3.2 Encryption Algorithm

- Convert plaintext to binary using ASCII values of characters and store itin a variable M.
- Generate a dynamic DNA key and convert it into binary code using 2-bit digitcodes 00, 01, 10, and 11 for A, C, G, and T, respectively.

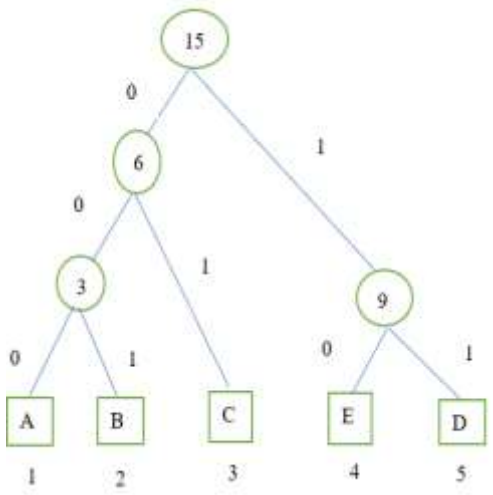


Figure 3. *Tree Structure Representation after assigning Code Words*

- Perform an XOR operation between plaintext M and the DNA key toobtain an intermediate result.
- Convert intermediate results into DNA sequences.
- Count the number of occurrences of A, T, G, and C from the DNA formof the ciphertext.
- Construct an optimal prefix code for DNA nucleotide bases using prefix coding. The two nucleotides with the lowest frequencies were removed from the set and a subtree was created with these two characters as leaves. The root of this subtree is assigned a frequency that is the summation of the frequencies of these two characters. The root of the subtree, along with the frequency, was added to the nucleotide set. This procedure was repeated until only one symbol remained in the set.
- The left edge of the subtree is labeled as 0 and the right edge as 1 to obtainthe prefix code for the nucleotide base.
- Encodes A, T, G, and C, as per their prefix code words, are used to obtain the final ciphertext.

Table 3 lists the algorithms with their input and output.

3.3 Decryption Algorithm

- The first step in decryption is to obtain the DNA key used in theencryption process.
- The encrypted message, which is in the form of DNA sequences,is converted back into binary
- using the prefix codes assigned to A, C, G, and Tduring the encryption process.
- Perform an XOR operation between the obtained binary code and the DNA key.

Table 3. *Encryption Process of Existing Method*

Step	Process	
1	Plain Text (Input)	Have a great day
2	Binary Value (M)	01001000011000010111011001100101001000000011000010 01000000110011101110010011001010110000101110100001 10000001100100011000010111100 1
3	DNA OTP (DNA Key)	ATCGACCGGATCACACAACAATCGACCGGATCACACAA CAATCGACCGGATCACACAACAATCG
4	Binary form of OTP(Key)	00110110000101101000110100010001000001000011011 00 00101101000110100010001000001000011011000010110 10 00110100010001000001000011011 0
5	$Ci = M_i \oplus k_e$	01111110011101111111110110111010000100100010101110 01101101110101001100011011000010101011101100010 10 10110101110101011001010100 1111
6	Convert C to DNA with frequency	CTTGCTCTTTGTCTCAAGCACCCTATCGTGGGCGATCG ACCCCTCGAGGGTCCTCCCGCCCATT {‘C’: 25, ‘T’:17, ‘G’:14 ‘A’:8}
7	Phase 2 operation	[[25, [‘C’,”]], [17, [‘T’,”]], [14, [‘G’,”]], [8, [‘A’,”]] [8, [‘A’,”]], [17, [‘T’,”]], [14, [‘G’,”]], [[25, [‘C’,”]] right = [8, [‘A’,”]] left = [14, [‘G’,”]] right = [17, [‘T’,”]] left = [22, [‘A’,’O’], [‘G’, ‘1’]] right = [[25, [‘C’,”]] left = [39, [‘T’,’O’], [‘A’,’10’], [‘G’,’11’]] [[‘C’,’0’], [‘T’,’10’], [‘A’,’110’], [‘G’,111’]]
8	Prefix code for nucleotide bases	{‘C’: bitarray (‘0’), ‘T’: bitarray (‘10’), ‘A’: bitarray (‘110’), ‘G’: bitarray (‘111)}
9	Final Ciphertext	bitarray(‘01010111010010101011110010011011011101100 00101101001111011111111101111101001111100000100 11 11101111111111000100001110001101010’)

Table 4. *Decryption Process for the Existed Method*

Step	Process	
1	Final Ciphertext	bitarray(‘01010111010010101011110010011011011101100 00101101001111011111111101111101001111100000100 11 11101111111111000100001110001101010’)
2	DNA OTP (Key)	ATCGACCGGATCACACAACAATCGACCGGATCACACAA CAATCGACCGGATCACACAACAATCG
3	Corresponding Nucleotide	CTTGCTCTTTGTCTCAAGCACCCTATCGTGGGCGATCG ACCCCTCGAGGGTCCTCCCGCCCATT
4	Binary Ci	01111110011101111111110110111010000100100010101110 01101101110101001100011011000010101011101100010 10 10110101110101011001010100 1111
5	$Mi = Ci \oplus ke$	01001000011000010111011001100101001000000011000010 01000000110011101110010011001010110000101110100001 10000001100100011000010111100 1
6	Plain text	Have a great day

- The result of the XOR operation is converted back to ASCII letters using the ASCII conversion table.
- The decrypted message is now plain text.
- This process is repeated for each block of the encrypted message to obtain a complete decrypted message.
- Verify the decrypted message to ensure that it matches the original plain-textmessage. Table 4 lists the decryption process of the existing method.

4. Proposed System

The proposed technique uses a DNA One-Time Pad method to transform ordinary text into DNA cipher text. The OTP is a binary stream cipher, where one bit of plain text is encrypted at a time by an exclusive or (XOR) addition with the corresponding bit in the secret key. In traditional OTP encryption, the absolute randomness of secure keys is required; thus, implementation has become more difficult. Owing to the huge storage capability of DNA[16], new

Table 5. *Decryption Process for the Existed Method*

Step	Process	
1	Final Ciphertext	bitarray('010101110100101010111100100110111011011001 00101101001111011111111101111101001111100000100111 11101111111111000100001110001101010')
2	DNA OTP (Key)	ATCGACCGGATCACACAACAATCGACCGGATCACACAA CAATCGACCGGATCACACAACAATCG
3	Corresponding Nucleotide	CTTGCTCTTTGTCTCAAGCACCCTATCGTGGGCGATCG ACCCCTCGAGGGTCCTCCCGCCCATT
4	Binary Ci	011111100111011111111011011010000100100010101110 011011011101010011000110110000101010111011000101 0 1011010111010101100101010011111
5	Mi = Ci ⊕ ke	010010000110000101110110011001010010000000110000010 010000001100111011100100110010101100001011101000 0 100000011001000110000101111001
6	Plain text	Have a great day

developments in OTP are needed. A binary stream sequence of any length can be easily developed. If the key used in the OTP is randomly generated and not used more than once, the size of the cipher text is also reduced with the help of the compression method (modified run length code). In the normal run-length code, repetitions are missed in the input string during expansion. This method resulted in cipher text whose size was smaller than the corresponding plain text. The observational results of the identification method showed that the cipher text to plain text ratio decreased[17,4].

4.1 Encryption

The encryption process in DNA cryptography using DNA one-timepads (OTPs) involves the following steps.

- Transform the given plaintext into binary text based on the ASCII value of the character.
- First, a random DNA series was generated as a one-time pad (OTP). The OTP must be at least as
- long as the message is encrypted.
- Convert the randomly generated DNA OTP key into binary text.
- Perform the XOR operation between the DNA OTP binary and plain-text binary.
- Convert the XORed binary into DNA nucleotides A, C, G, and T usingthe 2-digit binary 00, 01,10 and 11, respectively.
- Take the DNA-converted XOR string and apply the modified run-length code.
- We return this as the encoded output.

4.2 Modified Run length code

- Initialize 0 as a counter variable 'count' to 1 and an empty list 'result.'
- Each character of the input string is iterated from the second character.
- The current character was compared with the previous one. If they are equal, increment the counter
- Count by 1.
- Append the count (if count is greater than 1) if the present character is not equal to the existing
- Character and previous characters in the previous character to the 'result' list.
- Reset the count to 1.
- After the iteration, the last count and character were appended to the 'result' list.
- Joining the 'result' list into a string.

The encryption process is presented in Table 5.

Example: "HELLO"

1. H: 01001000 E: 01000101
   L: 01001100 L: 01001100
   O: 01001111

0100100001000101010011000100110001001111

2. Generate the OTP: OTP: "GATTA"
3. Convert the randomly generated DNA OTP key into binary text:
   G: 01000111 A: 01000001 T: 01010100
   T: 01010100 A:01000001

OTP binary:
0100011101000001010101000101010001000001

4. XOR operation between the DNA OTP binary and plain-text binary

   H: 01001000 XOR 01000111 = 00001111
   E: 01000101 XOR 01000001 = 00000100
   L: 01001100 XOR 01010100 = 00011000
   L: 01001100 XOR 01010100 = 00011000
   O: 01001111 XOR 01000001 = 00001110

00001111 0000010 000011000 0001100000001110

5. Convert the XORed binary into DNA nucleotides A, C, G, and T using the 2-digit binary 00, 01, 10, and 11, respectively:

   00: A 01: C 10: G 11: T

Encrypted DNA sequence:
AGGGCCCTGAGAGAGTTGAA

6. Apply modified run-length encoding
A3G3CTGAGAGAG2TG2A
7. Result of encoded:
A3G3CTGAGAGAG2TG2A

Table 6. *Steps for the Encryption*

step	Operation	
1	Plain Text	Have a great day
2	Binary Value (M)	01001000011000010111011001100101001000000110000 10010000001100111011001100100101010000101110100 00001000000110010001100001011110001
3	DNA OTP (DNAKey)	ATCGACCGGATCACACAACAATCGACCGGATCACACAACAATCG ACCGGATCACACAACAATCG
4	Binary form of OTP(Key)	00110110000101101000110100010001000001000011011 00000101101000110100010001000001000011011000001010 10100011010001000100000100000110110
5	C1=M⊕Ke	01111110011101111111101101110100001001000101011 10011011011010011000110110000101010111011000 10101011010111010101100101001111
6	Convert C to DNA	CTTGCTCTTTGTCTCAAGCACCCTATCGTGGGCGATCGACCCCTC GAGGGTCCTCCCGCCCATT
7	Phase 2 Operation Add run length code	C2TGCT3TGTCTC2AGCA3CTATCGT3GCGATCGATCG4CTCGA3 GT2CT3CG3CA2T

Figure 4 shows the encryption process. Initially, the input of plain text is converted into binary text and then a one-time pad (OTP) is generated. In this OTP, if the key generated is used more than once, it makes the system easily breakable; therefore, in our work, we do not use the key more than once. The randomly generated DNA OTP key was transformed into binary text. XOR operations are performed on the DNA OTP binary with plain text and then the XORed result is converted into DNA nucleotides[18,19].

For this, DNA nucleotides apply the run-length code, which is a form of lossless data compression that displays the sequence of redundant data as a single data value, which in turn represents the repeated block. In this run-length code, the original data are not accessed immediately; we must decode everything if we want to access the data. The proposed method uses a count variable that automatically increases if the current character is equal to the previous character of the input string [20,21]. If the count is greater than one, the present character is not equal to the existing character and the previous character appends the count value to the result and resets the count to one. After completing all iterations, the result list is joined into a string. With the help of this run-length code, the size of the cipher text is reduced compared to existing approaches[22,23].

4.3 Decryption

The steps of the decryption process are as follows.

1. Return the encrypted data for decomposition
2. Apply the decomposed modified run length code to encryption data

Convert DNA into binary format, where A, C, G, and T using the 2-digitbinary 00, 01,10 and 11,

3. respectively.
4. The same DNA key used in the encryption process was convertedinto a binary format.
5. Perform XOR for DNA binary and binary formats of the DNA key to obtainthe binary format of text.
6. The binary format of the text is converted into text using reverse ASCIIvalues.

The results are presented in plain text. Table 6 presents the decryption process.

Example

1. Return the encrypted data for decomposition.
A3G3CTGAGAGAG2TG2A
2. Apply the decomposed modified run-length code to the encrypted data.
AGGGCCCTGAGAGAGTTGAA
3. Convert DNA to binary format: 00 10 10 10 10 10 10 01 10 00 11 10 00 10 00 10 00 11 11 00
4. Convert DNA key GATTA to binary format: 0100011101000001010101000101010001000001
5. Perform XOR for DNA binary and binary formats of the DNA key and then obtain aplain text binary:
0100100001000101010011000100110001001111
6. The binary format of the text is converted into text using reverse ASCIIvalues:
HELLO
7. The plain text is HELLO.

5. Results

The ciphertext size of the proposed method was evaluated by taking samples for different file size categories ranging from 1MB to 5MB.The, which provides better storage performance and reduces transmission costs when DNA cryptography is the information is encrypted with the four characters A, T, G, and C. A few groups of DNA molecules are sufficient to store a large amount of information worldwide. DNA cryptography can also correct and collated with other classical methods [24,25]. In the past, encrypted text was always larger or the same size as plaintext. However, the results of the suggested technique demonstrate that the ciphertext size is consistently smaller than that of the plaintext. The below Table 7 shows that the size of plain text and cipher text generated for the existing method [26,27]. The problems with previous cryptography methods are the increased size of the cipher text,

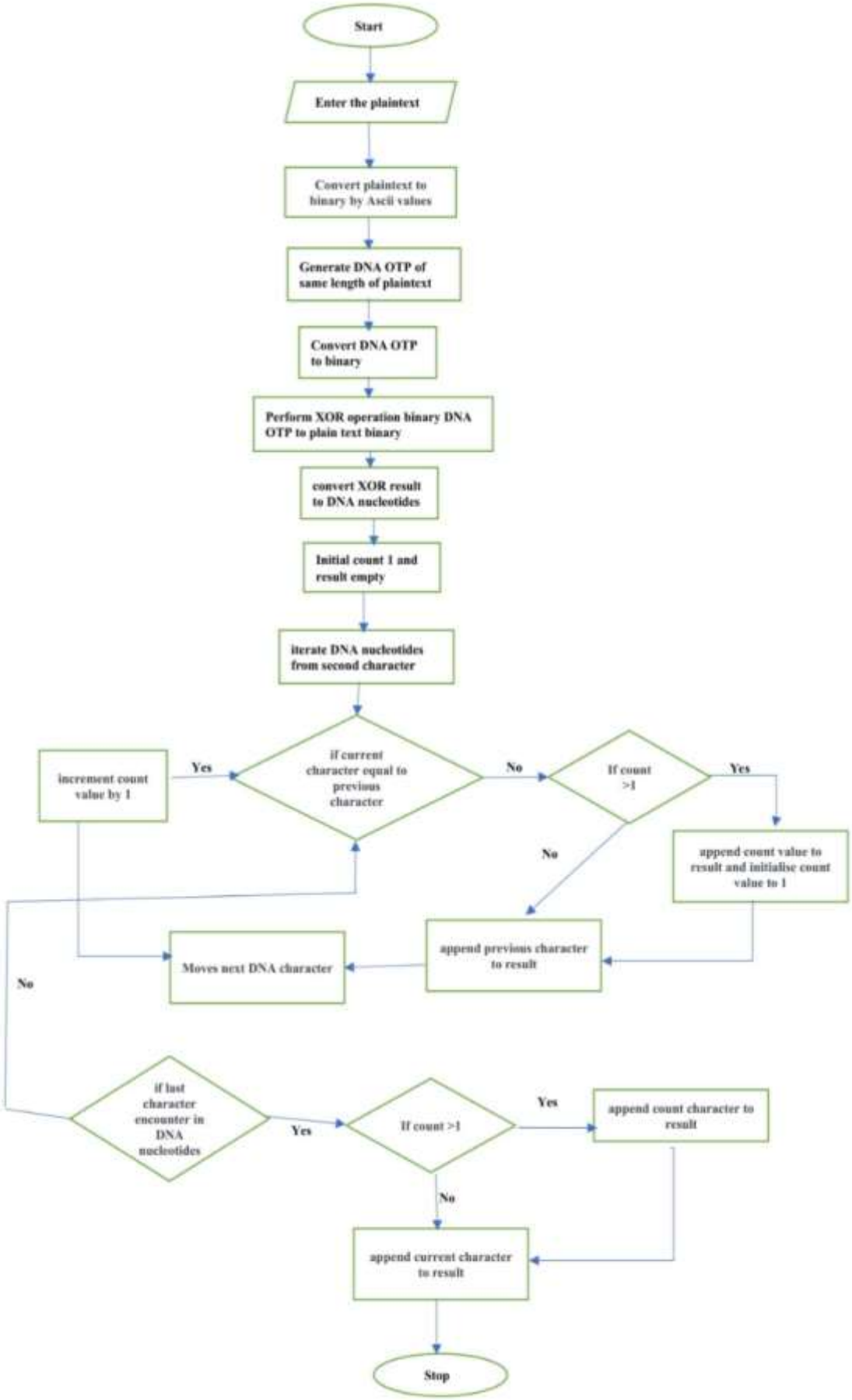


Figure 4. *Flowchart Representation for Encryption Process*

Table 7. *Steps for the decryption*

Step	operation	
1	Cipher text Run length code	C2TGCT3TGTCTC2AGCA3CTATCGT3GCGATCGATCG4CTCGA3GT2C T3CG3CA2T
2	Convert C to DNA	CTTGCTCTTTGTCTCAAGCACCCTATCGTGGGCGATCGACCCCTCGAGG GTCCTCCCGCCCATT
3	Binary form of C	01111110011101111111101101110100001001000101011100 1101011101010011000110110000101010111011000101010110 1010111010110010101001111
3	DNA OTP (DNAKey)	ATCGACCGGATCACACAACAATCGACCGGATCACACAACAATCGACCG GATCACACAACAATCG
4	Binaryform of OTP (Key)	00110110000101101000110100010001000001000011011000010 1011010001101000100010000010000110110000101101001100 0100010001000001000011011 0
5	Ci¬ Mi⊕Ke	01001000011000010111011001100101001000000110000100100 0000001100111011100100110010101110000101110100001000 0001100100011000010111100 1
6	Plain Text	Have a great day

Table 8. *Plain text size and ciphertext size of the proposed method.*

Cipher Text Size (KB)	970	1940	2910	3881	4850
Plain Text Size (KB)	1024	2048	3072	4096	5120

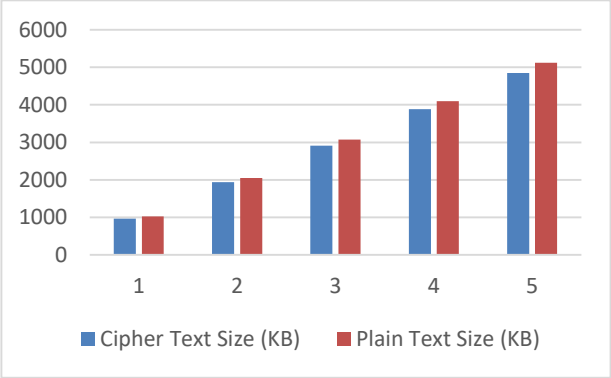


Figure 5. *Plain text size Vs Ciphertext size for the proposed method*

which results in increased storage and data transfer costs [28,29]. The proposed technique uses a DNA OTP method [30] to transform ordinary text into a DNA cipher text; the method used for compression of data is run-length code. This method resulted in cipher text whose size [31,32] was smaller than the corresponding plain text. Figure 5 is the plain text size Vs Ciphertext size for the proposed method. The experimental results in Table 8 and Table 9 show that the cipher text to plain text ratio is decreased among various methods as well as the existing method [33-35].

Figure 6 is the comparison between Traditional Methods & Proposed method. Comparison between Existing method & proposed method is shown in figure 7. Table 10 shows existing method versus proposed method.

6. Conclusion and future work

The traditional binary data uses two digits, 0 and 1, to code the data; however, for the DNA molecule, detect errors during data transmission. The size of the ciphertext plays a prominent role in regulating the cost of storage and transmission. In the past, encrypted text was always larger or the same size as plaintext. However, the experimental results show that after converting plaintext to cipher text, the cipher text size is reduced by upto 20kb for a 1024kb input text file. Similarly, for 2048kb of input text file 40kb, for 3072kb of input text file 60kb, for 5120kb of input text file 50kb of ciphertext size is reduced. The future work of the proposed work is extended to the implementation of DNA-based cryptography with the subject of organic data processing, which includes encryption of genetic statistics or safe organic information transmission and the implementation of DNA barcoding for authentication to improve encoded data safety and reliability. Cryptography is an important method and reported in literatüre [36-38].

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.

Table 9. *Proposed method VS Traditional methods*

Plain Text (MB)	Cipher Text (MB)			
	Majumder	Gupta	Blowfish	Proposed
1	2	1.33	1.8	0.97
2	4	2.67	3.7	1.94
3	6	4	5.28	2.91
5	10	6.65	9.23	4.85
10	20	13.5	18.43	9.7

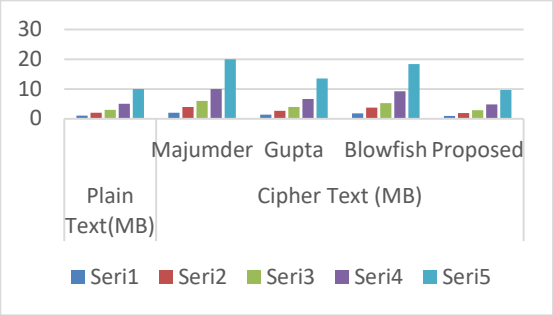


Figure 6. *Comparison between Traditional Methods & Proposed method*

Table 10. *Existing Method Vs Proposed Method*

Plain Text Size (KB)	Cipher Text Size (KB)	
Input file Size (KB)	Existing Method	Proposed Method
1024	990	970
2048	1980	1940
3072	2970	2910
5120	4900	4850
10240	9850	9700

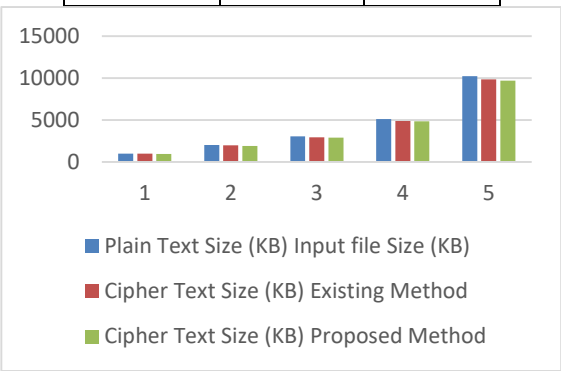


Figure 7. *Comparison between Existing method & proposed method*

- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Jacob, G., & Murugan, A. (2013). DNA based Cryptography: An Overview and Analysis. International *Journal of Emerging Sciences*, 3(1), 36–42.

[2] Vijayakumar, P., Vijayalakshmi, V., & Rajashree, R. (2018). Increased level of security using DNA steganography. *International Journal of Advanced Intelligence Paradigms*, 10(1/2), 74. https://doi.org/10.1504/IJAIP.2018.089490

[3] Mansouri, D., Yuan, X., & Saidani, A. (2020). A new lossless DNA compression algorithm based on a single-block encoding scheme. *Algorithms*, 13(4). https://doi.org/10.3390/A13040099

[4] Aieh, A., Sen, A., Dash, S. R., & Dehuri, S. (2015). Deoxyribonucleic acid (DNA) for a shared secret key cryptosystem with Diffie Hellman key sharing technique. *In Proceedings of the Third International Conference on Computer, Communication, Control and Information Technology (C3IT)* (pp. 1–6). https://doi.org/10.1109/C3IT.2015.7060130

[5] Biswas, M. R., Alam, K. M. R., & Morimoto, Y. (2019). A technique for DNA cryptography based on dynamic mechanisms. *Journal of Information Security and Applications*, 48, 102363. https://doi.org/10.1016/j.jisa.2019.102363

[6] Devi, K. R., & Prabakaran, S. (2016). An Enhanced Bilateral Information Security towards a Conventional Cryptographic System using DNA Sequences. *Indian Journal of Science and Technology*, 9(39). https://doi.org/10.17485/ijst/2016/v9i39/102067

[7] Gehani, A., LaBean, T., & Reif, J. (2003). DNA-based Cryptography. *In Proceedings of the Springer Conference on Advances in Cryptology* (pp. 167–188). https://doi.org/10.1007/978-3-540-24635-0_12

[8] Gupta, L. M., Garg, H., & Samad, A. (2019). An improved DNA Based Security Model using Reduced Cipher Text Technique. International *Journal of Computer Networks and Information Security*, 11(7), 13–20. https://doi.org/10.5815/ijcnis.2019.07.03

[9]Hammad, B. T., Sagheer, A. M., Ahmed, I. T., & Jamil, N. (2020). A comparative review on symmetric and asymmetric DNA-based cryptography. *Bulletin of Electrical Engineering and Informatics*, 9(6), 2484–2491. https://doi.org/10.11591/eei.v9i6.2470

[10]Kaundal, A. K., & Verma, A. K. (2015). Extending Feistel structure to DNA Cryptography. *Journal of Discrete Mathematical Sciences and Cryptography*, 18(4), 349–362. https://doi.org/10.1080/09720529.2014.995975

[11]Majumder, A., Majumdar, A., Podder, T., Kar, N., & Sharma, M. (2015). Secure data communication and cryptography based on DNA based message encoding. *In Proceedings of 2014 IEEE International Conference on Advanced Communication, Control and Computing Technologies* (pp. 360–363). https://doi.org/10.1109/ICACCCT.2014.7019464

[12]Akkasaligar, P. T., & Biradar, S. (2020). Selective medical image encryption using DNA cryptography. *Information Security Journal: A Global Perspective*, 29(2), 91–101.

https://doi.org/10.1080/19393555.2020.1718248

[13]Basu, S., Karuppiah, M., Nasipuri, M., Halder, A. K., & Radhakrishnan, N. (2019). Bio-inspired cryptosystem with DNA cryptography and neural networks. *Journal of Systems Architecture*, 94, 24–31. https://doi.org/10.1016/j.sysarc.2019.02.005

[14]Babaei, M. (2013). A novel text and image encryption method based on chaos theory and DNA computing. *Natural Computing*, 12(1), 101–107. https://doi.org/10.1007/s11047-012-9334-9

[15]Thomas, A. P., Rachel Jacob, J., & Nair, V. V. (2017). Secret Data Transmission Using Combination of Cryptography & Steganography. Retrieved from www.ijcert.org

[16]Raja, I., Reddy, S., Pradeep, R., & Reddy, K. (2014). DNA seed encryption using distributed polysubstitution choice based transposition techniques. 7(2):475-481

[17]Hebbale, S. B., Giridhar Akula, V. S., & Baraki, P. (n.d.). Tuna Swarm Optimization with 3D-chaotic map and DNA encoding for image encryption with lossless image compression based on FPGA. *International journal of electrical and computer engineering systems.* DOI:10.32985/ijeces.14.1.7

[18]Hemasri, S., Kiran, D. S., Ranichitra, D. A., & Kanna, D. A. R. (2023). Improved Data Encryption Standard Algorithm using Zigzag Scan for Secured Data Transmission. *International Journal of Innovative Technology and Exploring Engineering*, 12(6), 26–37. https://doi.org/10.35940/ijitee.F9516.0512623

[19]Ravichandran, D., Praveenkumar, P., Rayappan, J. B. B., & Amirtharajan, R. (2017). DNA Chaos Blend to Secure Medical Privacy. *IEEE Transactions on Nanobioscience*, 16(8), 850–858. https://doi.org/10.1109/TNB.2017.2780881

[20]UbaidurRahman, N. H., Balamurugan, C., & Mariappan, R. (2015). A Novel DNA Computing Based Encryption and Decryption Algorithm. *Procedia Computer Science*, 46, 463–475. https://doi.org/10.1016/j.procs.2015.02.045

[21]Mondal, M., & Ray, K. S. (2023). Review on DNA Cryptography. *International Journal of Bioinformatics and Intelligent Computing*, 2(1).

[22]Anusudha, K., Venkateswaran, N., & Valarmathi, J. (2017). Secured medical image watermarking with DNA codec. *Multimedia Tools and Applications*, 76(2), 2911–2932. https://doi.org/10.1007/s11042-015-3213-1

[23]M, S., & M, V. (2023). A novel and fast hybrid design of cryptosystems for image via 5-D chaos-based random keys and DNA. *Multimedia Tools and Applications.* 83, 58495–58514 https://doi.org/10.1007/s11042-023-17742-3

[24]Latha, H. R., & Ramaprasath, A. (2023). HWCD: A hybrid approach for image compression using wavelet, encryption using confusion, and decryption using diffusion scheme. *Journal of Intelligent Systems*, 32(1). https://doi.org/10.1515/jisys-2022-9056

[25]Rahul, B., Kuppusamy, K., & Senthilrajan, A. (2023). Dynamic DNA cryptography-based image encryption scheme using multiple chaotic maps and SHA-256 hash function. *Optik*, 289, 171253. https://doi.org/10.1016/j.ijleo.2023.171253

[26]Bhimani, P. (2018). A Review on Cryptography Techniques using DNA Computing. *International Journal of Computer Engineering and Research Trends.* https://doi.org/10.22362/ijcert/2018/v5/i6/v5i604

[27]Nath, I., Bhattacharyya, D., Mandal, A., Kundu, N., & De, O. (2017). NHSKCA: A New Heuristic for Symmetric Key Cryptographic Algorithm. *International Journal of Computer Engineering In Research Trends*, 4(12), 547-553 Retrieved from www.ijcert.org

[28]Bhirud, K., Kulkarni, D., Pawar, R., & Prachi, P. (2016). Data Security Using Elliptic Curve Cryptography. *International Journal of Computer Engineering and Research Trends*, 3(5), 222–225.

[29]Nath, I., Baidya, A., Biswas, S. K., Dam, S., & Singha, K. (2020). NISSC: A New Information Security System Using Cryptography. *International Journal of Computer Engineering and Research Trends.* https://doi.org/10.22362/ijcert/2020/v7/i07/v7i0704

[30]Phuc, H., Luu, H., Sakhi, A., & Latief, M. (2024). Optimizing Group Management and Cryptographic Techniques for Secure and Efficient MTC Communication. *International Journal of Computer Engineering and Research Trends*, 11(2). https://doi.org/10.22362/ijcert/2024/v11/i2/v11i201

[31]Maria Gonzalez, M. Bhavsingh & John Smith. (2024). Advanced DNA Cryptography for Enhanced Data Security Using Compression and Encoding Techniques. *Frontiers in Collaborative Research*, 2(2), 37-47. https://doi.org/10.70162/fcr/2024/v2/i2/v2i204

[32]Elena Petrova, & Ahmed Al-Farsi. (2024). Hybrid DNA Encryption with Adaptive Run-Length Encoding for Secure Big Data Applications. *Synthesis: A Multidisciplinary Research Journal*, 2(3), 22-31. https://doi.org/10.70162/smrj/2024/v2/i3/v2i303

[33]Michael Brown, & Li Wei. (2024). One-Time Pad-Based DNA Cryptography with Enhanced Ciphertext Reduction Strategies. *Synthesis: A Multidisciplinary Research Journal*, 2(1), 10-19. https://doi.org/10.70162/smrj/2024/v2/i1/v2i102

[34]Hiroshi Tanaka, M. Bhavsingh & Sarah Johnson. (2024). Optimized DNA Encryption Algorithms Leveraging Statistical and Run-Length Compression Methods. *Frontiers in Collaborative Research*, 2(3), 34-43. https://doi.org/10.70162/fcr/2024/v2/i3/v2i304

[35]Pierre Dupont, & Chitra Bansal. (2024). Secure DNA-Based Cryptography Using Novel Compression Algorithms for IoT Data Transmission. Macaw *International Journal of Advanced Research in Computer Science and Engineering*, 10(1), 100-109. https://doi.org/10.70162/mijarcse/2024/v10/i1/v10i111

[36]El-Taj, H. (2024). A Secure Fusion: Elliptic Curve Encryption Integrated with LSB Steganography for Hidden Communication. *International Journal of*

Computational and Experimental Science and Engineering, 10(3);434-460. https://doi.org/10.22399/ijcesen.382

[37]P., V., & A., M. R. (2024). A Scalable, Secure, and Efficient Framework for Sharing Electronic Health Records Using Permissioned Blockchain Technology. *International Journal of Computational and Experimental Science and Engineering*, 10(4);827-834. https://doi.org/10.22399/ijcesen.535

[38]S, P., & A, P. (2024). Secured Fog-Body-Torrent : A Hybrid Symmetric Cryptography with Multi-layer Feed Forward Networks Tuned Chaotic Maps for Physiological Data Transmission in Fog-BAN Environment. *International Journal of Computational and Experimental Science and Engineering*, 10(4);671-681. https://doi.org/10.22399/ijcesen.490