IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

**Revue d'Intelligence Artificielle**
Vol. 37, No. 6, December, 2023, pp. 1435-1440
Journal homepage: http://iieta.org/journals/ria

# Randomized Information Hiding in RGB Images Using Genetic Algorithm and Huffman Coding

Asraa Abdullah Hussein*, Rafeef M. Al Baity, Sheimaa A. Hadi

College of Science for Women, University of Babylon, Babil 51001, Iraq

Corresponding Author Email: wsci.israa.abdullah@uobabylon.edu.iq

https://doi.org/10.18280/ria.370607

**Received:** 4 August 2023
**Revised:** 12 September 2023
**Accepted:** 10 October 2023
**Available online:** 27 December 2023

***Keywords:***
*steganography, security, RGB image, LSB, compression, Huffman, hiding*

## ABSTRACT

Protecting information from manipulation and theft is a top priority as a result of progress in technology and the infrastructure of the multimedia network in addition to the development of illegal methods of obtaining information. One of the means of protecting and preserving information is to hide it in a digital medium. The motivation for introducing such a system is to enhance the security of confidential data by providing ways to protect the data and reduce attempts to attack it. The proposed system contains several steps summarized as follows: The sender side includes firstly the stage of generating hiding locations randomly depending on the genetic algorithm (GA) to generate rows and seed to generate columns. Secondly, the stage of including data after compressing it by the Huffman method. Data embedding depends on the pixel index as an indicator to choose one of the three bands to hide using LSB. The recipient side extracts the important information hiding in the two last rows which helps to extract the data and convert it into the original text. The proposed system gained efficiency and robustness with the help of genetic and Huffman where genetic chooses the best way for hidding among a set of suggested solution in addition to the randomness it possesses. The role of Huffman reduce data size and thus increase the cover capacity. System efficiency has been measured by PSNR through conducting a number of experiments that included using set of texts with different sizes and two types of standard colored cover images.

## 1. INTRODUCTION

The increase in the amount of data that is exchanged over the Internet has led to the importance of multimedia security greatly. Methods of protecting information varied between encryption and steganography. Encryption is the conversion of confidential data into an unreadable format [1] which is a sprinkling of letters that do not give any meaning only those who possess the encryption keys can decipher the text and return it to an understandable format [2].

In the field of information protection information steganography is one of the prominent addresses among researchers who have enriched this field with a lot of research [3]. Information steganography is the science of delivering confidential [4] information to the concerned parties through public channels in a hidden manner that no one can notice or realize and this is achieved by concealing it with a digital medium such as sound, image, video and text [5, 6].

Information hiding is the technique of protecting information from direct change or tampering with its content by unauthorized parties and delivering the information confidentially and securely to the requested party. The field of information hiding is considered important and vital due to its many applications in protecting copyrights, protecting personal information, credit cards, and exchanging confidential information [2].

The genetic algorithm is considered as one of the best optimization methods used by computer scientists and engineers. It relies on natural selection to solve practical problems [7]. The genetic algorithm starts with a random initial population of individuals where each individual represent a solution to a problem and it updates the population iteratively a number of times by performing a set of operations selection, crossover and mutation [8].

Text compression is still one of the most important fields despite companies trying to introduce and produce many devices dedicated to [9] dealing with huge amounts of information because it has become impossible to deal with data without compression in light of the increasing use of Internet networks [10]. Compression is a vital and important technology that cannot be [11] dispensed with as a result of the development in the age of technology and information in which we live. Its goal is to reduce the amount of data that we need to store or transmit over the Internet [12, 13]. There are two types of compression, the first type includes data loss when decompressing and is called lossy while the second type is called lossless because the data is retrieved after decompressing without loss any data such as Huffman coding and arithmetic coding [14, 15].

One of the classic compression algorithms that do not lose information during its work was invented by the world davied Huffman and is widely used in many fields [16]. The output of this method is codes of different lengths where the shortest code is allocated to the most frequent symbols and the longest code to less frequent so in this way there will be a reduction in size of the data to be compressed [17, 18].

The challenges and limitations facing hiding systems center around capacity, secrecy, and efficiency. The challenge of security has been overcome using GA which is consideration one of the optimization methods. It is characterized by proposing a set of solutions improving these solutions and then choosing the best solution. This helps to strengthen efficiency and confidentiality. Increasing the capacity of the media cover used to embedding data and improving its quality by reducing the size of the data using one of the compression methods such as Huffman [14].

## 2. RELATED WORK

Abdulwahed [19] presented a proposed hiding system called key adaptive LSB (NSKA-LSB) which consists of four stages pixel identification, secret data collection, embedding, and finally retrieval. The methods used in this system are the integration of random functions with a chaotic map, PSNR= 72.44.

In the year 2020 a hiding system was built by Al-Khateeb and Jader [20] dependent on encryption text with DNA to be hidden by hyper chaotic technology and then using this system to choose pixels from the cover image to hide an encrypted text inside it, PSNR=65.7599. The research suggests a way [21] to hide the data to protect and preserve it. The system starts by using RLE for compressing data to reduce its size and increase the efficiency of the system, then apply the LSB method to hide the data by adopting a random form by choosing pixels' sites, PSNR=57.16. The researchers in this work Barovih et al. [22] aim to combine the modified least significant bit method with the shape modification technique to determine the pixel of the cover image that will be used to hide the encrypted data inside it. The Message Length for the test is 96 and PSNR=54,88.

Sari et al. [23] proposed an AES algorithm to encrypt the data and DWT to hide it and noticed that the DWT method suffers from a lack of capacity so to overcome this problem they used Huffman to reduce the size of the data and thus increase the capacity, PSNR=46db. This method [24] combines techniques MSB with LSB based on color images. The researchers proposed a new method that used MSB as a check and then replace bits of LSB with secrets message, PSNR=72.023. Ali et al. [25] presented a data-hiding method based on the use of a Pseudo Random Number Generator (PRNG) to generate 7 locations randomly for each band of the color image. Make xor operation between the bit of randomly generated location with the data bits and the output is included in the 8 bit and the process is repeated for each band of the color cover image, PSNR=90 when embedding data" HelloWorld". The following Table 1 explains summary of related work.

**Table 1.** Illustrate summary of related work

| NO. Reference | Methods | Metrics | Year |
|---|---|---|---|
| [19] | Adaptive LSB | PSNR, MSE | 2020 |
| [20] | DNA with hyperchaotic system | PSNR, MSE, and CORRELATION | 2020 |
| [21] | RLE, LSB, Random Function | PSNR | 2020 |
| [22] | Modified LSB Method with shape Modification | PSNR | 2021 |
| [23] | AES, DWT and Huffman | PSNR,MSE and Histogram | 2019 |
| [24] | MSB and LSB | PSNR | 2021 |
| [25] | PRNG and XOR | PSNR | 2019 |
| Prposed System | Genetic algorithm, Huffman coding and index of pixels as indicator | | |

## 3. PROPOSED SYSTEM

The problem that the proposed system addresses is to provide a system to protect important information by hiding it randomly in RGB cover images. Reducing the size of the confidential data by compressing it using Huffman encoding to converting it into a code of (0, 1) then embedding the compressed string by going through two stages: the stage of selecting row locations randomly based on using genetic algorithm and stage two represent generate columns through the use of seed in addition to the process of including data within the image based on the index of the site. The idea of adopting genetic algorithm for selecting rows came as a result of the need to select a number of rows randomly without the intervention of the system designer and the same time it is difficult for manipulators and hackers to know which rows were chosen for hiding. The columns were identified using the seed to increase randomness and thus increase data protection procetures. Figure 1 shows the details of block diagram for sender.The following steps represent the sender and recipient side:

**Sender Side "generation hiding sites and embedding process"**

1. Read secret data and cover images.
2. Compress data based on Huffman encoding to convert the text into a codeword (0, 1).
3. Hiding process: This process includes the stage of generating hiding sites and the stage of embedding the data in the cover image.

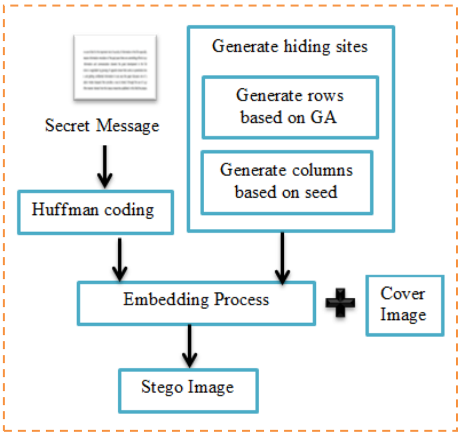

**Figure 1.** The block diagram of the sender side

### 3.1 Generate hiding locations

**Stage one**: **“generate rows randomly based on genetic”:**
A genetic algorithm is characterized by generating several generations and each generation contains a set of solutions this solution is called a chromosome which is represented by a vector that contains a group of cells called genes and an initial population generated randomly.

**Step one: Coefficients of Genetic**
Specify the parameters for the genetic that is a need in this step as follows:

1. Chromosome Length=No. of rows in RGB cover image (where each gene from chromosome allocate to each row in the image).
2. NO. Of chromosome in each generation=50.
3. NO. Of generations=30.
4.Suggested crossover probability=0.8.
5. Suggested mutation probability=0.2

Determine the values of the above parameters left to the designer of algorithm through the experiment except chromosome length is fixed which is equal to the NO. of rows of the cover image.

**Step two: Initial Population**
Randomly generate several chromosomes from (0 and 1) with a length equal to the number of rows for the cover image and then interpret this chromosome into numbers between (1 and the number of rows-2) where each gene represents a row of the cover rows except the last two rows. If the value of the gene is 0 the row number corresponding to it will be ignored and will not be included in the number of rows used for hiding while if its value is 1 we will take the row number corresponding to it to use for hiding as in the following example supposing the NO. rows of the cover image are 512:

| 1 | 2 | 3 | 4 | 5 | | 509 | 510 | 511 | $NO_{row\text{-}2}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 1 | 1 | ............ | 1 | 1 | 0 | 0 |

The chromosomes above=1,3,4,5,509,510 represent the rows that are used to hide data.

**Step three: Evolution of chromosome**
A PSNR measure is used to evaluate the chromosomes (representing the rows selected from the cover image) in each generation where the higher chromosome is the most candidate for selection to be the number of rows used for hiding.

**Step four: Generate the Number of Generations**
One of the advantages of the genetic algorithm is that it is not satisfied with one generation of solutions but rather continues for several generations by using its procedures to look for the best solutions. In this paper, the following genetic processes were used:

a. Selection process based on binary set method.
b. Perform crossover by using a uniform method.
c. Using (2m) mutation.

**Stage two: “Generate columns randomly based on seed”:**
Firstly, find the number of columns according to the equation:

$$\text{Columns} = \text{length}(\text{codeword})/\text{length}(\text{key1}) \qquad (1)$$

where, key1: represents the number of rows provided from genetic, codeword: represents text after compression, Columns: This represents the number of columns we need to hide in each row.

The seed is used with one of the simple shuffle functions to specify which columns will be selected in each row to hide the data.The following example shows how to generate columns(y) from the cover image.

Rng (seed).
Y=randperm (No. of columns cover, Columns(NO. of columns)).

### 3.2 Embedding

The distribution of confidential data on pixels selected randomly is done as follows:

**a.** If the index (i, j) of the selected pixel from the cover RGB image is even then it is selected Rchanel.

**b.** If the index (i, j) of the selected pixel from the cover RGB image is odd then it is selected Gchanel.

**c.** If one of the indexes (i, j) is odd and the other is even from the cover RGB image then it will be chosen Bchanel.

The pixel index was used as an indicator for choosing the band to distract hackers and attackers attention which bands contain data hidden and added some of randomness at embedding. The index pixels are either (even or odd) or one of them is odd and the other is even also there are three bands(R,G,B) for cover image. Therefore, we took advantage of this thing by adopting the index as an indicator .

The final step is to store the information that needs to be exchanged with the recipient in last two rows. This information represents the rows key and the seed value with the required number of columns in addition to the length of the secret text. The rows key(0,1) occupies the penultimate row where the recipient takes what corresponds to 1 and discards what corresponds to 0. The rest of the information is convert to binary and stored in the last row.

Assuming the following locations are randomly generated [(4,8), (3,1), (2,2), (10,5)], the process of distributing data bits to bands will be shown in the following Table 2.

**Table 2**. Explain the embedding process

| Locations | Type of Index | Chanel from Cover |
|---|---|---|
| **(4,8)** | I=4, j=8>>even | Rchanel |
| **(3,1)** | I=3, j=1>>odd | Gchanel |
| **(2,2)** | I=2, j=2>>even | Rchanel |
| **(10,5)** | I=10,j=1>> One of the indexes is odd and the other is even | Bchanel |

### Receiver side

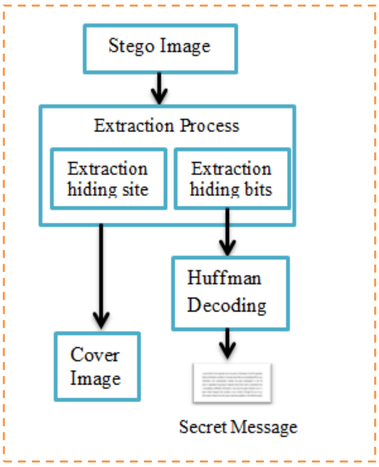

**Figure 2.** The block diagram of the receiver side

The first step that the recipient takes when receiving the hiding image is to extract the important keys from last two rows. After that, the data bits are extracted collected as a vector and then decompressed to return the original data. Convert the binary string to decimals and then letters. The Figure 2 illustrates the block diagram of receiver side.

## 4. RESULT AND DISCUSSION

### 4.1 Information collection

The proposed system is tested using a set of secret texts with different sizes and standard cover images of dimensions (512*512) that obtain from the Internet as shown in Figure 3.

**Figure 3.** Cover images for the proposed system

### 4.2 Quality measures

There are several ways to measure the efficiency of the proposed system such as Peak Signal-to-Noise Ratio (PSNR). It can be used to evaluate the level of distortion or noise introduced by the steganography technique so a higher PSNR means a more efficient method. The formula for calculating PSNR is as follows:

$$\mathrm{PSNR} = 10 * \log10((\mathrm{MAX}\hat{\ }2)/\mathrm{MSE}) \qquad (2)$$

where, MAX is the maximum possible pixel value of the image (typically 255 for 8-bit images).

### 4.3 Test proposed system

A set of experiments was built to check the proposed system through several texts and cover images of size 512*512. Table 3 and Table 4 present the results of the proposed system.

**Table 3.** Experiment (1) for the proposed system

| Cover Image | Text Size in Bits | Text Size after Compression in Bits | PSNR |
|---|---|---|---|
| | 408 | 176 | 88.0565 |
| | 1032 | 443 | 82.8550 |
| | 3608 | 1829 | 76.4316 |
| | 5048 | 2585 | 74.9842 |

**Table 4.** Experiment (2) for the proposed system

| Cover Image | Text Size in Bits | Text size after Compression in Bits | PSNR |
|---|---|---|---|
| | 408 | 176 | 89.3789 |
| | 1032 | 443 | 83.0734 |
| | 3608 | 1829 | 76.4842 |
| | 5048 | 2585 | 75.0734 |

The proposed system is characterized by following the method of randomness when choosing the locations of pixels in the hiding proses. This gives it strength in terms of even if the stego image falls into the hands of thieves and hackers the data inside it remains under protection because hackers still need to know the hidden locations in which the data is carried. The data is also present in one of the bands according to the index of pixels as previously described.

The size of the data plays an important and influential role in the embedding process because the size of the data if it is small will take less space from cover media and therefore the steganography system is strong and difficult for the human eye to notice a difference in the stego image. In this proposed system Huffman method was used to compress the data and the results of the difference in size are shown in Figure 4 whereas Figure 5 shows the histogram for the cover image before and after embedding 1000 bits.

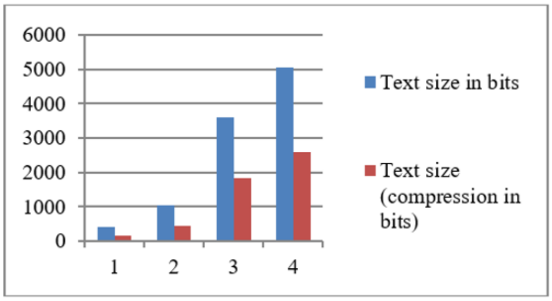

**Figure 4.** Text size before and after compression

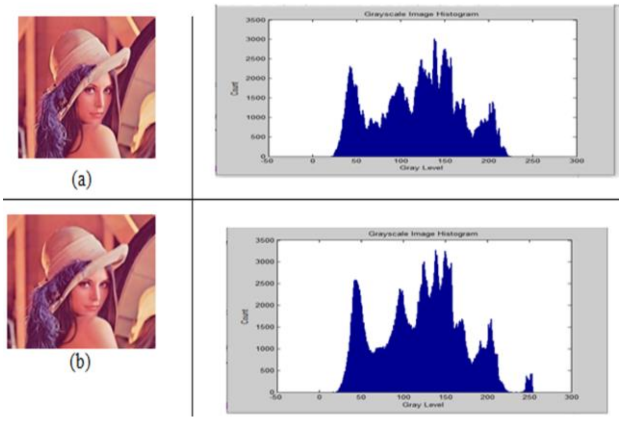

**Figure 5.** (a) The original cover image and histogram (b) The stego image and histogram

Table 5 displays the values of the best row chosen by genetic algorithm and the columns using the seed when hiding data in Lena cover image and for two cycles of execution where the first cycle hides data of size (408 bits) and the second cycle (3608 bits).

A sample of the data (408 bits) was taken and hidden inside Lena cover image. Making a change in the value of GA algorithm parameters for each execution as noted in Table 6 in order to study the effect of this change.

## 5. CONCLUSIONS

It is indispensable to exchange information between people via the Internet in various fields so there must be a way to protect information from theft and manipulation by hiding it in a digital medium. The proposed system includes hiding data after compressing it by the Huffman method to reduce the size of the data and thus increase the efficiency of the system. The compressed data is hidden randomly by selecting the pixel locations in two stages: the stage of selecting the row by adopting a genetic algorithm and the stage of selecting the column through the seed and one of the shuffle functions. The proposed system will have an impact on every field in the real world that specializes or is interested in encryption, information security and data transfer via media cover. Proposed future work is to increase the protection protocol by adding a method of encrypting the data before including it in the cover image. It is also possible to scatter the data using one of the chaotic map methods. The system achieved good results by conducting several experiments.

**Table 5.** Examples of better row and column

| NO. Execute | Texts in Bits | Row Based GA | Column Based Seed |
|---|---|---|---|
| 1 | Texts=408<br>Compresstion text=176 | 3 4 6 7 14 15 18 20 21 22 24 26 27 28 31 32 33 34 36 37 39 41 42 43 45 47 48 50 53 54 55 57 59 61 63 65 66 67 68 69 71 72 74 75 77 80 84 89 90 93 95 96 97 99 100 102 103 ..... 235 236 239 240 242 243 244 246 248 249 252 254 | 12 151 |
| 2 | Texts =3608<br>Compresstion text=1829 | 3 4 5 9 11 13 14 18 22 25 28 31 32 39 40 41 45 46 47 49 52 55 57 58 59 61 62 63 64 66 67 69 71 77 78 79 80 81 83 84 87 88 90 91 94 95 96 97 99 100 101 104 105 106 108 114 116 .... 244 246 247 248 252 254 255 256 | 12 151 53 22 7 104 97 14 207 98 97 193 140 134 245 |

**Table 6.** Results for vary GA parameters

| Detail of Parameter | Population Size | Generation | PSNR |
|---|---|---|---|
| | 30 | 30 | 88.4542 |
| Selection=Binary | 50 | 10 | 89.0940 |
| Crosscver=X1 | 60 | 10 | 89.0256 |
| Mutation=M2 | 50 | 30 | 89.3789 |
| | 7 | 40 | 88.1111 |
| | 30 | 30 | 88.5748 |
| Selection=Binary | 50 | 10 | 88.6989 |
| Crosscver=X2 | 60 | 10 | 88.7623 |
| Mutation=M2 | 50 | 30 | 89.1635 |
| | 7 | 40 | 88.0026 |
| | 30 | 30 | 88.3368 |
| Selection=Binary | 50 | 10 | 88.8267 |
| Crosscver=X1 | 60 | 10 | 89.0256 |
| Mutation=M1 | 50 | 30 | 88.7623 |
| | 7 | 40 | 88.1111 |

## REFERENCES

[1] Naser, M.A., Al-alak, S.M.K., Hussein, A.M., Jawad, M.J. (2022). Steganography and cryptography techniques based secure data transferring through public network channel. Baghdad Science Journal, 19(6): 1362-1362. https://doi.org/10.21123/bsj.2022.6142

[2] Alqadi, Z., Zahran, B., Jaber, Q., Ayyoub, B., Al-Azzeh, J., Sharadqh, A. (2019). Proposed implementation method to improve LSB efficiency. International Journal of Computer Science and Mobile Computing, 8(3): 306-319.

[3] Neamah, R.M., Abed, J.A., Abbood, E.A. (2020). Hide text depending on the three channels of pixels in color images using the modified LSB algorithm. International Journal of Electrical and Computer Engineering, 10(1): 809. http://doi.org/10.11591/ijece.v10i1.pp809-815

[4] Saleh, A.H., Yousif, A.S., Ahmed, F.Y. (2020). Information hiding for text files by adopting the genetic algorithm and DNA coding. In 2020 IEEE 10th Symposium on Computer Applications & Industrial Electronics (ISCAIE), Malaysia, pp. 220-223. http://doi.org/10.1109/ISCAIE47305.2020.9108842

[5] Al Hussien, S.S., Mohamed, M.S., Hafez, E.H. (2021). Coverless image steganography based on optical mark recognition and machine learning. IEEE Access, 9: 16522-16531. http://doi.org/10.1109/ACCESS.2021.3050737

[6] Kadhim, I.J., Premaratne, P., Vial, P.J., Halloran, B. (2019). Comprehensive survey of image steganography: Techniques, evaluations, and trends in future research. Neurocomputing, 335: 299-326. https://doi.org/10.1016/j.neucom.2018.06.075

[7] Cherian, C.S., Rasmi, P.S. (2019). Genetic algorithm and random number generation for symmetric encryption. International Journal of New Innovations in Engineering and Technology, 10(2): 1-5.

[8] Kalaiselvi, K., Gopika, S., Jacob, M. (2021). Optimized symmetric keys generated using genetic algorithm for fully homomorphic encryption system. Bioscience Biotechnology Research Communications, 14(6): 339-343. http://doi.org/10.21786/bbrc/14.6.70

[9] Ahmed, Z.J., George, L.E., Abduljabbar, Z.S. (2020). Fractal image compression using block indexing technique: A review. Iraqi Journal of Science, 61(7): 1798-1810. https://doi.org/10.24996/ijs.2020.61.7.29

[10] Motomura, R., Imaizumi, S., Kiya, H. (2021). A reversible data hiding method in encrypted images for

controlling trade-off between hiding capacity and compression efficiency. Journal of Imaging, 7(12): 268. https://doi.org/10.3390/jimaging7120268

[11] Rahman, M.A., Hamada, M. (2020). Burrows–Wheeler transform based lossless text compression using keys and Huffman coding. Symmetry, 12(10): 1654. https://doi.org/10.3390/sym12101654

[12] Rahman, M.A., Rabbi, M.F., Rahman, M.M., Islam, M.M., Islam, M.R. (2018). Histogram modification based lossy image compression scheme using Huffman coding. In 2018 4th International Conference on Electrical Engineering and Information & Communication Technology (iCEEiCT), Dhaka, Bangladesh, pp. 279-284. https://doi.org/10.1109/CEEICT.2018.8628092

[13] Rahman, M.A., Islam, S.M.S., Shin, J., Islam, M.R. (2018). Histogram alternation based digital image compression using base-2 coding. In 2018 Digital Image Computing: Techniques and Applications (DICTA), pp. 1-8. https://doi.org/10.1109/DICTA.2018.8615830

[14] Chuman, T., Sirichotedumrong, W., Kiya, H. (2018). Encryption-then-compression systems using grayscale-based image encryption for JPEG images. IEEE Transactions on Information Forensics and security, 14(6): 1515-1525. https://doi.org/10.1109/TIFS.2018.2881677

[15] Pandey, M., Shrivastava, S., Pandey, S., Shridevi, S. (2020). An enhanced data compression algorithm. In 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, pp. 1-4. https://doi.org/10.1109/ic-ETITE47903.2020.223

[16] Oswald, C., Sivaselvan, B. (2018). An optimal text compression algorithm based on frequent pattern mining. Journal of Ambient Intelligence and Humanized Computing, 9: 803-822. https://doi.org/10.1007/s12652-017-0540-2

[17] Portell, J., Iudica, R., García-Berro, E., Villafranca, A.G., Artigues, G. (2018). FAPEC, a versatile and efficient data compressor for space missions. International Journal of Remote Sensing, 39(7): 2022-2042. https://doi.org/10.1080/01431161.2017.1399478

[18] Rahman, M.A., Hamada, M. (2019). Lossless image compression techniques: A state-of-the-art survey. Symmetry, 11(10): 1274. https://doi.org/10.3390/sym11101274

[19] Abdulwahed, M.N. (2020). An effective and secure digital image steganography scheme using two random function and chaotic map. Journal of Theoretical and Applied Information Technology, 98(1): 78-91.

[20] Al-Khateeb, Z.N., Jader, M.F. (2020). Encryption and hiding text using DNA coding and hyperchaotic system. Indonesian Journal of Electrical Engineering and Computer Science, 19(2): 766-774. http://doi.org/10.11591/ijeecs.v19.i2.pp766-774

[21] Jaloud, I.S. (2020). Text image compression and hiding using random formula in color images. Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series, 42(4): 163-176.

[22] Barovih, G., Admojo, F.T., Hersaputra, Y. (2021). Steganographic techniques using modified least significant bit and modification reshape transposition methods. ILKOM Jurnal Ilmiah, 13(1): 67-77. https://doi.org/10.33096/ilkom.v13i1.848.67-77

[23] Sari, C.A., Ardiansyah, G., Rachmawanto, E.H. (2019). An improved security and message capacity using AES and Huffman coding on image steganography. TELKOMNIKA (Telecommunication Computing Electronics and Control), 17(5): 2400-2409. http://doi.org/10.12928/telkomnika.v17i5.9570

[24] Mahdi, S.A., Maisa'a, A.K. (2021). An improved method for combine (LSB and MSB) based on color image RGB. Engineering and Technology Journal, 39(1): 231-242. http://doi.org/10.30684/etj.v39i1B.1574

[25] Ali, U.A.M.E., Sohrawordi, M., Uddin, M.P. (2019). A robust and secured image steganography using LSB and random bit substitution. American Journal of Engineering Research (AJER), 8(2): 39-44.