Privacy-Preserving Method for Public Health Surveillance Data using Image Steganography

Ajay Kumar, Electronics & Communication Engineering Dept., Thapar Institute of Engineering & Technology Patiala, India. ajay.kumar@thapar.edu

Abhijit Karmakar, Integrated Systems Dept. CSIR-Central Electronics Engineering Research Institute Pilani, India. abhijit.karmakar@gmail.com

Alpana Agarwal, Electronics & Communication Engineering Dept., Thapar Institute of Engineering & Technology
Patiala, India. alpana@thapar.edu

Abstract: An algorithm of bio-inspired black widow is being optimised and is employed for design privacy-preserving methods for public health surveillance data using image steganography. The algorithm of the Black Widow Optimization is completely dependent upon the mating behaviour of the black spiders. The algorithm contains three basic steps namely, procreate, cannibalism, and mutation. The cannibalism step removes the inadequate solutions while finding the optimal solution. Thus, the BWO algorithm provides early convergence to find optimal solutions as compared to the existing optimization algorithm. In the proposed method, BWO algorithm is used for random key generation and optimized data hiding to enhance the imperceptibility. Initially, secret data is read and pre-processing is done to split the data into sensitive and non-sensitive attributes. After that, sensitive data is encrypted by performing the XOR operation with the random key generated using the BWO algorithm. Next, pre-processing of the cover image is done to select the most appropriate plane for data hiding based on the pixel intensities of the planes. After choosing the appropriate plane, optimized data hiding is done using the BWO algorithm. The BWO algorithm searches the optimal starting pixel and secret data order in the cover image. The simulation evaluation is done on the standard dataset images. The results show that the proposed method achieves superior results over existing optimization methods, namely, GA and Artificial bee colony. Besides that, the appropriate plane, optimal starting pixel, and secret data order is different for different cover images. Thus, it enhances the security of the proposed method.

Keywords: Black Widow Optimization, Data Hiding, Privacy-Preserving, Public Surveillance, Steganography.

Tob Regul Sci.™ 2021;7(6-1): 6814-6830
DOI: doi.org/10.18001/TRS.7.6.1.11

1. Introduction

The main motive of the system of public health surveillance is the data collection, analysis, and interpretation of the disease to control it from spreading [1]. In the last two years, coronavirus disease 2019 (COVID-19) spread globally and initially hit even the European countries with developed healthcare systems [2]. The Ministry of Health is taking preventive actions by analyzing the public data to control it. To achieve this goal, PCR tests are conducted in several laboratories to collect the samples [3]. Further, data used to predict the number of cases and to issue safety guidelines

for the public. However, when a patient gives the sample, an electronic medical record (EMR) is created that contains the patient information, such as name, contact details, date of birth, and symptoms record [4]. Thus, every minute, around the world, patient information is shared on the internet. The EMR contains the sensitive information of the patients and sharing it on the internet violates the privacy of the patient. Therefore, in the literature, privacy-preserving methods are designed to secure the patient's sensitive information on the internet [5]. Cryptography and steganography are the two security fields used for securing sensitive data [6]. Cryptography algorithms scramble the sensitive data using a private key whereas steganography hides the sensitive data in other media. Image is the most preferred cover media in steganography over other media such as text, audio, and video.

In the literature, triple DES [7], Advanced Encryption Standard (AES) [8], homomorphic encryption [9], and XOR operation [10] are the most preferred cryptography approaches to secure sensitive data. Random key plays an important role in cryptography algorithms [11]. In the literature, the random key generation methods are RSA [12], Elliptic Curve Cryptography [13], Linear/Non-linear feedback shift registers [14]. Further, in the current scenario, swarm intelligence algorithms are deployed for key generation based on the objective function. The most preferred algorithms are particle swarm optimization [15], ant colony optimization [16], genetic algorithm, harmony search [17]. However, these algorithms have a lower convergence rate to generate the complete random key.

On the other side, the least significant bit (LSB) based data hiding approaches are used in the steganography to hide the sensitive data in the cover media [18]. The algorithm of LSB needs to substitute the LSB bit of the cover image pixel with the secret data bit [19]. The substitution process generates variability in the cover image. Therefore, various approaches are designed to reduce the variability such as the flipping method [20] and deploying optimization algorithms to search the optimal starting pixel and optimal blocks in the cover image [21-22]. In the literature, the most preferred optimization algorithms are genetic algorithm (GA) [23], particle swarm optimization (PSO) [24], artificial bee colony (ABC) [22]. However, these algorithms have a lower convergence rate to search for the optimal solution. To overcome this issue, a black widow optimization algorithm is deployed in the proposed method. The black widow optimization algorithm is based on the unique mating behavior of black widow spiders [18]. The algorithm of the BWO keeps on providing the early set of convergence due to its cannibalism stage and due to this particular stage all of the solutions within an inappropriate range of fitness are being removed from that of the overall set of solutions. Within Thai literature all of the algorithm of the BWO is being deployed successfully for 51 various sets of benchmarking functionalities in order to verify all of its inefficiencies for obtaining the optimal amount of solutions for the problem [25].

The main contribution of this work is to design a privacy-preserving method using image steganography to secure the public health surveillance data. To achieve this goal, black widow optimization algorithms are taken under consideration. The black widow optimization algorithm is deployed for random key generation for encryption purposes and searching the optimal starting pixel in the cover image for data hiding. The black widow optimization algorithm comes under the swarm intelligence algorithm and provides a better exploration and exploitation rate as compared to the other swarm intelligence algorithms [18]. Proposed method, Initially, public health surveillance data is read. The data contains sensitive and non-sensitive attributes of the patients. Therefore, the data is split into sensitive and non-sensitive data. Next, the sensitive data is encrypted by performing the XOR operation with the random key. The random key is generated using a black widow optimization algorithm. After encrypting the sensitive data, it is concatenated with non-sensitive data. Further, before data hiding, the pre-processing of the cover image is done to select the most optimal plane for data hiding. The selection of the optimal plane is done by determining which plane of the cover image is contributed lesser in the entire image. The optimal plane and data are given to the BWO based data hiding algorithm. The BWO algorithm searches the optimal starting pixel in the cover image and optimal secret data order for data hiding. After searching the optimal starting pixel, the data hiding is done using the k-bit LSB method. The simulation evaluation is done on the standard dataset images. The results show that the proposed method

is superior to the existing method. Besides that, the optimal plane and optimal starting pixel are different for different cover images. Thus, it is difficult to predict which plane contains data and the optimal starting pixel for data extraction.

The rest of the paper is organized as follows. Section 2 describes the related work. Section 3 explains the Black Widow Optimization and K-bit LSB algorithm. Section 4 illustrates the proposed methodology and BWO algorithm deployment for key generation and optimal data hiding. Section 5 shows the simulation evaluation of the proposed method for the standard dataset. Finally, conclusion and future work presented in Section6.

2. Related Work

Section proposed , related work is shown for key generation and data hiding methods.

2.1 Key Generation Methods

Section, we have studied the various key generation methods.

Chandravathi et al. [12], used the RSA algorithm for key generation. The RSA algorithm requires exponential and modular functions. Thus, it consumes a large number of resources in terms of memory and time. Ghayvat et al. [13], used the elliptic curve cryptography (ECC) algorithm for key generation. The ECC is superior in terms of security as compared to the RSA algorithm. However, ECC requires a long lookup table for key generation. Thus, it consumes a large number of resources similar to the RSA algorithm. Next, Linear feedback shift register based key generation algorithm is designed for cryptography by Tuncer et al. [14]. However, prediction of next value is possible by determining the current value and feedback tapping positions. Further, in the current scenario, swarm intelligence algorithm based key generation algorithm gains popularity. These algorithms based on the objective function generate a completely random key. In the literature, genetic algorithm (GA), particle swarm optimization (PSO), ant colony optimization (ACO), harmony search, and differential evolution is used. Sreelaja N.K. and G.A. Vijayalakshmi Pai [15-16], deployed the Particle swarm optimization (PSO) and ant colony optimization algorithm for key generation for stream cipher algorithm. Krishna et al. [17], generated the random key for stream cipher algorithms using evolutionary algorithms. In their work, three algorithms are taken under consideration, namely, genetic algorithm, harmony search, and differential evolution algorithm. They have designed a multi-objective function to enhance the randomness of keys. The simulation results show that they achieve high entropy for the generated key. However, the swarm intelligence algorithms deployed in the literature have a low convergence rate to find the optimal solution.

2.2 Data Hiding Methods

The bio-inspired algorithms have been used in the image steganography for designing optimized data hiding methods to reduce variability. The most popular algorithms deployed for designing optimized data hiding methods in the spatial domains are genetic algorithm (GA), particle swarm optimization (PSO), and artificial bee colony (ABC).

Kanan et al. [21], presented an optimized data hiding method using the genetic algorithm (GA). The genetic algorithm is deployed for searching the optimal starting point in the cover image to enhance the visual quality. To achieve this goal, 7 parameters are used. However, these parameters need to communicate with the receiver to extract secret data. Aman Banharnsakun [22], developed an enhancing LSB algorithm by deploying the artificial bee colony (ABC) algorithm. In their method, the cover image is split into 16 blocks and optimal block order is determined using the ABC algorithm for data hiding. However, the optimal block order needs to communicate with the receiver to recover the original order of secret data. Wazirali et al. [23], used the genetic algorithm to determine the optimal secret data order to reduce variability. The secret data orders used in their proposed method are scanning, shifting, flipping, transposing, LSB matching checking, and secret data embedding. However, 5 genes have been used that need to communicate with the receiver. Besides that, the convergence rate of the genetic algorithm is low because in each iteration only two optimal solution generations are possible. Bedi et al. [24], designed an optimized data hiding method using the particle swarm optimization (PSO) algorithm. The PSO algorithm is deployed to search the optimal pixels

in the cover image for data hiding. However, all optimal pixel information needs to communicate with the receiver through a secure channel which is an overhead on the data hiding method.

3. Preliminaries

In this section, an overview of black widow optimization and the LSB method is given to understand the proposed method.

3.1 Overview of Black Widow Optimization (BWO) Algorithm

Inspired by the strange mating behaviour of black widow spiders, the Black Widow Optimization Algorithm (BWO) was developed. Cannibalism is a unique phase of this approach. This stage eliminates species that aren't well-suited to the circle, resulting in early convergence [25]. Using 51 benchmark functions, the BWO algorithm is tested for its ability to find the best possible solution to a given issue. Figure 1 depicts the algorithmic flow of the black widow spider. With every spider representing a possible solution, the BWO algorithm begins with a population of starting spiders. These first spiders, in pairs, are attempting to start a new colony. During or after mating, the female spider eats the male. Afterwards she releases stored sperm into egg sacs through her sperm thecae. The spiderlings emerge from the egg sacs as early as 11 days after already being placed. Sibling cannibalism may be witnessed at this phase, which can last from a few days to a week. They are then whisked away by the wind as they make their way back to civilization.

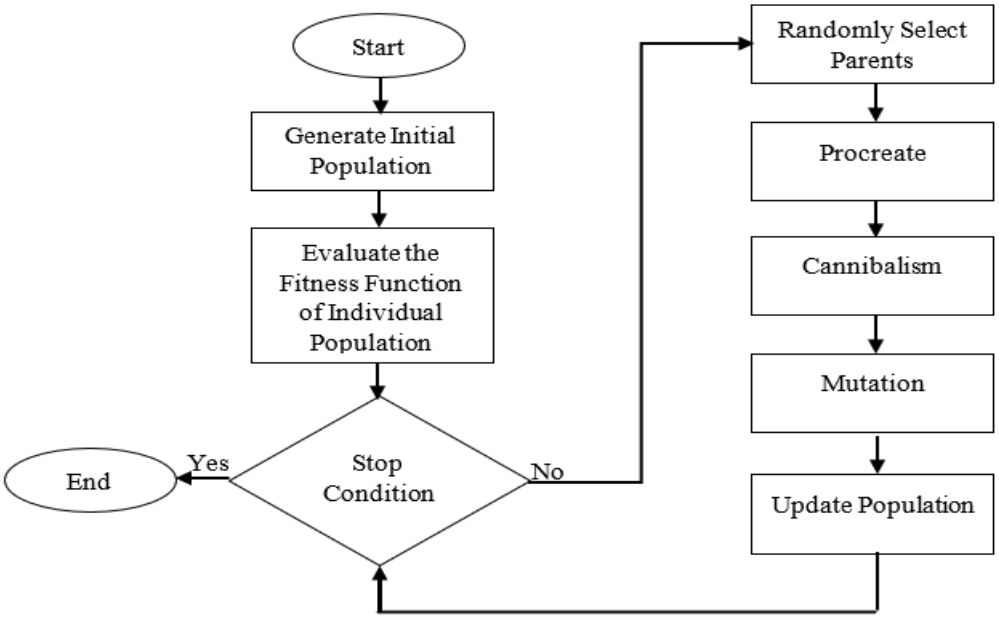


Figure1 Flow Chart for the BWO Algorithm

- Initial Population: It is necessary for the numbers of the problem variables to create an adequate structure in order to address the present issue while working on an optimization problem. Black widow optimization method (BWO) uses the word "widow" instead of "chromosome" or "particle position" to describe this structure. The possible answer to each issue has been modelled like a Black widow spider in the Black widow Optimization Algorithm (BWO). The values of the issue variables are shown on the backs of each Black widow spider. The structure should be viewed as an array for the purposes of this article's benchmark routines. A solution to an N_{var}-dimensional optimization issue is represented by a $1 \times N_{var}$ array called a widow. The following is the definition of this array:

$$Widow = [x_1, x_2 \dots\dots\dots\dots, xN_{var}] \quad (1)$$

Here, the value of variables $x_1, x_2 \dots\dots\dots\dots, xN_{var}$ are floating point numbers. The calculation of fitness function can be described as:

$$Fitness = f(Widow) = f(x_1, x_2 \dots\dots\dots\dots, xN_{var}) \quad (2)$$

Using an initial population of spiders, a potential widow matrix of size $N_{pop} \times N_{var}$ is created. After mating, the female black widow eats the black male widow, which is unpredictable.

- Procreate: While in nature, each couple mates in its own web, the genders of each pair begin mating simultaneously in order to generate the next generation as a whole. Each pairing produces roughly 1000 eggs, but only a few of the spider offspring survive, that are stronger than the others. After creating an array named alpha and filling it with random numbers from the widow array, the below equation (eqn1) is used to generate offspring, resulting in the generation of two new individuals, y_1 and y_2. Here x_1 and x_2 represents the parents of y_1 and y_2.

$$y_1 = alpha \times x_1 + (1 - alpha) \times x_2 \quad (3)$$

$$y_2 = alpha \times x_2 + (1 - alpha) \times x_1 \quad (4)$$

This method is repeated for a total of Nvar/2 times, with the exception that the numbers chosen at random must not be replicated. Finally, the offspring and mother are placed to an array and ordered according to their fitness value; then, as per the cannibalism grade, a few of the best people are introduced to the recently formed population to round out the population. These procedures are applicable to all couples.

- Cannibalism: There are three types of cannibalism. After mating, the black widow consumes her partner. This is the earliest kind of cannibalism. We were able to distinguish between male and females using their fitness scores in this method. Alternatively, the strongest spiderlings consume their lesser siblings in a kind of sibling cannibalism. A cannibalism rating (CR) is used in this algorithm to assess how many survivors there will be. Third-generation cannibalism, in which newborn spiders devour their mother, is common in certain circumstances. Spiderlings are judged on their fitness value.
- Mutation: In this step, randomly alter one population to develop a new solution. Whatever determines how many mutations are carried out is known as the mutation rate.

3.1.1 How it is Better than Other Optimization Algorithms

Within the process of producing a new set of generation of the algorithm of the BWOa large number of offspring is equal to the Nvar/2 while that of the GA, only for the two sets of descendants which are being generated [18]. A much higher range of the number of offsprings, which is of much more chance for discovering a much larger amount of the space of the search that will ensure that inorder to obtain a much higher range of performances for exploring all of the stages and that of the algorithm of the BWO and will be able in order to escape from that of the local set of optimisation of the problem. Moreover, the operator of the cannibalism which provides all of the ability in order to eliminate an improper amount of solutions on an immediate basis. Consequently, all of the next set of generation that will be needed to be reproduced through better parents guarantees for a much faster convergence of the solutions which needs to be nearby through the optimal solutions. All of the features are being known as per the exploitation. The algorithm of the NWO enables to provide a proper set of balance in between the stages of the exploitation and that of the exploration that is one of the most critical sets of features for the metaheuristics set of algorithms. In considerations for the arguments above , the BWO is able to obtain within the outstanding set of results in order to compare with other sets of experimental algorithms, especially in comparison to that of the GA.

3.2 LSB Method

In this method, initially, the cover image is read. Further, the least significant bit of the cover image pixel is replaced with a data bit that gives the stego image in the output, as shown in Figure 2 [22]. If the secret data bit is not the same as the LSB bit of the cover image then variability is generated.

10001000	00011110	10000000	10101001	01010101	10001010	10001000	11100001

(a) Cover Image Pixels

0	1	1	0	1	0	1	1

(b) Secret Data Bits

1000100**0**	0001111**1**	1000000**1**	1010100**0**	0101010**1**	1000101**0**	1000100**1**	1110000**1**

(c) Stego Image Pixels

Figure 2 LSB Method (a) Cover Image Pixels (b) Secret Data Bits (c) Stego Image Pixels

4. Proposed Method

A privacy-preserving method is designed for public health surveillance data using image steganography. The proposed method provides better imperceptibility and security as compared to the existing methods. The flowchart of the proposed method is shown in Figure 3. Initially, secret data is read and split into sensitive and non-sensitive data. The sensitive data is encrypted for security purposes. The encryption is done by performing the XOR operation between sensitive data and key generated using a black widow optimization (BWO) algorithm. The BWO algorithm generates a completely random key based on the objective function. We have taken entropy as an objective function and it is calculated using Eq. (5).

$$E = \sum_{i=1}^{n} -p_i \log p_i \quad (5)$$

After that, encrypted sensitive data and non-sensitive data are concatenated. Next, the cover image is read. After that, the superior, intermediate, and inferior planes of the cover image are determined based on the pixel intensity. Therefore, the pixel intensity of the RGB plane of the cover image is determined using Eq. (6-8) [26].

$$Sum(P_{red}) = \sum_{m=1}^{A} \sum_{n=1}^{B} P_{red}(m,n) \quad (6)$$

$$Sum(P_{green}) = \sum_{m=1}^{A} \sum_{n=1}^{B} P_{green}(m,n) \quad (7)$$

$$Sum(P_{Blue}) = \sum_{m=1}^{A} \sum_{n=1}^{B} P_{Blue}(m,n) \quad (8)$$

Whereas, $P_{red}, P_{green}, and\ P_{blue}$ represents the pixel values of the RGB plane. AB denotes the size of the images.

Based on the sum value of Eq.(6-8), superior, inferior, and intermediate planes are classified. The inferior plane contributes the minimum information of color in the cover image. Thus, the cover image is less impacted by a smaller variability in the inferior plane pixels that generated due to data hiding. Thus, the inferior plane is chosen for data hiding. Further, the inferior plane along with compressed data is given to the optimized data hiding using the BWO algorithm. The BWO algorithm searches the optimal starting pixel and secret data order in the inferior plane for data hiding to reduce the variability. After searching the optimal starting pixel and data order, data hiding is done using the k-bits LSB method in the inferior plane that gives the inferior stego plane in the output. Stego image is reconstructed by concatenating the inferior stego plane with the superior and intermediate planes of the cover image. In the last, the performance analysis of the proposed method is done to evaluate its performance over the existing methods.

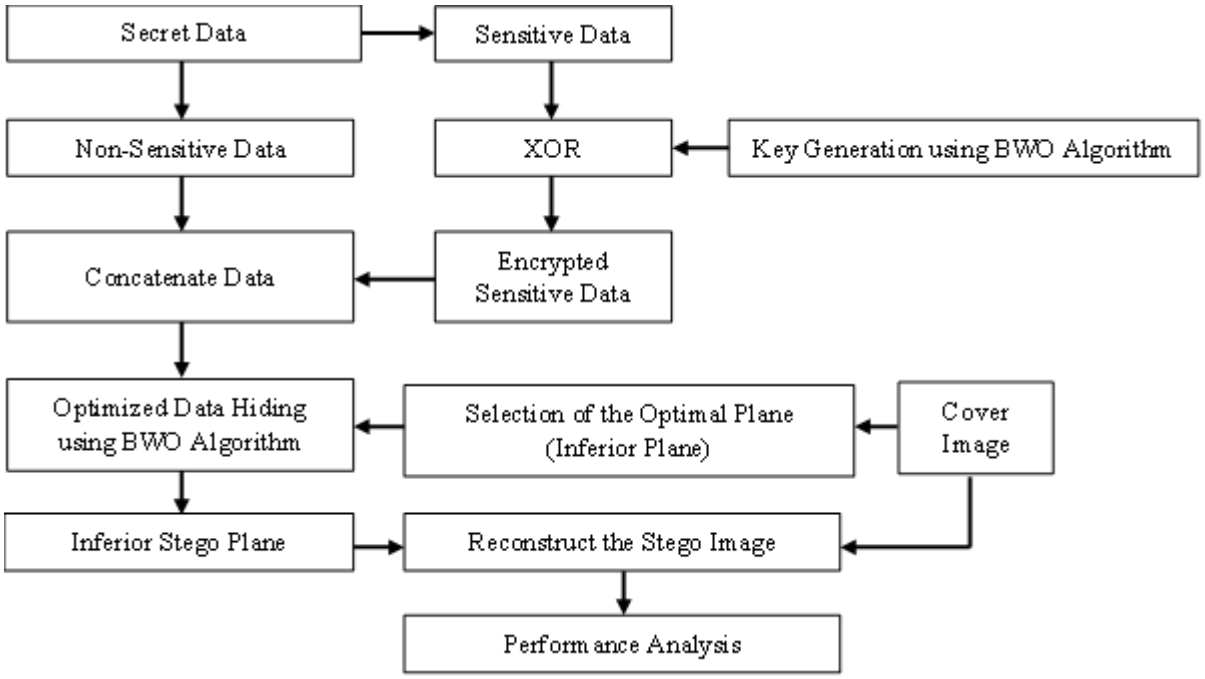


Figure 3 Flowchart of the Proposed Method

4.1 Key Generation using Black Widow Optimization Algorithm

The following steps are taken for key generation using the BWO algorithm.

1. Initially, define the size of the key, objective function, number of iterations, number of black widows, procreate, cannibalism, and mutation rate.
2. Next, initialize the black widow population matrix ($B \times D$). B represents the total number of black widows and D represents the dimension of each black widow.
3. After that, the fitness evaluation of each black widow is done based on the objective function to determine the initial best key.
4. Further, randomly $B/2$ black widows are selected as parents to generate offspring using procreate steps. Next, the fitness function is evaluated for generated offspring. Further, according to the fitness function, inadequate optimal solutions are removed using the cannibalism step and the population matrix is updated.
5. The mutation step is applied to the population matrix to generate a new solution after performing the procreate and cannibalism steps.
6. 3-5 steps are repeated until the stopping condition is not met to determine the optimal key.

4.2 Data Hiding using Black Widow Optimization Algorithm

The following steps are taken for optimal data hiding in the cover image using the BWO algorithm.

1. Initially, the cover image and secret data matrix read ($M \times N$) and transformed into a vector ($1 \times K$). The value of the K equal to $M \times N$.
2. Next, initialize the black widow population matrix ($B \times D$). B represents the total number of black widows and D represents the dimension of each black widow. In the proposed method, the value of D is 2. The first value represents the optimal starting pixel and the second value represents the secret data order. The possible secret data order value and its description are shown in Table 1.

Table 1 Secret Data Order Value and its Description

Secret Data Order Value	Description
0	No Operation on the Secret Data
1	Circular Shifting
2	Flipping the Secret Data
3	Reverse the Secret Data

3. The fitness function is evaluated for each black widow using the MSE parameter and the initial optimal solution is determined.
4. After that, randomly $B/2$ black widows are selected as parents to generate offspring using procreate steps. Next, the fitness function is evaluated for generated offspring. Further, according to the fitness function, inadequate optimal solutions are removed using the cannibalism step and the population matrix is updated.
5. The mutation step is applied to the population matrix to generate a new solution after performing the procreate and cannibalism step.
6. 3-5 steps are repeated until the stopping condition is not met to determine the optimal solution.
7. In the step, based on the optimal solution, the secret data is embedded in the cover image to generate a stego image.
8. In the last, the stego image vector $(1 \times K)$ is transformed into a matrix $(M \times N)$ to generate a stego image matrix. The stego image along with an optimal solution (optimal starting point and secret data order) information is communicated to the receiver.

4.3 Data Extraction

On the receiver side, the stego image and optimal solution information are read and the following steps are performed to recover the secret data.

1. The stego image matrix $(M \times N)$ is read and transformed into a vector $(1 \times K)$.
2. The secret data bits are extracted from the optimal starting pixel in the vector.
3. According to the secret data order information, the secret data bits are transformed into their original form.

5. Simulation Results

This section presents the simulation results for the proposed technique and compares with existing optimized data hiding techniques. The standard dataset colour images that include *Lena, Barbara, Baboon, Pepper, Female, Couple, Airplane, Rice, Cameraman, and Boat* are taken and employed for the proposed technique [27]. The cover size is 512×512 and format .jpg is used. The proposed algorithm is implemented on Intel (R) core (TM) i3-4005U CPU, 1.70GHz, with 8GB RAM, with 64-bit Windows 7 operating system. The code is written in MATLAB 2013a. Table 2 shows the parameter values for the proposed method.

Table 2 Parameter Values for the Proposed Method

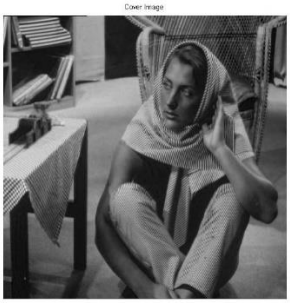
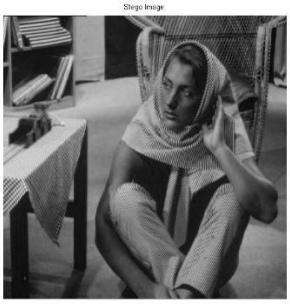
Sr No.	Parameter	Value
For Key Generation		
1	Objective Function for Key	Entropy
2	Key Size (in Byte)	64
3	Lower Limit	0
4	Upper Limit	255
5	Iterations	50

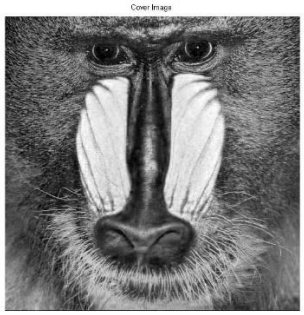
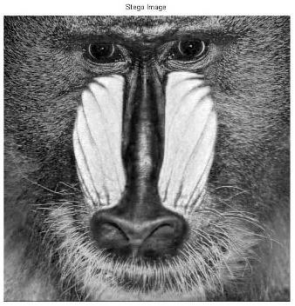
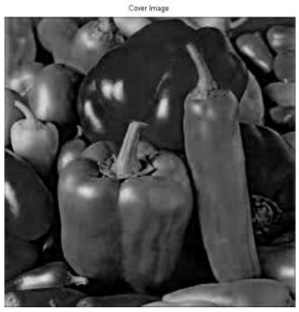
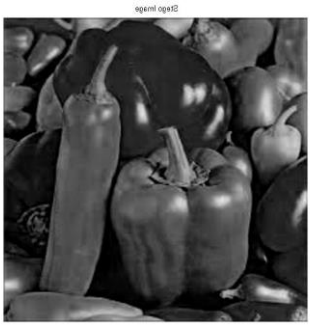
For Optimized Data Hiding		
1	Objective Function for Data Hiding	MSE
2	Iterations	20
3	Total Population (pop)	50
4	Procreate Rate	Pop/2
5	Cannibalism Rate	Pop/2
6	Mutation Rate	50%
7	Cover Image	512X512
8	Secret Data (in bits)	512X512
9	Data Hiding Method	LSB

5.1 Qualitative Analysis

The proper set of analysis which is being based upon the visual set of cover quality and that of the stego of images that are being compared. All of the analysis of the qualitative is being proposed for the methods which are being shown in Table 3. The results show that the images look similar.

Table 3 Qualitative Analysis

Images	Cover Image	Stego Image
Lena		
Barbara		

Baboon		
Pepper		
Female		
Couple		

Airplane		
Cameraman		
Boat		

5.2 Performance Metrics

The performance analysis of the proposed technique is evaluated on different metrics that are used for image steganography. These metrics are defined below.

5.2.1 Peak Signal to Noise Ratio (PSNR)

All of the PSNR function is being measured through the quality of the stego media after that of the data embedding. All of the PSNR is basically within the ratio of the maximum amount of power of the signal and that of the power of noise that affects the quality of its representation [28]. It is calculated as follow in Eq. (9)

$$PSNR = 10\frac{Peak^2}{MSE} \qquad (9)$$

where, $Peak^2$ denote the maximum power of the signal, MSE denote MSE function measures the commutative square error difference between the cover matrix and the residual matrix which is obtained from the cover image after data hiding at the suitable position (Wang et al. 2012). It is determined as Eq. (10)

$$MSE = \sum_{i=0}^{A-1}\sum_{j=0}^{B-1}(C_{i,j} - R_{i,j})^2 \qquad (10)$$

whereas, AB defined the size of the matrix $C_{i,j}$,$R_{i,j}$ is the cover and residual matrix.

The MSE and PSNR results for our proposed data hiding technique are presented in Table 4. It is observed that the PSNR varies from 54.1579dB to 54.3132dB for the different cover images. Table 4 shows that ***pepper*** Image has achieved the highest PSNR and ***boat*** the lowest. Due to this, the maximum optimal match is found for the ***pepper*** image as compared to the ***boat*** image for the secret data hiding.

Table 4 MSE and PSNR Analysis for the Proposed Method

Images	MSE	PSNR (in dB)
Lena	0.2482	54.1835
Barbara	0.2472	54.1998
Baboon	0.2482	54.1835
Pepper	0.2409	54.3132
Female	0.2458	54.2255
Couple	0.2488	54.1731
Airplane	0.2482	54.1831
Rice	0.2473	54.1982
Cameraman	0.2477	54.1922
Boat	0.2496	54.1579

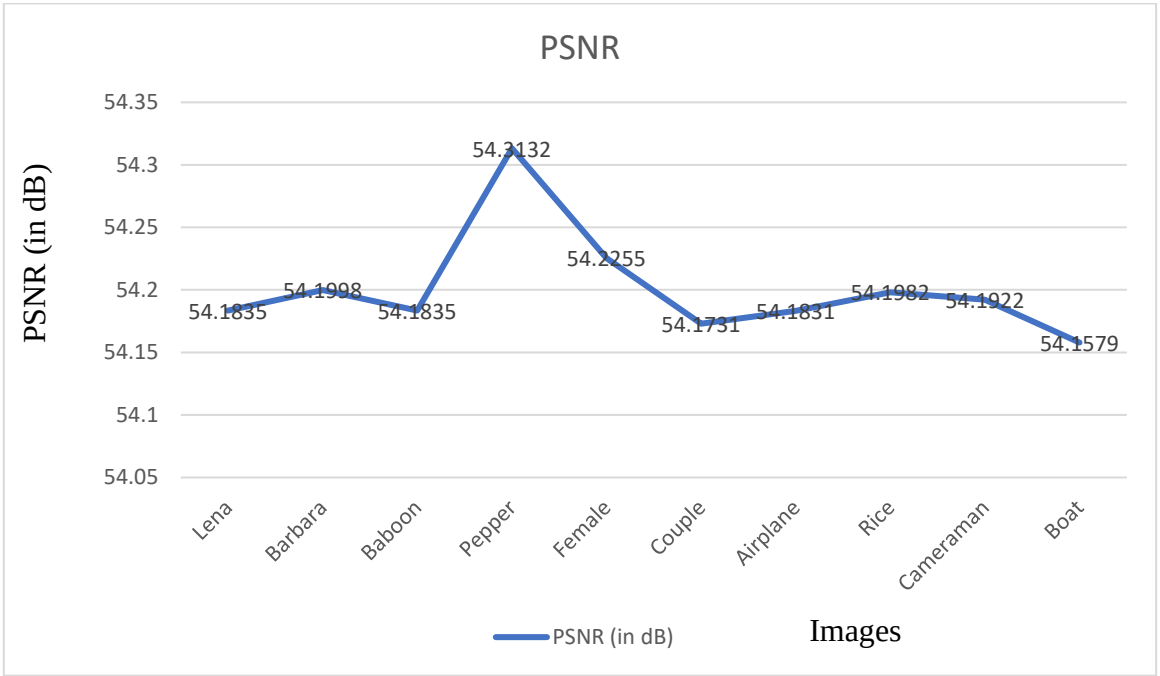


Figure 4 PSNR Analysis for Different Images

5.2.2 Correlation Factor

This parameter measures the correlation between cover and stego image [29,30]. It is calculated using Eq. (11)

$$r = \frac{\sum_{m=1}^{N}(X'_m-\mu_s)(X_m-\mu_c)}{\sqrt{\sum_{m=1}^{N}(X'_m-\mu_s)^2 \sum_{m=1}^{N}(X_m-\mu_c)^2}} \qquad (11)$$

where, r denotes the correlation factor X_m and X'_m cover and stego image pixel intensity μ_c and μ_s mean pixel values of cover and stego image and N is the total number of pixels.

The correlation factor for the proposed technique varies from 0.9997 to 0.9999 for the standard dataset images as shown in Table 5. From Figure 5, it can be observed that *Lena, Barbara, Baboon, Pepper, Airplane, Rice, Cameraman, and Boat* images have attained the highest PSNR and *couple* as lowest.

Table 5 Correlation Factor Analysis for the Different Images

Images	Correlation Factor
Lena	0.9999
Barbara	0.9999
Baboon	0.9999
Pepper	0.9999
Female	0.9998
Couple	0.9997
Airplane	0.9999
Rice	0.9999
Cameraman	0.9999
Boat	0.9999

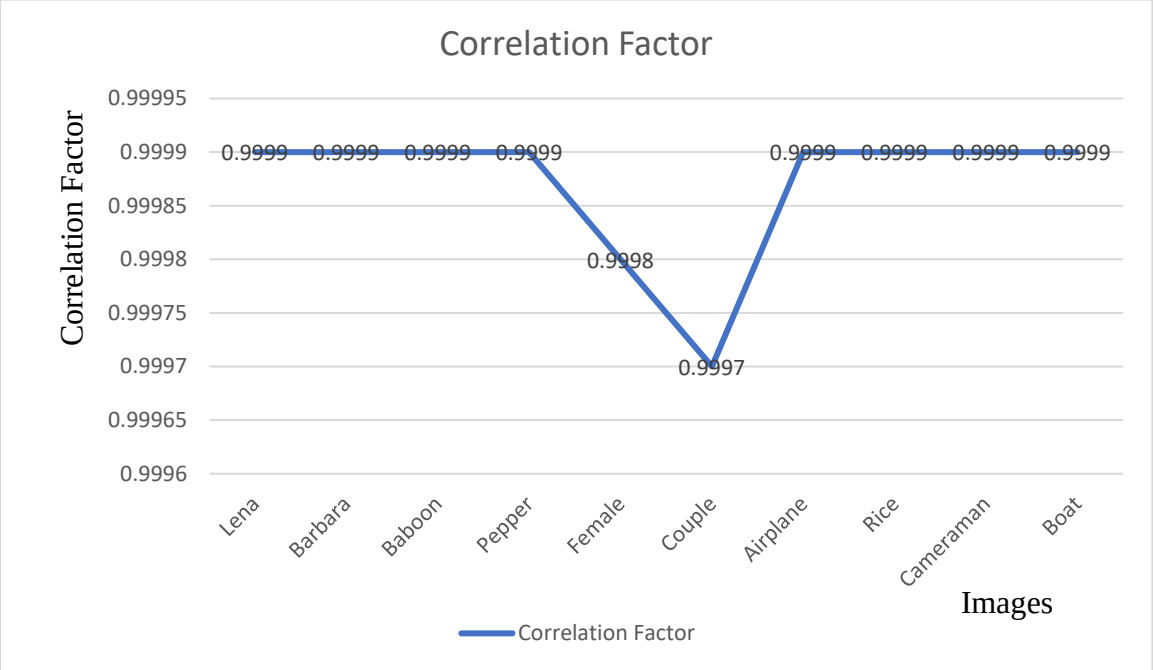


Figure 5 Correlation Factor for Different Cover Images

5.2.3 Embedding Capacity (EC)

Embedding capacity (EC) is defined as the amount of secret data bits concealed in the cover image [29]. It is commonly represented as bits per pixel (bpp) and calculated using Eq. (5).

$$EC(bpp) = \frac{Total\ number\ of\ secret\ bits\ embedded}{Total\ cover\ pixels\ in\ the\ image} \qquad (5)$$

In the proposed technique, 1 bpp embedding capacity has been achieved using the 1-bit LSB technique.

5.2.4 Selected Plane, Optimal Starting Pixel, and Secret Data Order for Data Hiding

In the proposed method, the pre-processing of the cover image is done to select the inferior plane for data hiding because the inferior plane contributes least to the cover image as compared to the superior and intermediate plane. On the other hand, optimal starting pixels and secret data order for data hiding is determined using black widow optimization algorithm to reduce the variability. Table 6 shows the selected plane, optimal starting pixel, and secret data order for the proposed method for the different sets of the cover images. The output shows that the selected plane, optimal starting pixel, and secret data is different for various images. Thus, it is difficult for an attacker to predict which plane contains information, secret data order, and optimal starting pixel. Thus, the proposed method enhances security.

Table 6 Selected Plane, Optimal Starting Pixel, and Secret Data Order for Proposed Method

Images	Selected Plane	Optimal Starting Pixel	Secret Data Order
Lena	2	69273	1
Barbara	3	38646	1
Baboon	3	177803	1
Pepper	3	245584	2
Female	3	65856	2
Couple	3	47136	2
Airplane	1	101501	2
Rice	1	82075	2
Cameraman	1	186668	1
Boat	1	75919	1

5.2.5 Comparative Analysis with the Existing Data Hiding Methods

In this section, the proposed technique is validated with the existing work based on the PSNR parameter results are comparison is made between them. In our work, we have compared with GA [21] and ACO [22] algorithms. For the purpose of comparison, same dataset images and secret images are preferred on which existing work has been reported on the black widow optimization algorithm. In Table 7 the results are provided. As can be observed that we have acquired better average PSNR for various standard dataset images. This shows that BWO algorithm finds the more optimal pixels and secret data order as compared to the existing algorithm in the pixels which reduces the mean square error between the cover and stego image.

Table 7 Comparative Analysis with the Existing Data Hiding Methods

Optimization Technique		**ABC**	**BWO**
Secret Image (128 X128)	Cover Image (512X512)	Aman Banharnsakun [22] PSNR (dB)	Proposed Technique PSNR (dB)
	Lena.jpg	56.40	57.20
	Jet.jpg	56.39	57.24
	Lake.jpg	56.40	57.22
	Elaine.jpg	56.36	57.21
	Baboon.jpg	56.39	57.22
	Average	**56.39**	**57.22**
Optimization Technique		**GA**	**BWO**

Secret Image (256X256)	Cover Image (512X512)	Kanan et al. [21] PSNR (dB)	Proposed Technique PSNR (dB)
	Lena.jpg	45.12	47.19
	Jet.jpg	45.18	47.17
	Pepper.jpg	45.13	47.36
	Sailboat.jpg	45.10	47.20
	Baboon	45.12	47.22
	Average	**45.13**	**47.23**

6. Conclusion and Future Scope

A privacy-preserving method is designed for public health surveillance data using image steganography. Within the proposed set of the method and that of the pre-processing's onto the cover image and that of the secret set of data is being done. The pre-processing on the cover image is done to select the most appropriate plane for data hiding whereas pre-processing on the secret data determines the sensitive and non-sensitive attributes of the data. After that, the algorithm optimization of the black widow is being deployed within the proposed set of methods for key generation and for optimized data hiding. The key generation is done to encrypt the sensitive attributes of the secret data. Further, optimized data hiding is done by determining the optimal starting pixel and secret data order. The simulation evaluation is done on standard dataset images. The output of the proposed method provides better security and PSNR as compared to the existing methods. In the future, data compression and that of the correction of the errors code algorithms are being hybridized with the proposed set of methods for enhancing the data hiding capacity and that of robustness against attacks.

7. Acknowledgement

The authors are grateful to MeitY for the financial support through the 'Visvesvarya Fellowship' and SMDP chips to system design' project. The authors also want to express their sincere gratitude towards the Director, Thapar Institute of Engineering and Technology, Patiala for his persistent support and encouragement.

References

[1] Nsubuga, P., White, M.E., Thacker, S.B., Anderson, M.A., Blount, S.B., Broome, C.V., Chiller, T.M., Espitia, V., Imtiaz, R., Sosin, D. and Stroup, D.F., 2006. Public health surveillance: a tool for targeting and monitoring interventions. *Disease Control Priorities in Developing Countries. 2nd edition.*

[2] Hussein, M.R., Apu, E.H., Shahabuddin, S., Shams, A.B. and Kabir, R., 2020. Overview of digital health surveillance system during COVID-19 pandemic: public health issues and misapprehensions. *arXiv preprint arXiv:2007.13633.*

[3] Cleveland Clinic (2020). *PCR Test for COVID-19: What it Is, How it's Done, What the Results Mean.* [online] Cleveland Clinic. Available at: https://my.clevelandclinic.org/health/diagnostics/21462-covid-19-and-pcr-testing.

[4] Nguyen, L., Stoové, M., Boyle, D., Callander, D., McManus, H., Asselin, J., Guy, R., Donovan, B., Hellard, M. and El-Hayek, C., 2020. Privacy-preserving record linkage of deidentified records within a public health surveillance system: evaluation study. *Journal of Medical Internet Research*, *22*(6), p.e16757.

[5] Jin, H., Luo, Y., Li, P. and Mathew, J., 2019. A review of secure and privacy-preserving medical data sharing. *IEEE Access*, *7*, pp.61656-61669.

[6] Salomon, D., 2003. *Data privacy and security: encryption and information hiding*. Springer Science & Business Media.

[7]Devi, R.R. and Chamundeeswari, V.V., 2020. Triple DES: privacy preserving in big data healthcare. *International Journal of Parallel Programming*, *48*(3), pp.515-533.

[8] Ghayvat, H., Pandya, S.N., Bhattacharya, P., Zuhair, M., Rashid, M., Hakak, S. and Dev, K., 2021. CP-BHDCA: Blockchain-based Confidentiality-Privacy preserving Big Data scheme for healthcare clouds and applications. *IEEE Journal of Biomedical and Health Informatics*.

[9] Aono, Y., Hayashi, T., Wang, L. and Moriai, S., 2017. Privacy-preserving deep learning via additively homomorphic encryption. *IEEE Transactions on Information Forensics and Security*, *13*(5), pp.1333-1345.

[10] Shin, M., Hwang, C., Kim, J., Park, J., Bennis, M. and Kim, S.L., 2020. Xor mixup: Privacy-preserving data augmentation for one-shot federated learning. *arXiv preprint arXiv:2006.05148*.

[11] Rukhin, A., Soto, J., Nechvatal, J., Smid, M., Barker, E., Leigh, S., Levenson, M., Vangel, M., Banks, D., Heckert, A., Dray, J. and Vo, S. (n.d.). *Special Publication 800-22 Revision 1a A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*. [online] Available at: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-22r1a.pdf.

[12] Chandravathi, D. and Lakshmi, P.V., Privacy Preserving Using Extended Euclidean Algorithm Applied To RSA-Homomorphic Encryption Technique.

[13] Ghayvat, H., Pandya, S.N., Bhattacharya, P., Zuhair, M., Rashid, M., Hakak, S. and Dev, K., 2021. CP-BHDCA: Blockchain-based Confidentiality-Privacy preserving Big Data scheme for healthcare clouds and applications. *IEEE Journal of Biomedical and Health Informatics*.

[14] Tuncer, T. and Avaroğlu, E., 2017, May. Random number generation with LFSR based stream cipher algorithms. In *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (pp. 171-175). IEEE.

[15] NK, S. and Pai, G.V., 2009. Design of stream cipher for text encryption using Particle Swarm Optimization based key generation. *Journal of Information Assurance and Security*, pp.30-41.

[16] Sreelaja, N.K. and Pai, G.V., 2008, January. Swarm intelligence based key generation for text encryption in cellular networks. In *2008 3rd International Conference on Communication Systems Software and Middleware and Workshops (COMPARE 08)* (pp. 622-629). IEEE.

[17] Krishna, G.J., Ravi, V. and Bhattu, S.N., 2018. Key generation for plain text in stream cipher via bi-objective evolutionary computing. *Applied Soft Computing*, *70*, pp.301-317.

[18] Sajedi, H. and Yaghobi, S.R., 2020. Information hiding methods for E-Healthcare. *Smart health*, *15*, p.100104.

[19] Chang, C.C., Hsiao, J.Y. and Chan, C.S., 2003. Finding optimal least-significant-bit substitution in image hiding by dynamic programming strategy. *Pattern Recognition*, *36*(7), pp.1583-1595.

[20] Sahu, A.K., Swain, G. and Babu, E.S., 2018. Digital image steganography using bit flipping. *Cybernetics and Information Technologies*, *18*(1), pp.69-80.

[21] Kanan, H.R. and Nazeri, B., 2014. A novel image steganography scheme with high embedding capacity and tunable visual image quality based on a genetic algorithm. *Expert systems with applications*, *41*(14), pp.6123-6130.

[22] Banharn Sakun, A., 2018. Artificial bee colony approach for enhancing LSB based image steganography. *Multimedia Tools and Applications*, *77*(20), pp.27491-27504.

[23] Wazirali, R., Alasmary, W., Mahmoud, M.M. and Alhindi, A., 2019. An Optimized Steganography Hiding Capacity and Imperceptibly Using Genetic Algorithms. *IEEE Access*, *7*, pp.133496-133508.

[24] Bedi, P., Bansal, R. and Sehgal, P., 2013. Using PSO in a spatial domain based image hiding scheme with distortion tolerance. *Computers & Electrical Engineering*, *39*(2), pp.640-654.

[25] Hayyolalam, V. and Kazem, A.A.P., 2020. Black widow optimization algorithm: a novel meta-heuristic approach for solving engineering optimization problems. *Engineering Applications of Artificial Intelligence*, *87*, p.103249.

[26] Azmi, K.Z.M., Ghani, A.S.A., Yusof, Z.M. and Ibrahim, Z., 2019. Natural-based underwater image color enhancement through fusion of swarm-intelligence algorithms. *Applied Soft Computing*, *85*, p.105810.

[27] The USC-SIPI Image Database, 1977. http://sipi.usc.edu/database/database.php?volume=misc&image=10#top. Accessed 25 January 2019

[28] Kordov, K. and Stoyanov, B., 2017. Least significant bit steganography using Hitzl-Zele chaotic map. *International Journal of Electronics and Telecommunications*, *63*.

[29] Kadhim, I.J., Premaratne, P., Vial, P.J. and Halloran, B., 2019. Comprehensive survey of image steganography: Techniques, Evaluations, and trends in future research. *Neurocomputing*, *335*, pp.299-326.

[30] Subhedar, M.S. and Mankar, V.H., 2014. Current status and key issues in image steganography: A survey. *Computer science review*, *13*, pp.95-113.