Multimed Tools Appl (2019) 78:1473–1494
https://doi.org/10.1007/s11042-018-6155-6

# Feature selection for image steganalysis using levy flight-based grey wolf optimization

**Yadunath Pathak[1] · K. V. Arya[1] · Shailendra Tiwari[2]**

Received: 24 August 2017 / Revised: 7 March 2018 / Accepted: 16 May 2018 /
Published online: 25 June 2018


**Abstract** Image steganalysis is the process of detecting the availability of hidden messages in the cover images. Therefore, it may be considered as a classification problem which categorizes an image either into a cover images or a stego image. Feature selection is one of the important phases of image steganalysis which can increase its computational efficiency and performance. In this paper, a novel levy flight-based grey wolf optimization has been introduced which is used to select the prominent features for steganalysis algorithm from a set of original features. For the same, SPAM and AlexNet have been used to generate the high dimensional features. Furthermore, the random forest classifier is used to classify the images over selected features into cover images and stego images. The experimental results show that the proposed levy flight-based grey wolf optimization shows preferable convergence precision and effectively reduces the irrelevant and redundant features while maintaining the high classification accuracy as compared to other feature selection methods.

**Keywords** Image steganalysis · Feature selection · Grey wolf optimization · Swarm intelligence

✉ Yadunath Pathak
yadunath@iiitm.ac.in

K.V. Arya
kvarya@iiitm.ac.in

Shailendra Tiwari
shailendra@thapar.edu

1 Multimedia and Information Security Lab, Atal Bihari Vajpayee Indian Institute of Information Technology and Management, Gwalior, India

2 Thapar Institute of Engineering and Technology, Patiala, India

## 1 Introduction

Steganography is the process of concealing information into a media like images, videos and other forms in such a way that the hidden message is not noticeable. On the other hand, steganalysis algorithms try to reveal the presence of the secret messages. Generally, images are considered more appropriate for steganographic carrier as human cannot easily percept the minor changes in the images. A number of algorithms have been proposed for steganography on images, such as prediction-based reversible steganographic scheme using image inpainting [37] and data hiding scheme with reversibility based on exploiting modification direction [36]. A steganographic image, having hidden message, is known as stego-image while clean images are termed as cover-images. Thus, steganalysis algorithms may be considered as a classification problem in which an image is categorized into either cover image or stego image. Although, many steganalysis methods have been proposed in the literature, one of the most important concern in information forensics is the improvement of steganalysis performance.

Steganalysis methods are broadly categorized into steganalysis for specific embedding and universal blind steganalysis [27]. Methods of steganalysis for specific embedding reveal the hidden messages for known steganographic methods. For example, regular singular (RS) [15], sample pair analysis (SPA) [12], difference image histogram (DIH) [46], and least square method (LSM) [26] detect the spatial LSB steganography. Further, Pevny et al. [34] proposed a method to find the existence of secret message in a discrete cosine transform (DCT) domain image. As it is hard to know the steganography method used, steganalysis for specific embedding are rarely preferred. In the contrary, universal blind steganalysis methods can reveal the secret messages without having the knowledge of steganography method and are practically used in information forensics.

Universal blind steganalysis is generally considered as binary classification problem which is a two step process. First step extracts the features from the input images and second classifies these images into cover and stego images based on the absence or presence of hidden messages respectively [32]. Support vector machine (SVM) [9], neural network [42], k-nearest neighbor (kNN) [17], fisher linear discriminant (FLD) [38], random forest (RF) [22] are some of the popular classification methods used in universal blind steganalysis methods to classify the input images into cover or stego images. The performance of a classification method is highly dependent on the extraction of high dimensional significant features from different domains. The feature extraction methods used for image steganalysis are subtractive pixel adjacency model (SPAM) [33] in spatial domain, Markov features using intra- and inter-block dependencies (CHEN) [4], CHEN features enhanced by Cartesian calibration (CC-CHEN) [20], the union of diff-abs NJ-ratio and ref-diff-abs NJ features (LIU) [25], Cartesian-calibrated PEV feature set (CC-PEV) [35], the high-dimensional feature space (CC-C300) [21], compact rich model for DCT domain (CFn) [22] and many more. However, "curse of dimensionality" [8] is the major issue in large dimensional features space. With the increase of the feature dimensions, exponentially order of training data is required which causes high computational complexity of the classifier. Further, there are the chances of irrelevant and redundant features which may degrade the performance of a classifier. Hence, there is a requirement of an efficient feature selection method to overcome this problem [2].

A feature selection method uses an evaluation measure to find an optimal or suboptimal feature subset. There are various search methods for the same. An exhaustive search method compares $2^N$ feature subsets with a complexity of $O(2^N)$ for $N$ dimensional feature space which becomes impractical for large $N$ [23]. Therefore, to solve this problem, many feature selection methods have been developed and categorized into filters, wrappers, and embedded methods [10]. Filter methods are computationally efficient as set of features are considered as class variables. However, for a given classifier, it may perform poorly [40]. On the other hand, wrapper methods evaluate the feature subset using predictive models and are more common than filter methods [10]. The most commonly used wrapper method is sequential backward selection (SBS) [7] which uses greedy hill-climbing search method to iteratively eliminate the least promising features one by one till the performance of learning model drops below a given threshold. The embedded methods, such as support vector machine with recursive feature elimination (SVM-RFE) [18], use the information obtained from a supervised learner to select the features. In SVM-RFE, features having lowest weight obtained from a trained SVM are eliminated. However, both the wrapper and embedded methods are computationally intensive methods [40].

Recently to overcome this problem, meta-heuristic algorithms have widely been employed by the researchers. Meta-heuristic algorithms use the natural phenomena to solve complex optimization problems [41] and are widely used for feature selection with steganalysis algorithms. Particle swarm optimization (PSO) [5], artificial bee colony (ABC) [30], firefly algorithm (FA) [6], gravitational search algorithm (GSA) [39], fast evolutionary programing (FEP) [45], grey wolf optimization (GWO) [13] are some of the meta-heuristic algorithms used for feature selection in image steganalysis.

GWO [29] is a recently proposed meta-heuristic algorithm based on the concept of grey wolf society. In comparison with popular nature inspired algorithms, GWO has performed better in searching the solution of nonlinear functions in multidimensional space. GWO uses a collective behavior of wolves for finding the optimal solution. It starts with the exploration of search space and exploits gradually, using the grey wolf hunting procedure. To manage the trade-off between exploration and exploitation, GWO uses a parameter known as **A**. **A** is a linearly decreasing function that defines the step size of wolves at any iteration. Although, GWO has advantage of low computational cost, it still has some drawbacks like slow convergence rate and traps in local optima at times. Therefore, to enhance the performance of GWO, many variants have been proposed in literature such as improved Grey Wolf Optimizer (IGWO) [31], binary grey wolf optimization (bGWO) [14] and many more. However, it has been observed that the performance of GWO can further be improved by controlling it's exploration and exploitation trade-off defined by **A**.

Therefore, this paper proposes a novel variant of GWO, levy flight-based grey wolf optimization (LFGWO), where **A** has been modified which shows preferable convergence precision, quick convergence rate, and better global search ability. Further, the proposed variant, LFGWO, has been applied in feature selection method used for image steganalysis. AlexNet, a convolution neural network (CNN), and SPAM are two different methods which are used to extract the features from the input images. The extracted features are fed to the proposed feature selection method. The selected features are fed to different classifiers to classify the images into cover and stego images. The performance of the proposed method has been compared with existing popular meta-heuristic-based feature selection methods namely; PSO, GSA, FEP, and GWO. Moreover, results of image steganalysis have been analyzed on BOSSbase ver 1.01 data set using support vector machine (SVM), linear discriminant analysis (LDA), random forest (RF), k-nearest neighbor (kNN), and ZeroR classifiers.

The main contributions of this paper are as follows:

(1) A novel levy flight-based grey wolf optimization (LFGWO) method has been proposed.

(2) AlexNet, a convolution neural network (CNN), and SPAM are used to extract the features from the input images.
(3) A novel feature subset selection method based on LFGWO has been introduced to select the prominent features.
(4) For discriminating cover images from stego images in BOSSbase ver 1.01 dataset, SVM, LDA, RF, kNN, and ZeroR classifiers are used and compared.

The rest of the paper is organized as follows. Section 2 briefly introduces CNN and GWO. Section 3 describes the different phases of image steganalysis method i.e. feature extraction, feature selection method along the novel LFGWO, and classification. Experimental results of applying LFGWO on standard benchmark functions and the proposed staganalysis method on BOSSbase ver 1.01 dataset along with statistical analyses and comparison with other metaheuristics are included in Section 4. Finally the paper is concluded in Section 5.

## 2 Prelimiaries

This section describes the convolution neural network (CNN) and basic grey wolf optimization (GWO) method to be used for feature extraction and feature selection method respectively.

### 2.1 Convolutional neural network

A convolutional neural network (CNN) [24] is in general contains a set of convolutional layers, followed by non-linear activation layer and pooling layer. For classification of images, CNN uses some fully-connected layers of neurons along with softmax layer at the end. However, in this paper CNN is used for feature extraction and hence does not use fully-connected layers and softmax layer. The three steps, convolution, non-linear activation, and pooling produce sets of arrays called feature maps. The convolution layer performs filtering by using $k$ kernals to generate $k$ new feature maps. After applying the convolution layer, nonlinear activation layer is applied to introduce the nonlinearity into the system. Generally, Rectified Linear Units (ReLU) [16] activation function is used for the same which is shown in Fig. 1. ReLU activation function has a value zero when $x < 0$ and then linear with slope 1 when $x > 0$. After some ReLU layers, pooling layer (max pooling) is used which is also known as down sampling layer. In max pooling layer, a filter (normally of size 2x2) is applied to the output of previous layer and returns the maximum value from the region that the filter convolves around.

Let $F^n(X)$ denote the output feature map of layer $n$, having $W^n$ kernel (filter) and $B^n$ bias. If $F^0(X)$ represents the input data (image) then $F^n(X)$ can be calculated by (1).

$$F^n(X) = pooling(f^n(F^{n-1}(X) * W^n + B^n)) \tag{1}$$

where $f^n$ is a non-linear activation function (ReLU). A CNN having multiple convolutional layers is trained to extract high-level features from the original data that are used as input to the feature selection method in steganalysis algorithm.

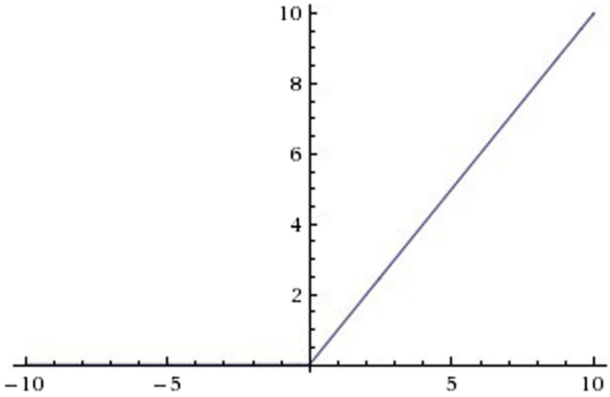



**Fig. 1** Rectified Linear Unit (ReLU) activation function [16]

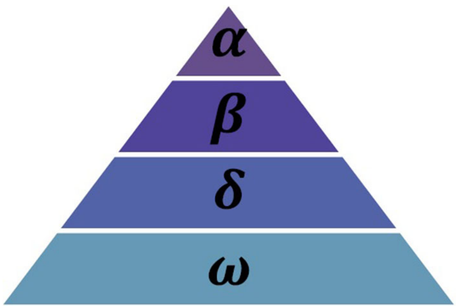
## 2.2 Grey wolf optimization (GWO)

Grey Wolf Optimization (GWO) [29] is the meta-heuristic algorithm based on the mathematical model of leadership hierarchy and hunting process of gray wolves. In general, grey wolves live in a group of size 5-12 and have a strict social dominant hierarchy. According to dominance, they are divided into four groups namely; alpha ($\alpha$), beta ($\beta$), delta ($\delta$), and omega ($\omega$) as shown in Fig. 2 [29].

Grey wolves who lead the group for hunting are termed as alpha. All other members of the group follow their orders. The next level in the hierarchy includes the Grey wolves who are subordinates to alpha and help them to take decisions. Delta is the subsequent level that contains the Grey wolves who follow $\alpha$ and $\beta$ groups and dominant the lowest level group known as omega. The delta group members may work as scouts, sentinels, elders, hunters, and caretakers. The omegas are submissive to $\alpha$, $\beta$, and $\delta$ wolves. Therefore, $\alpha$, $\beta$, and $\delta$ are the important wolves in the process of group hunting. Group hunting is another social activity of Grey wolves whose main phases include tracking, encircling, and attacking prey.

GWO mathematically modeled the social hierarchy and hunting procedure of Grey wolves for optimization problem as discussed in the following section.

**Fig. 2** Dominance hierarchy of grey wolf , increasing from bottom to top [29]



### *2.2.1 Social hierarchy*

Social hierarchy of grey wolves is mathematically modeled by considering the fittest, second fittest, and third fittest solution as alpha ($\alpha$), beta ($\beta$), and delta ($\delta$) respectively. Rest other solutions are termed as omega ($\omega$).

### *2.2.2 Encircling prey*

Following formulas are considered to mathematically model the encircling process [17]:

$$\mathbf{D} = \left|\mathbf{C}.\mathbf{X_p}(t) - \mathbf{X}(t)\right| \tag{2}$$

$$\mathbf{X}(t+1) = \mathbf{X_p}(t) - \mathbf{A}.\mathbf{D} \tag{3}$$

where, $\mathbf{X_p}(t)$ and $\mathbf{X}(t)$ are the position of prey and a Grey wolf at iteration $t$. $\mathbf{A}$ and $\mathbf{C}$ are coefficient vectors, calculated by (4) and (5) respectively.

$$\mathbf{A} = 2\mathbf{a}.\mathbf{r}_1 - \mathbf{a} \tag{4}$$

$$\mathbf{C} = 2.\mathbf{r}_2 \tag{5}$$

here $\mathbf{r_1}$ and $\mathbf{r_2}$ represent the random vectors in [0, 1]. $\mathbf{a}$ is a parameter which decreases linearly from 2 to 0 over the course of iterations. This parameter is used to control the step size ($\mathbf{D}$) of a Grey wolf.

### *2.2.3 Hunting*

In general, hunting is guided by alpha and supported by beta and delta. Therefore, to simulate this behavior, the first three best solutions ($\alpha$, $\beta$, $\delta$) in an iteration are used by other search agents (omegas) to update their positions by using following equations:

$$\overrightarrow{D}_\alpha = \left|\overrightarrow{C_1}.\overrightarrow{X_\alpha} - \overrightarrow{X}\right|, \quad \overrightarrow{D}_\beta = \left|\overrightarrow{C_2}.\overrightarrow{X_\beta} - \overrightarrow{X}\right|, \quad \overrightarrow{D}_\delta = \left|\overrightarrow{C_3}.\overrightarrow{X_\delta} - \overrightarrow{X}\right| \tag{6}$$

$$\overrightarrow{X_1} = \overrightarrow{X_\alpha} - \overrightarrow{A_1}.\left(\overrightarrow{D_\alpha}\right), \quad \overrightarrow{X_2} = \overrightarrow{X_\beta} - \overrightarrow{A_2}.\left(\overrightarrow{D_\beta}\right), \quad \overrightarrow{X_3} = \overrightarrow{X_\delta} - \overrightarrow{A_3}.\left(\overrightarrow{D_\delta}\right) \tag{7}$$

$$\overrightarrow{X}(t+1) = \frac{\overrightarrow{X_1} + \overrightarrow{X_2} + \overrightarrow{X_3}}{3} \tag{8}$$

### *2.2.4 Attacking prey (exploitation)*

The hunting by the Grey wolves is finished by attacking the prey and they stop moving. This phenomena is implemented by decreasing the value of $\mathbf{A}$ in (3) which depends on $\mathbf{a}$. As described in Section 2.2.2, the value of $\mathbf{a}$ decreases linearly from 2 to 0 and hence play an important role to exploit the search space whereas very little exploration which may lead to trap into local optimum.

### *2.2.5 Search for prey (exploration)*

To enhance the divergence behavior of GWO, the value of $\mathbf{A}$ is considered either greater than 1 or less than -1. This increases the exploration capability of GWO. Further, the value of $\mathbf{C}$ takes random values in all iterations which shows better exploration not only during initial iterations but also in final iterations. This escapes GWO from stagnation. The pseudocode of the GWO is presented in Algorithm 1 [29].

**Algorithm 1** Grey Wolf Optimization (GWO)

**Input:** $N$ search agents $X_i$ $(i = 1, 2, \ldots, n)$ having $n$ dimensions.
**Ouput:** Best solution $(X_\alpha)$.
Randomly initialize the initial population of $N$ search agents
Initialize the value of a, A, and C
Evaluate the fitness of each search agent
$X_\alpha$ = the best search agent
$X_\beta$ = the second best search agent
$X_\delta$ = the third best search agent
**while** stopping criteria is not satisfied **do**
  **for** i=1 to $N$ **do**
    Update the position of $X_i$ by (8)
  **end for**
  Modify a, A, and C
  Evaluate the fitness of all search agents
  Modify $X_\alpha$, $X_\beta$, and $X_\delta$
**end while**
return $X_\alpha$

## 3 Proposed approach

The proposed image stegnalysis method consists of three phases as shown in Fig. 3: (1) First phase extracts the features from the input images using either SPAM or AlexNet, (2) second phase reduces the high dimensional features by selecting the prominent features using the proposed novel feature selection method based on LFGWO, and (3) finally, the selected features are given to the classifier for classifying the images into cover and stego images. A detailed description of each phase has been presented in the following sections.

### 3.1 Feature extraction

Feature extraction is one of the most important phase of steganalysis. The accuracy of a steganalysis algorithm is highly dependent on the extracted features. For an effective stegnalysis algorithm, features must be multi-dimension and dissimilar for both the cover image and stego image. For last decade, many feature extraction methods have been introduced for steganalysis algorithms where each produces different number of features for the same set of images. In this paper, SPAM and AlexNet have been used to extract the feature from BOSSbase ver 1.01 dataset for steganalysis and described in following sections.

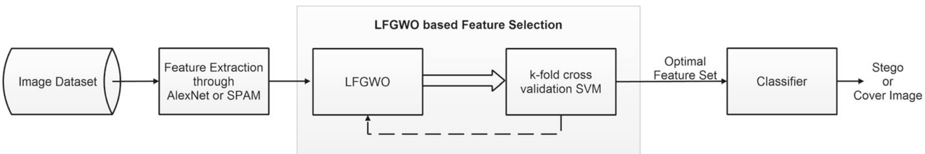



**Fig. 3** The proposed image stegnalysis method

### 3.1.1 Subtractive pixel adjacency matrix (SPAM)

Subtractive Pixel Adjacency Matrix (SPAM) [33] is the popular feature extraction method for spatial domain images. It has shown better performance for both the steganalysis for specific embedding and universal blind steganalysis. It is one of the most efficient methods for LSB matching steganalysis as compared to state-of-the-art. It is a variant of Markov chain features and utilizes the fact that the images generally do not contain the stego noise, which is an independent random component. On the other hand, there are short-range dependences among noise components within an image [33]. SPAM finds the local dependences between differences of neighboring cover elements and models them as a Markov chain to extract the feature vector for steganalysis. In this paper, SPAM extracts 686 features for BOSSbase ver 1.01 dataset.

### 3.1.2 AlexNet

The second method used for feature extraction in this paper is AlexNet [24] which is a pre-trained deep convolutional network proposed in 2012. This network was first successful convolutional neural network to win the 2012 ILSVRC (ImageNet Large-Scale Visual Recognition Challenge). AlexNet relies on the deep neural networks and extracts the features directly from raw image pixels without human intervention i.e., there is no need for handcrafted image descriptors. Therefore, it shows more semantic properties which are only based on the network construction. The architecture of AleNet is shown in Fig. 4 [24]. AlexNet has a simple layout having 5 convolution layers, max-pooling layers, dropout layers, and 3 fully connected layers. As visualized from Fig. 4, the architecture of AleNet is divided into two parts, each executing on a separate graphics processing unit (GPU) to achieve good overall performance. First layer is a convolution layer which uses 96 filters, each of size $11 \times 11 \times 3$. The output of first layer is given to second layer which performs max pooling to reduce the computation, followed by convolution with 256 filters of size $5 \times 5 \times 48$. The next three layers follow the similar operations. Layer 6 is a fully connected layer which maps the input matrix to a vector of size $1 \times 2048$. Similarly, layers 7 & 8 are also fully connected layers. The output of AlexNet is a feature vector of size $1 \times 1000$.

Since, this paper uses a pre-trained AlexNet for feature extraction and the features extracted after $8^{th}$ layer is used in staganalysis. Therefore, the extracted 1000 features, by using AlexNet, for BOSSbase ver 1.01 dataset are used for feature selection process.

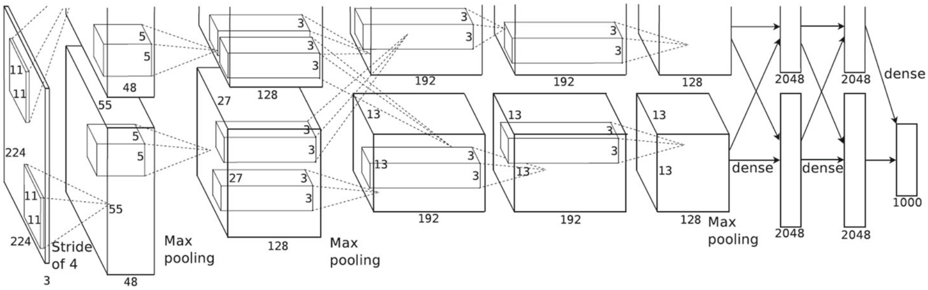



**Fig. 4** AlexNet architecture [24]

### 3.2 Feature selection

The extracted features from AlexNet or SPAM are high dimensional feature set containing 1000 and 686 features respectively which require high computational cost and may reduce the performance of a classifier. Therefore, these features are fed to the next phase i.e. feature selection to filter out the redundant and irrelevant features. The different stages of proposed feature selection method is depicted in Fig. 5.

The proposed feature selection method uses novel levy flight-based grey wolf optimizer (LFGWO) to reduce the irrelevant and redundant features. In the proposed LFGWO, firstly the initial positions of each grey wolf in the population is randomly initialized in between [0, 1]. The dimension of each individual is equal to the number of features extracted from

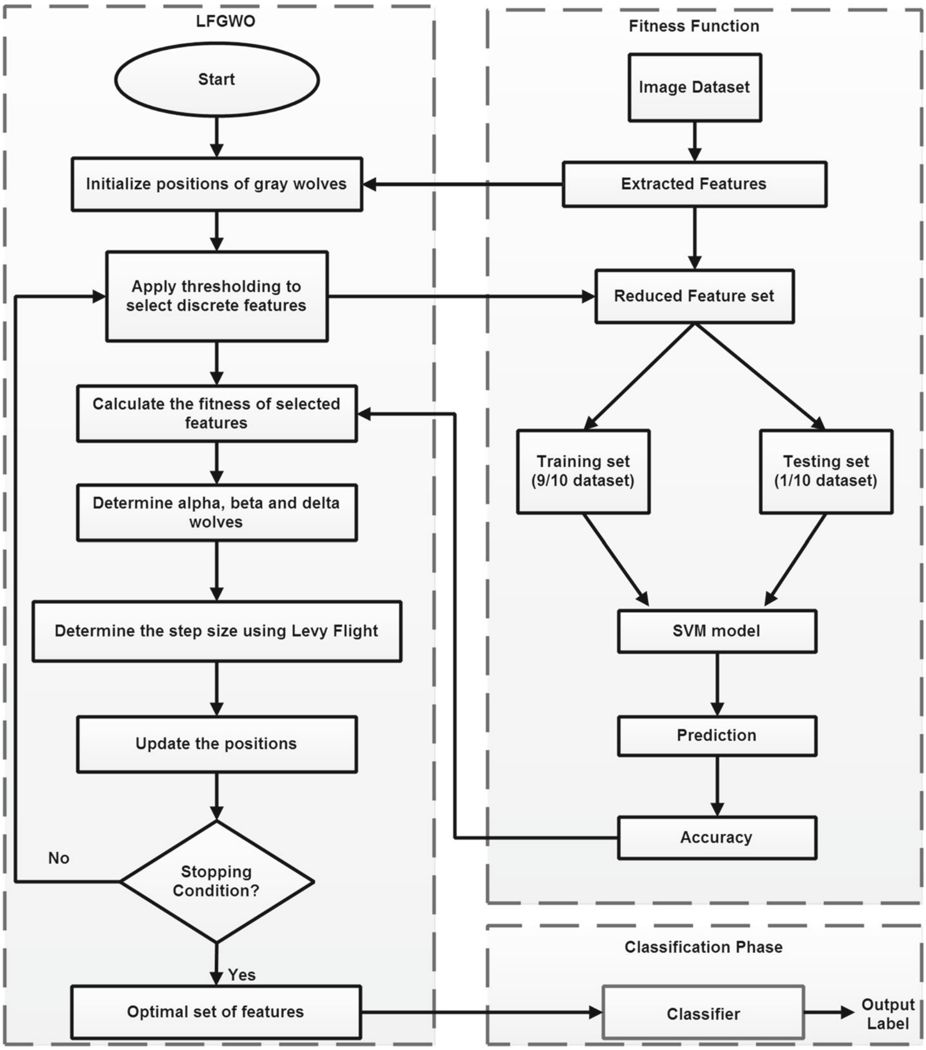



**Fig. 5** The proposed LFGWO based feature selection method

AlexNet or SPAM. Let there are $n$ extracted features and $N$ wolves then $i^{th}$ individual will be represented as (9).

$$X_i = \{x_1, x_2, \ldots, x_n\}; \quad i = 1, 2, \ldots, N \tag{9}$$

Each $x_i$ will have a random value in between 0 and 1 which is discretized to either 1 or 0 for calculating fitness. In this paper, the threshold value is selected to 0.9 which is set empirically. This can be visualized from Fig. 6 that for 0.9 threshold value, the fitness value of the system is the best. Therefore, the value of $x_i$ is set to 1 if its value is greater that or equal to 0.9 otherwise to 0. Thus, only those features are fed to calculate the fitness function value whose corresponding $x_i$ is 1. The fitness value is calculated using SVM through 10-fold cross validation. After computing the fitness value, the actual value of grey wolves along with their fitness values are given to subsequent phases of LFGWO which are explained in the following section.

### *3.2.1 Levy flight-based grey wolf optimization (LF-GWO)*

As exploration and exploitation are two important aspects of an optimization algorithm, their good trade-off helps in achieving precise solution by escaping the local optima. In GWO, the step size of a search agent linearly decreases with the iterations. **A** is the parameter that is used to control this step size. However, it has been sheen that at later iterations GWO traps into local optimum due to poor divergence. Therefore, in the proposed algorithm, the value of **A** is modified by using the levy flight which enhances the ability of the exploration and exploitation of GWO simultaneously.

Levy flight [3] uses the Levy probability distribution function which is function of power-law and used for calculating the jump size. The mathematical formula of the Levy distribution is shown in (10).

$$L(s, \gamma, \mu) = \begin{cases} \sqrt{\frac{\gamma}{2\pi}} \exp\left[-\frac{\gamma}{2(s-\mu)}\right] \frac{1}{(s-\mu)^{\frac{3}{2}}} & \text{if } 0 < \mu < \infty \\ 0 & \text{if } s \le 0 \end{cases} \tag{10}$$

where $\mu$, $\gamma$, and $s$ are the position parameter, scale parameter which controls the scale of distribution, and the collection of samples in this distribution respectively. An example of

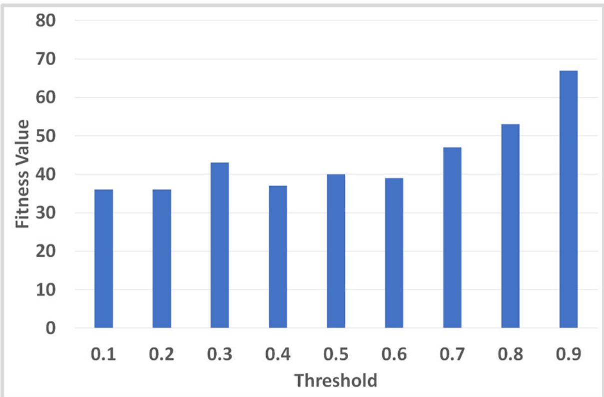



**Fig. 6** Chart showing fitness value vs threshold value

the Levy flight is shown in Fig. 7 [1]. As it could be observed, first movements are around a point, and then, it face sudden jumps. This factor causes to have local and global search simultaneously.

In the proposed method, the parameter **A** is modified by levy flight as shown in (11). This makes the parameter **A** to take its values according to levy flight distribution instead of linear decrease.

$$\mathbf{A} = LEVY(S) * \mathbf{r}_1 \tag{11}$$

where, $S$ is position of wolves and $\mathbf{r_1}$ is random vector. Levy flight-based **A** balances the abilities of local search and global search behavior simultaneously in GWO as depicted in Fig. 7. Rest of the steps of GWO remains same as original GWO as described in Algorithm 1.

### 3.3 Classification

After the selection of prominent features, the performance of the different classifiers is measured to classify the images into cover and stego images. For the same, support vector machine (SVM), linear discriminant analysis (LDA), random forest (RF), k-nearest neighbor (kNN), and ZeroR classifiers are used and analyzed.

## 4 Experimental results

This paper proposes a novel feature selection method based on LFGWO for image stegnalysis. Therefore, the results are depicted in two phases; First, the performance of proposed LFGWO method has been presented and second, feature selection method used in image steganalysis has been analyzed.

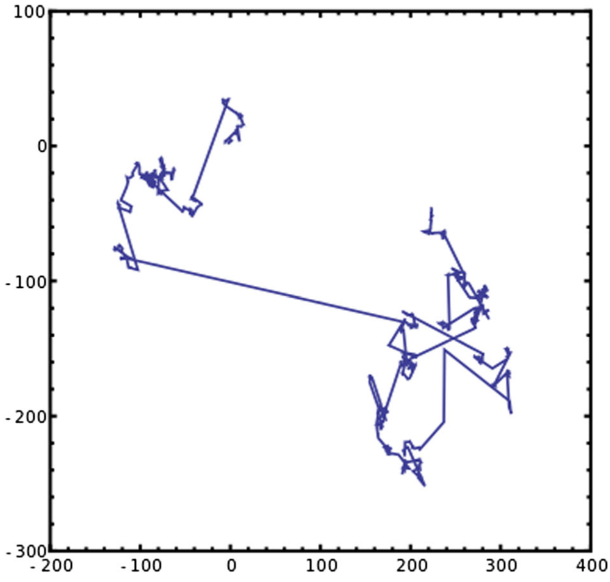



**Fig. 7** An example of a Levy flight for 1000 steps in two dimensions. The motion starts at [0,0] and the step size is distributed according to a Levy [1]

### 4.1 Performance evaluation of LFGWO method

To compare the performance of proposed LFGWO, 15 representative benchmark functions are used as listed in Table 1 [19, 43, 44]. In the table, the range of decision variables and optimal value of each benchmark function is also presented. The results of LFGWO are compared with PSO, GSA, FEP, and GWO meta-heuristic algorithms. For all the algorithm, default value of parameters have been set as described in their respective papers. However, population size ($N$) and number of iterations ($itr$) are kept fixed for all methods to 100, and 1000 respectively. As these algorithms are stochastic in nature, hence each executed 30 times and are compared in terms of average fitness, standard deviation, and convergence behavior.

**Table 1** Represented benchmark functions

| Sr. No. | Equation | Dimensions | Range | Optimal value |
|---|---|---|---|---|
| 1 | $F_1(X) = \sum_{i=1}^{d} x_i^2$ | 30 | [−100, 100] | 0 |
| 2 | $F_2(X) = \sum_{i=1}^{d} \mid x_i \mid + \prod_{i=1}^{d} \mid x_i \mid$ | 30 | [−10, 10] | 0 |
| 3 | $F_3(X) = max_i\{\mid x_i \mid, 1 \leq i \leq d\}$ | 30 | [−100, 100] | 0 |
| 4 | $F_4(X) = \sum_{i=1}^{d-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$ | 30 | [−30, 30] | 0 |
| 5 | $F_5(X) = \sum_{i=1}^{d} ([x_i + 0.5])^2$ | 30 | [−100, 100] | 0 |
| 6 | $F_6(X) = \sum_{i=1}^{d} ix_i^4 + random[0, 1)$ | 30 | [−1.28, 1.28] | 0 |
| 7 | $F_7(X) = -20exp\left(-0.2\sqrt{\frac{1}{d}\sum_{i=1}^{d} x_i^2}\right) - exp\left(\frac{1}{d}\sum_{i=1}^{d} cos(2\pi x_i)\right) + 20 + e$ | 30 | [−32, 32] | 0 |
| 8 | $F_8(X) = \frac{\pi}{d}\{10\sin(\pi y_1) + \sum_{i=1}^{d-1}(y_i - 1)^2[1 + 10\sin^2(\pi y_{(i+1)})] + (y_d - 1)^2\} + \sum_{i=1}^{d} u(x_i, 10, 100, 4)$ $where\ y_i = 1 + \frac{(x_i+1)}{4},\ and\ u(x_i, a, k, m) = \begin{cases} k(x_i - a)^m & x_i > a \\ 0 & -a \leq x_i \leq a \\ k(-x_i - a)^m & x_i < -a \end{cases}$ | 30 | [−50, 50] | 0 |
| 9 | $F_9(X) = 0.1\{\sin^2(3\pi x_1) + \sum_{i=1}^{d}(x_i - 1)^2[1 + \sin^2(3\pi x_i + 1)] + (x_d - 1)^2[1 + \sin^2(2\pi x_d)]\} + \sum_{i=1}^{d} u(x_i, 5, 100, 4)$ | 30 | [−50, 50] | 0 |
| 10 | $F_{10}(X) = \sum_{i=1}^{11}\left[a_i - \frac{x_1(b_i^2+b_i x_2)}{b_i^2+b_i x_3+x_4}\right]^2$ | 4 | [−5, 5] | 0.0003 |
| 11 | $F_{11}(X) = 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4$ | 2 | [−5, 5] | −1.0316 |
| 12 | $F_{12}(X) = (x_2 - \frac{5.1}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6)^2 + 10(1 - \frac{1}{8\pi})cos\ x_1 + 10$ | 2 | [−5.5] | 0.398 |
| 13 | $F_{13}(X) = [1 + (x_1 + x_2 + 1)^2(19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)] \times [30 + (2x_1 - 3x_2)^2 \times (18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2)]$ | 2 | [−2, 2] | 3 |
| 14 | $F_{14}(X) = -\sum_{i=1}^{4} c_i\ exp(-\sum_{j=1}^{3} a_{ij}(x_j - p_{ij})^2)$ | 3 | [1,3] | −3.86 |
| 15 | $F_{15}(X) = -\sum_{i=1}^{4} c_i\ exp(-\sum_{j=1}^{6} a_{ij}(x_j - p_{ij})^2)$ | 6 | [0,1] | −3.32 |

The comparative results of average fitness values, generated by proposed and other methods, are presented in Table 2. Table 2 depicts that the proposed LFGWO computes the best average values among PSO, GSA, FEP, and GWO for all the considered benchmark functions except $F_5$ and $F_{14}$ where, FEP and PSO performs slightly better than LFGWO respectively.

To validate the results of Table 2, statistical analysis has been performed using wilcoxon rank sum test [11]. For the same, NULL hypothesis is considered that two algorithms are similar at the 5% significance level for all the benchmark functions. The results of wilcoxon rank sum tests has been shown in Table 3 which depicts $p - value$ and $h - value$ as computed by comparing the fitness values of 30 runs for existing and proposed algorithms. It is considered that if $p < 0.05$ then null hypothesis will be rejected and symbolized by '$+$' or '$-$' otherwise accepted and represented by '$=$'. The symbol '$+$' shows that LFGWO computes significantly different and better results as compared to existing algorithms while '$-$' represents significantly different and poor results. From Table 3, it is visualized that the proposed algorithm shows significantly different and better results or no significant difference for maximum number of functions except $F_5$ where FEP shows comparatively better results. If we analyze both the Tables 2 and 3, the wilcoxon rank sum test shows that the difference for $F14$ in Table 2 is not significant. Therefore, it validates that the proposed LFGWO outperforms existing algorithms in terms of mean fitness values.

Moreover, the variations among the optimal values in 30 runs have been analyzed using standard deviation values which are also shown in Table 2. From the table, it is confirmed that LFGWO shows more consistent results for maximum number of functions. This consistency has also been validated by drawing box-and-whisker diagrams [28] in Fig. 8 against each considered and proposed algorithms over 30 runs. The box-and-whisker diagrams are only shown for four representative functions. The name of the algorithms are represented in the horizontal axis of the box-plot diagram, while the vertical axis denotes the best fitness values. All the vertical axes in the figures are represented in a linear scale. From Fig. 8, it can be observed that LFGWO represents more consistent results for all the representative functions. Therefore, the statistical analysis also validates that in almost all cases, the proposed LFGWO generates more consisting and high precision results as compared to other existing algorithms.

The convergence behavior of a meta-heuristic algorithm is also an important component to be compared. Therefore, the proposed LFGWO has also been analyzed in terms of its convergence behavior by plotting the convergence graph. The convergence graph are also shown for four representative functions. Figure 9 depicts the convergence trends of all the algorithms for 200, 400, 600, 800, and 1000 iterations. The y-axes of the convergence graphs show the corresponding best fitness values and are represented in a logarithmic scales. From the figures, it can be observed that the LFGWO also shows better convergence rate for all the representative benchmark functions than existing algorithms. This validates that the proposed LFGWO algorithm not only finds the better fitness values but also shows better convergence behavior for standard benchmark functions.

### 4.2 Performance evaluation of feature selection method

The proposed feature selection method based on LFGWO is compared and analyzed with PSO and GWO based feature selection method with respect to image steganalysis. For

**Table 2** Comparative analysis of existing and proposed methods for mean and standard deviation values on the benchmark functions

| Function | LFGWO | | GWO | | PSO | | GSA | | FEP | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | STD | Mean | STD | Mean | STD | Mean | STD | Mean | STD |
| $F_1$ | 3.77E-39 | 8.04E-39 | 1.43E-27 | 2.41E-27 | 1.36E-04 | 2.02E-04 | 2.53E-16 | 9.67E-17 | 5.70E-04 | 1.30E-04 |
| $F_2$ | 1.38E-24 | 1.40E-24 | 8.62E-17 | 6.50E-17 | 0.0421 | 0.0454 | 0.0557 | 0.1941 | 8.10E-03 | 7.70E-04 |
| $F_3$ | 1.80E-10 | 5.86E-10 | 1.02E-06 | 9.64E-07 | 1.0865 | 0.3170 | 7.3549 | 1.7415 | 0.3 | 0.5 |
| $F_4$ | 4.3906 | 3.3969 | 27.0420 | 0.8882 | 96.7183 | 60.1156 | 67.5431 | 62.2253 | 5.06 | 5.87 |
| $F_5$ | 0.5562 | 0.2333 | 0.7515 | 0.3627 | 1.02E-04 | 8.28E-05 | 2.50E-16 | 1.74E-16 | 0 | 0 |
| $F_6$ | 3.57E-04 | 2.04E-03 | 2.01E-03 | 1.08E-03 | 0.1229 | 0.0450 | 0.0894 | 0.0434 | 0.1415 | 0.3522 |
| $F_7$ | 1.69E-14 | 3.70E-15 | 1.01E-13 | 1.43E-14 | 0.2760 | 0.5090 | 0.0621 | 0.2363 | 0.018 | 2.10E-03 |
| $F_8$ | 3.71E-07 | 0.0155 | 0.0430 | 0.0154 | 6.92E-03 | 0.0263 | 1.7996 | 0.9511 | 9.20E-06 | 3.60E-06 |
| $F_9$ | 7.15E-05 | 2.21E-01 | 0.6711 | 0.2856 | 6.68E-03 | 8.91E-03 | 8.8991 | 7.1262 | 1.60E-04 | 7.30E-05 |
| $F_{10}$ | 3.89E-04 | 7.50E-03 | 4.47E-03 | 9.30E-03 | 5.77E-04 | 2.22E-04 | 3.67E-03 | 1.65E-03 | 5.00E-04 | 3.20E-04 |
| $F_{11}$ | −1.0316 | 1.36E-06 | −1.0316 | 2.19E-08 | −1.0316 | 6.25E-16 | −1.0316 | 4.88E-16 | −1.03 | 4.90E-07 |
| $F_{12}$ | 0.3979 | 6.29E-05 | 0.3979 | 8.86E-05 | 0.3979 | 0 | 0.3979 | 0 | 0.3980 | 1.50E-07 |
| $F_{13}$ | 3.0000 | 1.33E-05 | 3.0001 | 4.40E-05 | 3.0000 | 1.33E-15 | 3.0000 | 4.17E-15 | 3.02 | 0.11 |
| $F_{14}$ | −3.8625 | 1.70E-04 | −3.8613 | 2.81E-03 | −3.8628 | 2.58E-15 | −3.8628 | 2.29E-15 | −3.86 | 1.40E-05 |
| $F_{15}$ | −3.3812 | 0.0545 | −3.2702 | 0.0726 | −3.2663 | 0.0605 | −3.3178 | 0.0231 | −3.27 | 0.059 |

**Table 3** Results of the wilcoxon rank sum test for statistically significance level at $\alpha$ = 0.05 on benchmark functions

| Function | LFGWO-GWO | | | LFGWO-PSO | | | LFGWO-GSA | | | LFGWO-FEP | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $p-value$ | $h-value$ | $SGFNT$ | $p-value$ | $h-value$ | $SGFNT$ | $p-value$ | $h-value$ | $SGFNT$ | $p-value$ | $h-value$ | $SGFNT$ |
| $F_1$ | 1.17E- 12 | 1 | + | 1.17E- 12 | 1 | + | 1.17E- 12 | 1 | + | 1.17E- 12 | 1 | + |
| $F_2$ | 1.91E- 09 | 1 | + | 1.17E- 12 | 1 | + | 1.26E- 05 | 1 | + | 4.71E- 08 | 1 | + |
| $F_3$ | 1 | 0 | + | 1.17E- 12 | 1 | + | 1 | 0 | = | 1.91E- 09 | 1 | + |
| $F_4$ | 2.87E- 11 | 1 | + | 1.17E- 12 | 1 | + | 1.91E- 09 | 1 | + | 5.62E- 11 | 1 | + |
| $F_5$ | 1 | 0 | = | 2.87E- 11 | 1 | + | 1.70E- 10 | 1 | - | 2.87E- 11 | 1 | - |
| $F_6$ | 2.31E- 08 | 1 | + | 6.32E- 07 | 1 | + | 8.47E- 02 | 0 | = | 2.87E- 11 | 1 | + |
| $F_7$ | 1.17E- 12 | 1 | + | 1.17E- 12 | 1 | + | 1.17E- 12 | 1 | + | 1.17E- 12 | 1 | + |
| $F_8$ | 1.17E- 12 | 1 | + | 1.17E- 12 | 1 | + | 1.17E- 12 | 1 | + | 1.17E- 12 | 1 | + |
| $F_9$ | 1.47E- 07 | 1 | + | 2.87E- 11 | 1 | + | 2.09E- 05 | 1 | + | 8.60E- 07 | 1 | + |
| $F_{10}$ | 1.26E- 05 | 1 | + | 1.17E- 12 | 1 | + | 1.36E- 03 | 1 | + | 1.10E- 02 | 1 | + |
| $F_{11}$ | 1 | 0 | = | 1.17E- 12 | 1 | + | 1 | 0 | = | 2.87E- 11 | 1 | + |
| $F_{12}$ | 1 | 0 | = | 1 | 0 | = | 1 | 0 | = | 1 | 0 | = |
| $F_{13}$ | 1 | 0 | = | 1 | 0 | = | 1 | 0 | = | 1 | 0 | = |
| $F_{14}$ | 1.06E- 11 | 1 | + | 1 | 0 | = | 1 | 0 | = | 1 | 0 | = |
| $F_{15}$ | 2.87E- 11 | 1 | + | 2.87E- 11 | 1 | + | 4.92E- 08 | 1 | + | 2.87E- 11 | 1 | + |

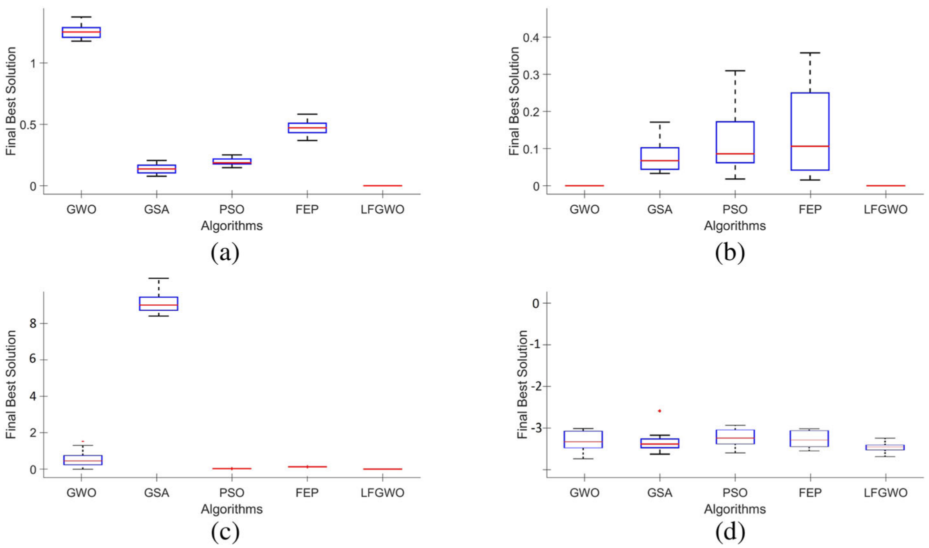



**Fig. 8** The box-plot graphs for benchmark functions **a** $F_1$, **b** $F_6$, **c** $F_9$, and **d** $F_{15}$

the same, 1000 gray scale images from BOSSbase ver 1.01 steganographic image dataset has been considered containing 10000 cover images and 10000 stego images. Each image goes through the feature extraction process by SPAM and AlexNet methods which generate 646 and 1000 features respectively. Further, the extracted features are fed to the proposed LFGWO based feature selection method.

The performance of LFGWO based feature selection method has been compared with PSO and GWO based feature selection method in terms of the number of features selected

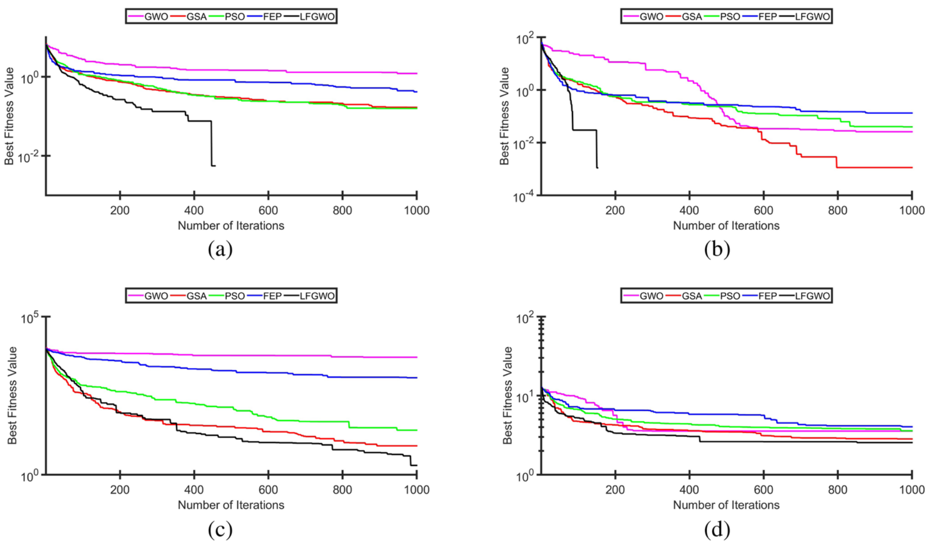



**Fig. 9** The convergence graphs for benchmark functions **a** $F_1$, **b** $F_6$, **c** $F_9$, and **d** $F_{15}$

**Table 4** Comparison of different feature selection methods and accuracy of classifiers

| Feature selection method | Number of features selected (SPAM) | Number of features selected (AlexNet) | Classification method | Accuracy on SPAM | Accuracy on AlexNet |
|---|---|---|---|---|---|
| | | | SVM | 40.28 | 39.36 |
| | | | LDA | 37.39 | 40.23 |
| None | 686 | 1000 | RF | 52.21 | 47.39 |
| | | | kNN | 49.36 | 42.33 |
| | | | ZeroR | 32.12 | 31.32 |
| | | | SVM | 51.56 | 49.97 |
| | | | LDA | 53.22 | 50.06 |
| PSO | 95 | 94 | RF | 59.63 | 56.35 |
| | | | kNN | 54.15 | 53.05 |
| | | | ZeroR | 34.33 | 32.15 |
| | | | SVM | 57.56 | 54.15 |
| | | | LDA | 57.68 | 55.23 |
| GWO | 87 | 92 | RF | 61.15 | 57.25 |
| | | | kNN | 58.92 | 54.13 |
| | | | ZeroR | 36.45 | 32.22 |
| | | | SVM | 67.23 | 65.49 |
| | | | LDA | 66.31 | 63.12 |
| LFGWO | 84 | 89 | RF | 69.12 | 67.11 |
| | | | kNN | 67.33 | 64.39 |
| | | | ZeroR | 39.42 | 35.39 |

and classification accuracy after feature selection for both the SPAM and AlexNet feature extraction methods. The results are shown in Table 4. To calculate the classification accuracy, six classifiers namely; SVM, LDA, RF, kNN, and ZeroR are used. Among all the classifiers, the accuracy returned by ZeroR classifier is considered as a baseline for all the feature selection method. From the table, it has been observed that the proposed feature selection method selects minimum number of features for both the SPAM and AlexNet feature extraction methods as compared to others. Moreover, all the considered classifiers produce better accuracy for the features selected by the proposed method. However, the accuracy returned by RF classifier is the best among others.

Moreover, the impact of the selected prominent features has also been studied for the computational efficiency of a classifier. Table 5 shows the computational time taken by each considered classifier for training before and after the feature selection methods. From the table, it can be visualized that the proposed LFGWO takes the least amount of time to train each considered classifier. This validates that the selection of prominent features not only increases the classification accuracy, but also increases the computational efficiency of a classifier. Therefore, it can be stated that the proposed LFGWO based feature selection method gives less number of features, maintaining the high classification accuracy that is it is capable to remove irrelevant and redundant features.

**Table 5** Analysis of the computational time (in seconds) of the classifiers, before and after the features selection

| Feature selection method | SPAM | | | | | | AlexNet | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Selected features | SVM | LDA | RF | kNN | ZeroR | Selected features | SVM | LDA | RF | kNN | ZeroR |
| None | 686 | 3.23 | 1.04 | 4.12 | 0.21 | 0 | 1000 | 4.98 | 1.21 | 7.31 | 0.29 | 0 |
| PSO | 95 | 1.05 | 0.27 | 0.91 | 0.18 | 0 | 94 | 1.02 | 0.27 | 0.89 | 0.18 | 0 |
| GWO | 87 | 0.95 | 0.23 | 0.78 | 0.14 | 0 | 92 | 0.98 | 0.25 | 0.87 | 0.16 | 0 |
| LFGWO | 84 | 0.91 | 0.21 | 0.67 | 0.11 | 0 | 89 | 0.95 | 0.22 | 0.71 | 0.14 | 0 |

## 5 Conclusion

In this paper, a feature selection method for Image steganalysis has been introduced which is based on novel levy flight-based grey wolf optimizer (LFGWO). The method is tested on BOSSbase ver 1.01 image dataset, containing cover and stego images. Two feature extraction methods, SPAM and AlexNet, have been used which extract 686 and 1000 features respectively. The performance of LFGWO is compared with GWO, PSO, GSA, and FEP meta-heuristic algorithms in terms of mean fitness, standard deviation values, and convergence behavior. The experimental and statistical results validate that the proposed LFGWO outperforms existing meta-heuristic algorithms. Further, the performance of proposed feature selection method based on LFGWO is compared with PSO and GWO based based feature selection methods. The proposed method extracts 84 and 89 features from SPAM and AlexNet extracted features respectively which are the lowest among other considered methods. Moreover, five classifiers (SVM, LDA, RF, kNN, ZeroR) are used to compare the performance of image steganalysis method over selected features. Among all the classifiers, RF outperforms the other classifiers over the proposed feature selection method to classify the images into cover and stego images. This validates that the proposed feature selection method reduces the irrelevant and redundant features, maintaining the high classification accuracy.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

1. Amirsadri S, Mousavirad SJ, Ebrahimpour-Komleh H (2017) A levy flight-based grey wolf optimizer combined with back-propagation algorithm for neural network training. Neural Comput & Applic 28:1–14
2. Bhattacharyya S, Sengupta A, Chakraborti T, Konar A, Tibarewala D (2014) Automatic feature selection of motor imagery eeg signals using differential evolution and learning automata. Med Biol Eng Comput 52:131–139
3. Chechkin AV, Metzler R, Klafter J, Gonchar VY et al (2008) Introduction to the theory of lévy flights, Anomalous transport: Foundations and applications 129–162
4. Chen C, Shi YQ (2008) Jpeg image steganalysis utilizing both intrablock and interblock correlations. In: Proc of IEEE international symposium on circuits and systems, pp 3029–3032
5. Chhikara RR, Sharma P, Singh L (2016) A hybrid feature selection approach based on improved pso and filter approaches for image steganalysis. Int J Mach Learn Cybern 7:1195–1206
6. Chhikara RR, Sharma P, Singh L (2016) An improved dynamic discrete firefly algorithm for blind image steganalysis. Int J Mach Learn Cybern 6:1–15
7. Cotter SF, Kreutz-Delgado K, Rao BD (2001) Backward sequential elimination for sparse vector subset selection. Signal Process 81:1849–1864
8. Dash M, Liu H (1997) Feature selection for classification. Intelligent Data Analysis 1:131–156
9. Deepa S (2017) Steganalysis on images using svm with selected hybrid features of gini index feature selection algorithm. Int J Adv Res Comput Sci 8:1503–1509
10. Deng H, Runger G (2012) Feature selection via regularized trees. In: Proc of international joint conference on neural networks, pp 1–8
11. Derrac J, García S, Molina D, Herrera F (2011) A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. Swarm Evol Comput 1:3–18
12. Dumitrescu S, Wu X, Wang Z (2003) Detection of lsb steganography via sample pair analysis. IEEE Trans Signal Process 51:1995–2007
13. Emary E, Zawbaa HM, Grosan C, Hassenian AE (2015) Feature subset selection approach by gray-wolf optimization. In: Proc of Afro-European conference for industrial advancement, pp 1–13

14. Emary E, Zawbaa HM, Hassanien AE (2016) Binary grey wolf optimization approaches for feature selection. Neurocomputing 172:371–381
15. Fridrich J, Goljan M (2004) On estimation of secret message length in lsb steganography in spatial domain. In: Security, steganography, and watermarking of multimedia contents VI, pp 23–35
16. Glorot X, Bordes A, Bengio Y (2011) Deep sparse rectifier neural networks. In: Proc of international conference on artificial intelligence and statistics, pp 315–323
17. Guettari N, Capelle-Laizé AS, Carré P (2016) Blind image steganalysis based on evidential k-nearest neighbors. In: Proc of IEEE international conference on image processing, pp 2742–2746
18. Guyon I, Weston J, Barnhill S, Vapnik V (2002) Gene selection for cancer classification using support vector machines. Mach Learn 46:389–422
19. Jamil M, Yang X-S (2013) A literature survey of benchmark functions for global optimisation problems. International Journal of Mathematical Modelling and Numerical Optimisation 4:150–194
20. Kodovskỳ J, Fridrich J (2009) Calibration revisited. In: Proc of ACM workshop on multimedia and security, pp 63–74
21. Kodovskỳ J, Fridrich J (2011) Steganalysis in high dimensions: fusing classifiers built on random subspaces. In: Media watermarking, security, and forensics III, p 78800L
22. Kodovsky J, Fridrich J, Holub V (2012) Ensemble classifiers for steganalysis of digital media. IEEE Trans Inf Forensics Secur 7:432–444
23. Kohavi R, John GH (1997) Wrappers for feature subset selection. Artif Intell 97:273–324
24. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105
25. Liu Q (2011) Steganalysis of dct-embedding based adaptive steganography and yass. In: Proc of ACM multimedia workshop on multimedia and security, pp 77–86
26. Lu P, Luo X, Tang Q, Li S (2004) An improved sample pairs method for detection of lsb embedding. In: Information hiding, vol 3200. Springer, pp 116–127
27. Luo X-Y, Wang D-S, Wang P, Liu F-L (2008) A review on blind detection for image steganography. Signal Process 88:2138–2157
28. Massart DL, Smeyers-verbeke AJ et al (2005) Visual presentation of data by means of box plots, Practical data handling. LC-GC Europe 18:215–218
29. Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. Adv Eng Softw 69:46–61
30. Mohammadi FG, Abadeh MS (2014) Image steganalysis using a bee colony based feature selection algorithm. Eng Appl Artif Intell 31:35–43
31. Muangkote N, Sunat K, Chiewchanwattana S (2014) An improved grey wolf optimizer for training q-gaussian radial basis functional-link nets. In: Proc of international conference on computer science and engineering, pp 209–214
32. Nissar A, Mir A (2010) Classification of steganalysis techniques: a study. Digital Signal Process 20:1758–1770
33. Pevny T, Bas P, Fridrich J (2010) Steganalysis by subtractive pixel adjacency matrix. IEEE Trans Inf Forensics Secur 5:215–224
34. Pevnỳ T, Fridrich J (2005) Towards multi-class blind steganalyzer for jpeg images. In: Proc of international workshop on digital watermarking, pp 39–53
35. Pevny T, Fridrich J (2007) Merging markov and dct features for multi-class jpeg steganalysis. In: Proc of electronic imaging, pp 650503–650503
36. Qin C, Chang C-C, Hsu T-J (2015) Reversible data hiding scheme based on exploiting modification direction with two steganographic images. Multimedia Tools and Applications 74:5861–5872
37. Qin C, Chang CC, Huang YH, Liao LT (2013) An inpainting-assisted reversible steganographic scheme using a histogram shifting mechanism. IEEE Trans Circuits Syst Video Technol 23:1109–1118
38. Ramezani M, Ghaemmaghami S (2010) Towards genetic feature selection in image steganalysis. In: Proc of IEEE consumer communications and networking conference, pp 1–4
39. Rashedi E, Nezamabadi-Pour H, Saryazdi S (2009) Gsa: a gravitational search algorithm. Inform Sci 179:2232–2248
40. Saraswat M, Arya K (2014) Feature selection and classification of leukocytes using random forest. Med Biol Eng Comput 52:1041–1052
41. Saraswat M, Arya K, Sharma H (2013) Leukocyte segmentation in tissue images using differential evolution algorithm. Swarm Evol Comput 11:46–54
42. Sheikhan M, Pezhmanpour M, Moin MS (2012) Improved contourlet-based steganalysis using binary particle swarm optimization and radial basis neural networks. Neural Comput & Applic 21:1717–1728
43. Simon D (2013) Evolutionary optimization algorithms. Wiley, Hoboken

44. Yang X-S (2014) Nature- inspired optimization algorithms. Elsevier, New York
45. Yao X, Liu Y, Lin G (1999) Evolutionary programming made faster. IEEE Trans Evol Comput 3:82–102
46. Zhang T, Ping X (2003) A new approach to reliable detection of lsb steganography in natural images. Signal Process 83:2085–2093


**Yadunath Pathak** has done his Master in Technology in computer science and engineering from Atal Bihari Vajpayee Indian Institute of Information Technology and Management, Gwalior, Madhya Pradesh, India (2014). Currently, he is pursuing his PhD degree in computer science and engineering from Atal Bihari Vajpayee Indian Institute of Information Technology and Management, Gwalior, Madhya Pradesh, India. His research interest includes Bio-Medical Engineering, Digital Image Processing and Machine Learning techniques.

**Dr. K. V. Arya** received the B. Sc. from Rohilkhand University- Bareilly, India. He received M.E. (Integrated) from Indian Institute of Science–Bangalore, India. He received the Ph.D. degree in Computer Science and Engineering from Indian Institute of Technology–Kanpur, India. He has been an Associate Professor with the Department of Computer Science and Engineering at Atal Bihari Vajpayee Indian Institute of Information Technology and Management, Gwalior, Madhya Pradesh, India. He has more than 25 years of teaching and research experience. He has more than 120 research papers in international journals, book chapters, and conference proceedings. His main research focuses on Image Processing: Image Enhancement, Image Registration, Biometrics: Face Recognition, Iris Recognition, Gait, Wireless Ad Hoc Networks: Energy Aware routing, Secure routing and InformationSecurity: Intrusion Detection and Prevention.

**Dr. Shailendra Tiwari** an Assistant Professor in the CSE Department, Thapar Institute of Engineering & Technology, Patiala, India. Dr. Tiwari received the Master of Engineering (M.E.) degree in Computer Engineering from Punjab Engineering College (PEC), Chandigarh, India and earned Ph. D. degree in Computer Science & Engineering from Indian Institute of Technology (BHU), Varanasi. He has more than 9 years of teaching and research experience. He has more than 30 research papers in International Journals, Book Chapters, and Conference Proceedings. His research interests include Medical Image Reconstruction, Segmentation Algorithm in Medical Imaging, OptimizationTechniques and other non-invasive problem estimation.