Chapter 4
Improved Detection of Steganographic Algorithms in Spatial LSB Stego Images Using Hybrid GRASP-BGWO Optimisation

S. T. Veena, S. Arivazhagan and W. Sylvia Lilly Jebarani

Abstract With the success of passive steganalysis, active steganalysis proceeds with its first step to reveal the steganographic algorithms being used to create the stego images. This process needs to be modelled as a multi-class classification problem. Stem to stern analysis of the literature points out that the existing universal steganalytic features are a thorn in the flesh because of their dimensionality curse (34,671). Hence this work concentrates on detection of steganographic algorithms by optimal novel features christened Local Residual Pattern (LRP) and Local Distance Pattern (LDiP). LRP captures first order derivatives of the high pass filtered output, while LDiP exploits the multi scaled radii neighbourhood to capture deformities at a distance. Acquiring LRP and LDiP from fifteen different kernels, this work focuses to find optimal features by the proposed hybrid technique of Greedy Randomised Adaptive Search—Binary Grey Wolf Optimisation (GRASP-BGWO). Deriving confidence from the bio-inspired algorithm and the divide and conquer approach of the proposed optimisation, this work succeeds in improving the performance of the employed ensemble logistic regression classifier with minimal features. Experimentations conducted using five representative algorithms of spatial Least Significant Bit (LSB) embedding category for eight different payloads show that the developed minimal feature steganalyser outperforms the state-of-the-art steganalysers.

Keywords Active steganalysis · Hybrid optimisation · Local descriptors
Multi-class classification · Ensemble classifier

S. T. Veena (✉) · S. Arivazhagan · W. Sylvia Lilly Jebarani
Department of Electronics and Communication Engineering,
Centre for Image Processing and Pattern Recognition,
Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India
e-mail: veena_st@mepcoeng.ac.in

S. Arivazhagan
e-mail: sarivu@mepcoeng.ac.in

W. Sylvia Lilly Jebarani
e-mail: wsylvia@mepcoeng.ac.in

4.1 Introduction

In the age of digitisation, it is easier to hide secret messages into any image using steganography which can be a challenge to national or international security. Globalisation has made a large number of steganographic tools freely and easily available even to illegitimate users for example a single website [1] contains more than 110 free steganographic tools. It is the job of steganalysis to surveil these secret communications. Steganalysis starts off with simple detection of stego images from the innocent cover images and proceeds to extract or decipher the secret hidden within them. The former task is known as passive steganalysis and the latter processes are collectively termed as active steganalysis. A large number of literature exists for passive steganalysis of both targeted and universal nature [2–5]. Though the targeted steganalysers are found to be more accurate, the universal steganalysers enjoy favouritism in the context of being able to work on a large range of steganographic algorithms. Particularly, universal steganalysis of spatial LSB steganography in raw image formats has attracted researchers because of their very low embedding change rates and poses a tougher challenge than the JPEG steganalysis. The low volume payload and the content adaptive LSB steganography are two open challenges in spatial LSB steganography [6, 7]. Also, there are not many literary works in active steganalysis as in passive steganalysis. And the first task of active steganalysis is that of identification of tools or the algorithm involved in creating the stego images. Identification of the tool is taken up as a branch of forensic study and most of them are signature based steganalysis [8, 9], while identification of algorithms is handled as a pattern recognition process. However, not much of work in literature supports identification of the steganographic algorithm involved.

The first step in this direction of detecting algorithm used in creating stego images was that of classification of the JPEG steganographic techniques [10]. Here, the Discrete Cosine Transform (DCT) features previously developed were used along with a multi-class Support Vector Machine (SVM) with Gaussian kernel trained with images from four JPEG techniques namely F5, MB1, MB2 and Outguess. The multi-class classifier was built on one against one strategy and named in the paper as Max-Wins. They were able to classify images with large messages reliably and when tested with new schemes, they were assigned to closely-related trained schemes. The authors extended their work to double compressed JPEG images with six techniques using calibrated DCT features with the same classifier [11]. They reported that as the JPEG quality factor of compression increases, the reliability of the classifier deteriorates. The technique with low embedding rate was the worst to be identified amongst all. They also inferred that due to two similar embedding algorithms, there may be a merging of results making them indistinguishable. The training sets need to be very dense in the context of techniques and quality factors to give a more reliable result.

Later, Pevny and Fridrich used the average of the DCT features along with Markov features instead of simple concatenation of features to develop a reduced

set of features to classify embedding technique in JPEG images [12]. The challenge that the ability of a classifier trained on diverse algorithms may fail to identify unseen images from closely related methods, even as stego, was the inspiration. They built a forerunner for estimating the quality factor and this bi layered double compression detector was followed by the multi-class classifier. They made an interesting note that the multi classifier will not be able to detect steganographic methods with entirely different types of embedding changes. Dong et al. proposed run-length feature based SVM multi classifier for classification of algorithms in both spatial and JPEG images [13]. They also studied hierarchical and non-hierarchical multi-class schemes. In the hierarchical scheme, a separation of the cover and stego images was done followed by separation of the stego classes. The results of the experimentation conducted showed that the hierarchical scheme performed better. The misclassification mostly existed within the intra domain techniques rather than within the inter domain. This was the first scheme that included tested images on the spatial domain.

In [14], the multi-class classification was also carried out with Logistic Regression (LR) classifier and five classes (cover + four spatial algorithms— LSBR, LSBM, LSBR2, LSBRmod5) on three databases. The authors used Subtractive Pixel Adjacency Matrix (SPAM) features and t-test to validate the detection accuracy. They found that LIBSVM was more efficient than LR in passive steganalysis but LR was the best for multi classification. The single bit and multi bit embedding made no difference in the performance with SPAM features. The authors caution that claim on improvement should be on equal footing in all aspects of steganalysis. Zhu et al. suggested an ensemble multi-class classifier for steganalysis of JPEG images with Cartesian Calibrated JPEG domain Rich Model (CC–JRM) features with linear SVM as the base classifier [15]. They used two schemes for ensemble classification and claim less computation cost than other classifiers.

All the reported works or literature for algorithm detection were for JPEG images and the only literature that exists for spatial LSB is that of Lubenko and Ker, which suggest the difficulty of the task in spite of its need. This stays as a motivation to perform algorithm detection in spatial LSB stego images using machine learning. The existing passive steganalytic features [2, 16, 17] are mostly extracted from residuals such that it is rich in stego content and devoid of the cover content. Then, co-occurrence matrices from the quantised and thresholded residual, is formed as a pattern to distinguish stego from cover. However, while moving to the higher order, the co-occurrence matrices become sparsely populated; truncation and quantisation lead to the loss of the minute changes produced by steganographic embedding. Shi et al. suggested Local Binary Pattern as more capable operator than co-occurrence matrices [18]. Following this course, this paper presents a residual based local descriptor for steganalysis.

Similarly, the performance of classification is improved by simple union or concatenation of diverse individual models [16]. However, this leads to a feature which is very huge in dimension. One of the existing state-of-art steganalytic features—Spatial Rich Model (SRM) formed using this technique has a very huge

dimension of 34,671. This makes classification task difficult by requiring special classifiers to handle that dimensionality. Also, it was shown by Lyu and Farid that type and number of features being concatenated are crucial to improve the quality of performance and a simple concatenated feature model will not yield optimal efficiency [19]. Therefore it is necessary to obtain both optimally concatenated model from individual models and also to reduce the dimensionality of the so obtained concatenated model for algorithm steganalysis. Hence, optimisation is done in this paper in two phases or as a hybrid. The first phase of optimisation finds out the optimal combination of discriminant individual feature models and the second phase of optimisation proceeds to reduce dimension within the obtained combination of features. The authors in their previous ventures proposed a similar hybrid optimisation algorithm–Greedy Randomised Adaptive Search—Recursive Feature Elimination (GRASP-RFE (GR)) for selection/reduction of features which are based on the principle of divide and conquer to estimate the size of payload in spatial LSB stego images. The proposed GRASP-RFE was found to be very efficient; however the limitation was that the dimension is user defined [20]. Therefore a dynamic hybrid optimisation–Greedy Randomised Adaptive Search—Binary Grey Wolf Optimisation (GRASP-BGWO (GB)) involving a more powerful bio-inspired algorithm is proposed in this paper. This hybrid optimisation is applied for algorithm detection steganalysis of spatial LSB algorithms using the proposed local descriptors. Thus, necessary and tough task of identification of spatial LSB algorithms using minimal optimally concatenated features of novel local descriptors by the proposed hybrid feature selection of GRASP-BGWO is presented in this paper. The paper is organised as follows: The basics of the steganographic algorithms to be detected is presented in Sect. 4.2. Section 4.3 explains the proposed features and Sect. 4.4 presents the proposed hybrid optimisation technique in detail. The experiments conducted and the results are discussed in Sect. 4.5. The paper concludes in Sect. 4.6 with scope for future enhancements.

4.2   Basics of Spatial LSB Algorithms

This section introduces five spatial LSB algorithms—LSB Replacement (LSBR), LSB Matching (LSBM), LSBM Revisited (LSBMR), Two bit LSBR (LSBR2 or 2LSB) and Modulo 5 LSBR (LSBRmod5). In LSBR, a random secret data bit replaces the LSB of the cover image to give the corresponding stego image, while in LSBR2, the last two least significant bits are replaced [14, 21]. Embedding in LSB leads to inherent asymmetry with even values either unchanged or increased by 1 and odd values either unchanged or decreased by 1. To counteract this, LSBM (also known as ±1 embedding) embeds 1 randomly by either adding to or subtracting from the cover image, if the secret data bit does not match the LSB of the cover image [22]. In LSBMR, the embedding is performed using a pair of pixels as a unit so that fewer pixel change rate is encountered than LSBM [23]. In LSBRmod5 embedding, the least significant digits are adjusted such that the

remainder of dividing stego pixel by 5 gives the embedding secret digit [14, 24]. The models are represented in Eq. (4.1).

$$LSBR(X) = 2 \times \lfloor X/2 \rfloor + M$$
$$LSBM(X) = 2 \times \lfloor X/2 \rfloor \pm M$$
$$LSBMR(X) = LSBR(f(p,q)) \qquad (4.1)$$
$$LSBR2(X) = 4 \times \lfloor X/4 \rfloor + M$$
$$LSBRmod5(X) = argmin_{Y mod 5 = M}|X - Y|$$

where X, Y are pixels $\in \{0, 1, \ldots, 255\}$, M is the secret message in bits and $f(p,q)$ is the function defined on pixel pairs (p,q). All the LSB based algorithms embed the secret at random location based on the stego key.

4.3 Proposed Features

The victory of textural co-occurrence features [2] in spatial LSB steganalysis led to the search of other textural features that may help in steganalysis. Local Binary Pattern (LBP) is one such textural feature used in various applications, but its application in steganalysis is not fully exploited [25, 26]. Also, Shi et al. [18] demonstrated that LBP features are better than the co-occurrence features since they are sensitive to noise and are able to capture the deformities of the embedding algorithm in a local neighbourhood. But LBP is a first order statistic and is non-directional in the sense that it encodes the first order derivative difference in all directions. The authors of the paper in their previous venture have proposed a local descriptor called Local Filter Pattern (LFP) for passive steganalysis and found it effective [5]. So, a local descriptor Local Residue Pattern (LRP) that captures LSB distortion using directional and high order information is proposed for LSB steganalysis. To capture subtle distortion patterns that exist within a neighbourhood at varying distance, Local Distance Pattern (LDiP) is proposed. The features are explained in detail in the following subsection.

4.3.1 Local Residue Pattern (LRP)

A local descriptor which acts upon the residues of high pass filters is presented for steganalysis of LSB based steganography. High pass filtering plays an inevitable role in steganalysis since stego signals are additive noises and the image content is suppressed by filtering. Thus, a residue Re is formed from the high passed filtered output, which is independent of the image content but contains the noise or the information embedded inside it.

$$Re = I * k \tag{4.2}$$

where I is the input image, k is the high pass filter and $*$ is the convolution operation. The proposed LRP is developed on this residue as magnitude LRP and sign LFP as extended forms of local filter pattern [5] with additional kernels. The first order derivative differences between the residue values are captured by the magnitude LRPs (mLRPs). The sign LRPs (sLRPs) capture the first order derivative differences of the sign (direction) change in residue.

Various linear high pass filter kernels for computing residues used in this research are shown in Fig. 4.1.

The choice of the kernels has been found suitable for steganalysis in various available literature [18, 27, 28]. For the computation of LRP, the first step is to find the residues Re_θ of the image using filter kernels $k1$ to $k15$ in different directions θ using Eq. (4.2).

In case of kernels $k1$ to $k10$—out of the eight different directions, only four of them—horizontal, vertical, major and minor diagonal directions are considered (i.e.) $\theta = \{0°, 45°, 90°, 135°\}$ because of the symmetric nature of residues. In case of residuals from kernels $k11$ to $k14$, two possible directions are considered (i.e.) $\theta = \{0°, 180°\}$. In case of kernel $k15$, processing in a single direction is considered (i.e.) $\theta = \{0°\}$. Then, the magnitude part of LRP (mLRP) is encoded on the residue output $Re_{\theta,c}$ of a local neighbourhood (pixels in a local window) with c as its centre pixel as shown by Eq. (4.3).

$$mLRP_{B,R}(Re_{\theta,c}) = \sum_{i=1}^{B} f(Re_{\theta,i} - Re_{\theta,c})2^{i-1}$$

$$where,$$

$$f(Re_{\theta,i} - Re_{\theta,c}) = \begin{cases} 0 & if \quad Re_{\theta,i} < Re_{\theta,c} \\ 1 & otherwise \end{cases} \tag{4.3}$$
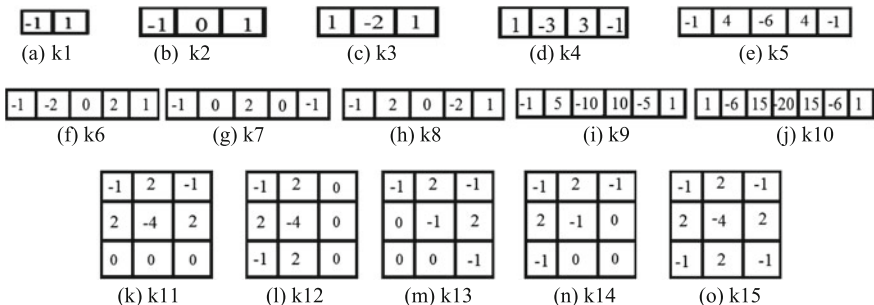

Fig. 4.1 Various high pass filter kernels used

and θ in D and $D = \{0°, 45°, 90°, 135°\}$ or $D = \{0°, 180°\}$ or $D = \{0°\}$ depending on the kernel, B is the number of neighbours in the local window considered and R is the radius of the local neighbourhood from its centre pixel for which binary coding is done using function f. An example illustrating the LRP binary coding is given as Fig. 4.2.

The histogram of the mLRP, $Hist(mLRP_{B,R})$ is the image feature that is constructed by concatenating the encoded output from all applicable directions and binning the occurrences of the concatenated output as given by Eq. (4.4).

$$Hist\left(mLRP_{B,R},j\right) = Hist\left(\{mLRP_{B,R}\left(Re_{\theta,c}\right)|\theta \in D\},j\right) \qquad (4.4)$$

In this study, the value of B and R is taken as 8 and 1 respectively. As a result, the feature vector is 256 in dimension. To further reduce the dimension, rotation invariant form of LBP is also used, since the starting order of the binary sequence is immaterial for steganalysis. The histogram of the rotation invariant form $mLRP_{B,R}^{ri}$ given by Eq. (4.5) has a dimension of 36.

$$mLRP_{B,R}^{ri}\left(Re_{\theta,c}\right) = \min_{0 \leq i \leq 2^{B-1}} ROR\left(mLRP_{B,R}\left(Re_{\theta,c}\right),i\right) \qquad (4.5)$$

where $ROR(x,i)$ denotes 'i' right bitwise rotations on number 'x'. Thus, a total of 30 (15 rotation variant and 15 rotation invariant) mLRPs are proposed as feature sets for mLRP feature model.

Similarly, the sign or direction based LRP (sLRP) also known as Local Filter Pattern (LFP) [5] is defined as shown in Eq. (4.6).

$$sLRP_{B,R}\left(Re_{\theta,c}\right) = \sum_{i=1}^{B} f'\left(Re_{\theta,i}, Re_{\theta,c}\right)2^{i-1} \qquad (4.6)$$

26	38	23
28	22	113
34	22	64

(a) Sample

-1	2	-1
2	-4	2
-1	2	-1

(b) Kernel k15

6	-151	188
-8	167	-294
-58	11	-8

(c) Filtered Output

0	0	1
0		0
0	0	0

(d) mLRP value: 00100000

1	0	1
0		0
0	1	0

(e) sLRP value: 10100100

Fig. 4.2 Illustration of LRP binary encoding

where $D = \{0°, 45°, 90°, 135°\}$ or $D = \{0°, 180°\}$ or $D = \{0°\}$ depending on the kernel. The histogram for sLRP is encoded in the same way as mLRP using Eq. (4.4). The rotation invariant form of sLRP, $sLRP_{B,R}^{ri}$ is given by Eq. (4.5) replacing mLRP with sLRP. Thus, thirty (15 + 15) feature models of sLRP capture the higher order gradient information from the residuals. Thus, a total of 60 feature sets exist for LRP feature model.

4.3.2 Local Distance Pattern (LDiP)

To further capture the dependencies that exist between pixels within a distance, the following arrangement of neighbouring pixels are considered as shown in Fig. 4.3b. The value indicates the sequence of the neighbours in forming the binary pattern. This rectangular pattern of considering neighbours rather than the conventional square type helps in capturing dependencies that exist over sequential neighbours at a distance. Also, alternate left and right numbering of neighbours, help in giving weightage to the neighbour dependencies directly proportional to their distance from centre pixel. Thus, pixels near to the centre pixel will form Most Significant Bits in the binary pattern, thereby contributing more to capturing distortions by embedding changes. The vertical, horizontal and two diagonal directions of the operator are indicated by 0LDiP, 90LDiP, 45LDiP and 135LDiP. The sign and magnitude form of LDiP are constructed in the same way as LRP as in Eqs. (4.3 and 4.6). An example illustrating the LDiP binary encoding is given in Fig. 4.3.

The histograms of LDiP are constructed using Eq. (4.7).

$$Hist(\theta LDiP_B, j) = Hist(LDiP_B(Re_{\theta,c}), j) \tag{4.7}$$

The rotation invariant form of LDiP is also constructed. Here again, the value of B is taken to be 8. Total of 16 feature sets (8 rotation variant + 8 rotation invariant) are formed as LDiP feature sets.

The LRP and LDiP represent the histogram features of the LRP and LDiP (both sign and magnitude) respectively, while LRPri and LDiPri represent the rotation invariant LRP and LDiP histogram features. The sign or magnitude representation is done by 's' or 'm' preceding them. In case of LRP, the kernel from which the feature has been arrived is represented at the posterior. While in LDiP, the direction is represented preceding the sign or magnitude representation. The naming convention and the 76 proposed feature sets formed using LRP and LDiP feature models are summarised in Table 4.1 with their dimensions along with other LBPs found in literature.

(a) Sample   (b) Sequence of Neighbours in LDiP   (c) 0mLDiP value: 01101001

Fig. 4.3 Illustration of LDiP binary encoding

Table 4.1 Summary of the proposed 76 feature models with their dimensionality along with other existing LBP models

Feature models	Dimensionality
$\{s, m\}LRP_{8,1} - k\{1-15\}$, $\{0, 45, 90, 135\}\{s, m\}LDiP_{8,1}$	256
$\{s, m\}LRP_{8,1}^{ri} - k\{1-15\}$, $\{0, 45, 90, 135\}\{s, m\}LDiP_{8,1}^{ri}$	36
LBP [29]	256
LBP^{u2} [30]	59
LBP^{ri} [30]	36
LBP^{riu2} [30]	10

4.4 Proposed Feature Selection Technique

Universal steganalysis is generally done by combining features from different models to form a mega model. This is because a single model generally leads to under populated bins, which hampers the task of universally detecting a wide spectrum of embedding algorithms. However, forming a mega model introduces curse of dimensionality. Optimisation techniques help to reduce dimensionality and thereby save CPU time [31, 32]. Global optimisation techniques like evolutionary algorithms are powerful and robust [33], but consume high CPU time and are poor in terms of convergence. On the other hand, local search algorithms converge faster, but get caught in local minima/maxima. A hybrid or bi-level technique proves to be strong in terms of converging time, thus reducing computation time, at the same time increasing solution quality [34, 35]. A Bi-level optimisation approach (Greedy Random Adaptive Search Procedure–Recursive Feature Elimination (GRASP-RFE)) was proposed by the authors for quantitative steganalysis and was found to be effective. However, the RFE method suffers from two main limitations. The first one is that the dimensionality of the selected features is user defined and second it is time consuming. So, a hybrid algorithm using GRASP and a bio inspired evolutionary algorithm—Binary Grey Wolf Optimisation (BGWO) is proposed. The GRASP algorithm is used for obtaining the optimal concatenated model and is explained in detail in [20]. The second level of the proposed optimisation, the Binary grey wolf optimisation is explained in the following subsection.

4.4.1 Binary Grey Wolf Optimisation (BGWO)

Nature inspired Meta heuristic algorithms are best suited for feature selection which leads to dimensionality reduction. Grey Wolf optimisation technique is a recent swarm-based technique which imitates the leadership ranking and hunting strategy of the Grey Wolf pack [36]. The detailed Binary Grey Wolf Optimisation (BGWO) is given by Algorithm 1.

Algorithm 1 BGWO
INPUT: N - Number of Grey Wolf in the pack
MaxIter- Number of Iterations,
OUTPUT: BestPos - Optimal Grey Wolf binary positions

1: function BGWO (N,MaxIter)
2: Initialise a population of *N* Grey Wolves whose positions is $Pos_{i,j}$ where i = {1, 2,…, N} and j = {1, 2, …, dim}
3: Find the alpha, beta and gamma wolves based on the fitness function given in Eq. (4.11)
4: Initialise a = 2, and calculate A and C as per Eq. (4.8)
5: for iter = 1 to MaxIter do
6: for each Wolf 'i' in the Pack do
7: Update $Pos_{i,:}$ according to Eq. (4.10)
8: end for
9: Update alpha, beta and gamma wolves based on previous step
10: end for
11: BestPos ← $Pos_{alpha,:}$
12: return BestPos
13: end function

COMMENT: rand () produces random number in the range (0,1].

Here the pack is led by social ordering of wolves–alpha, beta, delta and omega. Alpha wolves are the dominant ones and they lead the pack. Beta and delta wolves assist alpha in hunting decisions and omega are the followers. In other words, the best wolf is the alpha followed by beta (second), delta (third) and lastly by omega (others). The encircling of prey during hunting is determined by adjusting the position of the *k*th wolf, X_k with respect to the prey *p* in *i* + 1th iteration given by Eq. (4.8).

$$X_{p,k}(i+1) = X_p - A \times |C \times (X_p - X_k(i))|$$
$$\text{where,} \quad A = 2 \times a \times rand() - a \tag{4.8}$$
$$C = 2 \times rand()$$

and the parameters—*a* is linearly decreased from 2 to 0 for each iteration, *A* and *C* help to converge the algorithm globally and *rand()* is a random number (0,1] generation function. Since alpha, beta and delta are the best wolves that give the best position from the prey, the optimum location of prey is determined by alpha, beta and delta wolves' positions and the positions of all wolves are updated according to Eq. (4.9).

$$X_k(i+1) = \frac{X_{alpha,k}(i) + X_{beta,k}(i) + X_{delta,k}(i)}{3} \tag{4.9}$$

This bio inspired technique has been remodelled for feature selection by Emary et al. using two squashing functions [37]. The role of the squashing function is to retain the population position values as binary. One of them is the sigmoid squash function which helps to maintain the binary input needed, has more potential for the feature selection process in steganalysis and given by Eq. (4.10) is used.

$$BinX_k(i+1) = \begin{cases} 1 & \text{if} \quad sigmoid(X_k(i+1)) \geq rand() \\ 0 & \text{otherwise} \end{cases}$$

$$\text{where,} \; sigmoid(j) = \frac{1}{1 + e^{-10(j-0.5)}} \tag{4.10}$$

The fitness function for BGWO is the selection of the best feature (i.e.) the one with maximum classification accuracy and minimum number of features. So, the fitness function f for classification, is set as in Eq. (4.11)

$$f = \alpha \times Accuracy + \beta \times \frac{|T - L|}{T} \tag{4.11}$$

where, *Accuracy* is the classification accuracy using the features, T is the total number of features and L is the length of the selected features. α and β are the two parameters that determine the quality of the classification and length respectively, where $\alpha = 0.99$ and $\beta = 1 - \alpha$ as in base paper [37].

4.5 Experimental Results and Discussion

4.5.1 Experimental Setup

The goal is to establish a universal low complex steganalytic feature for identifying the commonly used (Traditional LSB) spatial steganography. Bossbase v1.01 [38] images of size 512×512 embedded with five LSB algorithms and eight different payloads are used. The samples of the stego images for one such random cover image for various algorithms and the statistical metrics—Mean Square Error (MSE) and Entropy show the embedding distortion caused are given in Table 4.2.

It can be inferred from Table 4.2 that even with a high volume secret payload of 1.0 bpp (262,144 bits), the stego images are not visually distinguishable from the cover image and amongst themselves. The variations in the measures are also so small which depicts the challenge in identifying the algorithms of same nature using their stego images. The train-test ratio for the experimentation is fixed as 50%, i.e., for each payload, random 500 images of each cover and stego images of each algorithm ($500 + 500 \times 5 = 3000$) are trained using ensemble One Against One (OAO) Logistic Regression (LR) classifier and the remaining unseen 3000 images are tested. The statistics for all the experiments are the median of the statistics collected by repeating the experiment ten times with different random train/test datasets.

Table 4.2 Comparison of the stego images from various algorithms for a payload of 1.0 bpp from its cover image

	Cover image	LSBR stego image	LSBM stego image	LSBMR stego image	LSBR2 stego image	LSBRmod5 stego image
Visual comparison						
MSE	0	0.5003	0.4994	0.3760	3.2641	0.8512
Entropy	6.6207	6.6248	6.6276	6.6263	6.6426	6.5844

4.5.2 Algorithm Detection Using Individual LRP and LDiP Feature Models

The 76 individual feature sets discussed in Sect. 4.2 are trained and tested individually on stego images with a low volume payload of 0.1 bits per pixel (bpp). The Receiver Operating Curve (ROC) plots of the individual group by micro averaging, portray the experimental results in Fig. 4.4 along with Area Under Curve (AUC) measures.

Figure 4.4 shows the excellency of LRP over the LDiP features and rotation variant forms are slightly better than rotation invariant forms. Particularly, considering sign LRP and magnitude LRP, the mLRPs are more contributing than the sLRPs. It is because the considered training set consists of images from both single and multi-bit embedding algorithms. It can be noted that *mLRP* with kernel *k*14 is the best among the proposed models with an accuracy of 58.23%. It is important to note the difference between the accuracy and the AUC shown in ROC plots. This is because in multi-class classification, the number of negative classes is greater than the number of positive classes. The feature models that give the maximum accuracy for other payloads are given in Table 4.3.

As payload increases, the magnitude LDiP feature captured in vertical direction (*90mLDiP*) is better performing because of its spatial multi resolution property. Figure 4.5 shows the individual class ROC plot of both the features.

It can be seen that as payload increases, the order of detection of algorithms becomes different. In low payload (Fig. 4.5a), LSBR2 is easily detected, followed by LSBRmod5, Cover, LSBR, LSBM and lastly by LSBMR, while in high payload (Fig. 4.5b), the order is LSBR2, LSBR, Cover, LSBRmod5, LSBM and LSBMR. Thus, universal active steganalysis with similar algorithms is a true challenge with detection accuracy slightly greater than random choice. This stretches the experiment to move towards the improvement in performance which is sought by optimal concatenation of features.

4.5.3 Algorithm Detection Using Optimally Concatenated LRP + LDiP Features

The concatenation of LRP and LDiP feature models by GRASP is done to improve the detection accuracy of the individual feature models. The optimal solution obtained from experimentation on a low volume payload of 0.1 bpp is the concatenation of features–*LRPri-k3, k4, k15, mLRPri-k1, k3, k7, k8, k9, k11, k12, k13, sLRPri-k2, 90mLRPri* with a dimensionality of 576. An increase of 12.5% detection accuracy is achieved for 0.1 bpp payload and the ROC of the proposed concatenated model (LRP + LDiP) for various payloads is given in Fig. 4.6.

(a) LDiP & LDiP^ri

(b) mLRPs

(c) mLRP^ris

(d) sLRPs

(e) sLRP^ris
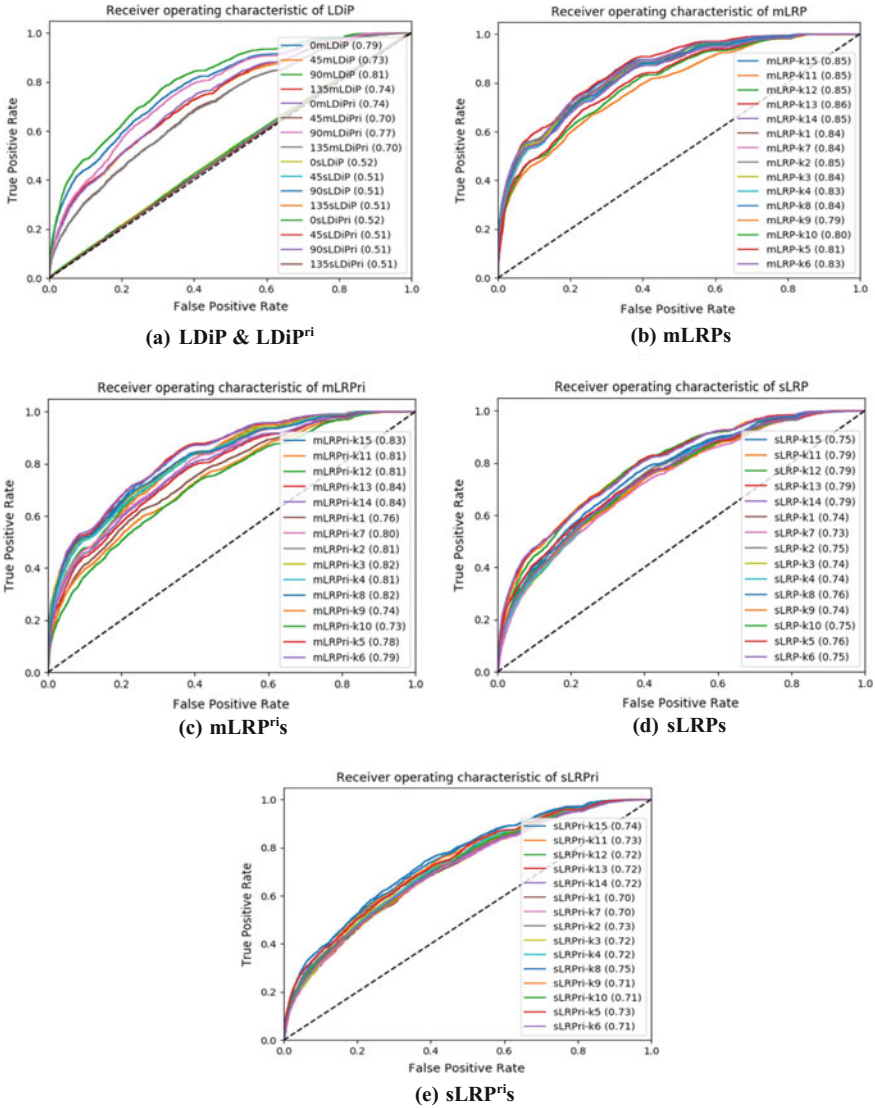
Fig. 4.4 ROC plots for algorithm steganalysis of traditional LSB using LRP and LDiP features

Performance analysis of the proposed LRP + LDiP feature in varying groups: To illustrate the difficulty of the algorithm detection task and its dependence on choice of algorithms chosen for training, three groups of LSB algorithms have been designed–First group G1 consists of images from all the above said algorithms, the second group G2 consists of the cover and LSBR, LSBM, LSBR2 and LSBmod5 stego images (most difficult algorithm removed) and the last group G3 consists of the cover, stego images from LSBR, LSBM and LSBMR algorithms

Table 4.3 Traditional LSB Algorithm detection for various payloads using proposed individual features

Payload (bpp)	Overall accuracy in (%)	Max. feature
0.1	58.23	mLRP-k14
0.2	66.8	mLRP-k1
0.25	70.33	90mLDiP
0.3	73.37	90mLDiP
0.4	79.03	90mLDiP
0.5	82.63	90mLDiP
0.75	88.4	90mLDiP
1.0	90.2	90mLDiP

(only single bit embedding). The ROC plots for these groups embedded with a low volume payload of 0.1 bpp are given in Fig. 4.7.
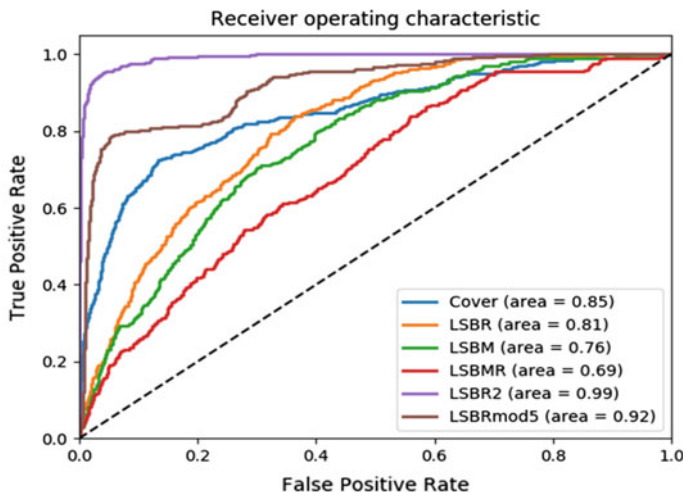
It can be seen from Fig. 4.7 that the most difficult is the group consisting of single bit schemes (G3 accuracy 61.35%, Cover–429, LSBR&LSBMR-281, LSBM-236), followed by the scheme where all algorithms are considered (G1 accuracy 70.73%, Cover-410, LSBR-286, LSBM-226, LSBMR-273, LSBR2-483, LSBRmod5-444). The easiest is the scheme where LSBMR is exempted (G2 accuracy 77.68%, Cover-420, LSBR-294, LSBM-302, LSBR2-482, LSBRmod5-444).

The confusion matrix of the LRP + LDiP feature in algorithm classification for a low volume payload of 0.1 bpp using different groups is given as Table 4.4.
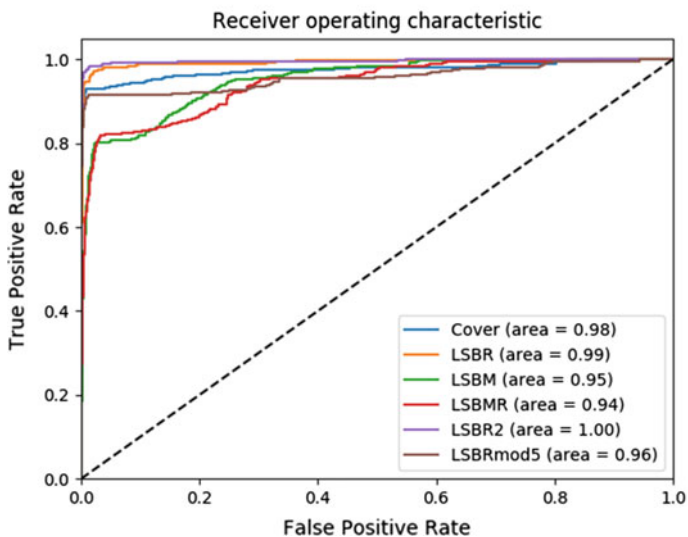
Thus, the choice of the training stego algorithms mainly affects the algorithm detection process and the intermediate group G1 is considered for further experimentation.

Performance analysis of Classifier: To compare the effectiveness of classifier against other classifiers in LSB steganographic algorithm detection, experimentations are carried out with various classifiers on the obtained optimal concatenated feature model of dimensionality 576. Two groups of classifiers are considered. The first group is the simple classifier models. The classifiers considered are Logistic Regression (LR), Naive Bayes, K-Nearest Neighbour (KNN), Linear Support Vector Machine (LinearSVM) and Decision Tree. The second group consists of ensemble classifiers like Random Forest, Extremely Randomised Tree (Extra Tree), Adaboost, Gradient boosting and Bagging with the default base learner and One Against All (OAA) with LR as base learner.

The results tabulated in Table 4.5 show that among simple classifiers, LR provides twice more accurate results than other simple classifiers. However, the ensemble form of LR (OAO) produces 5% more accuracy in low volume of 0.1 bpp than simple LR. Again, among various ensemble classifiers like Trees, Boosting and Bagging, OAO (LR) is better and gives nearly twice more accuracy than the tree based algorithms and about 7% more accuracy than Gradient Boosting, the best among Boosting and Bagging classifiers. Though OAA (LR) is a simpler model compared to OAO, it performs at par only for 1.0 bpp payload. Thus, OAO (LR) is better suited for LSB algorithm classification.

Receiver operating characteristic

True Positive Rate

Cover (area = 0.85)
LSBR (area = 0.81)
LSBM (area = 0.76)
LSBMR (area = 0.69)
LSBR2 (area = 0.99)
LSBRmod5 (area = 0.92)

False Positive Rate

(a) mLRP-k14 for 0.1 bpp

Receiver operating characteristic

True Positive Rate

Cover (area = 0.98)
LSBR (area = 0.99)
LSBM (area = 0.95)
LSBMR (area = 0.94)
LSBR2 (area = 1.00)
LSBRmod5 (area = 0.96)

False Positive Rate

(b) 90mLDiP for 1.0 bpp

Fig. 4.5 ROC plots of the feature for each class of algorithm detection

4.5.4 Algorithm Detection Using Optimised LRP + LDiP Feature Using GB

Though performance of the model has been increased from 58.23 to 70.73% for 0.1 bpp payload, this has indirectly led to the increase of dimensionality from 256

Fig. 4.6 ROC plots of optimised LRP + LDiP features for various payloads of algorithm detection
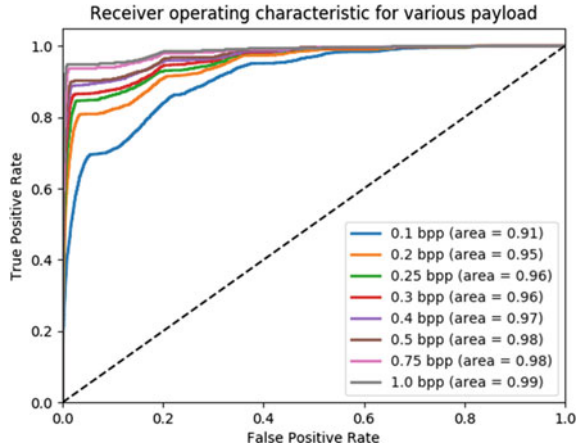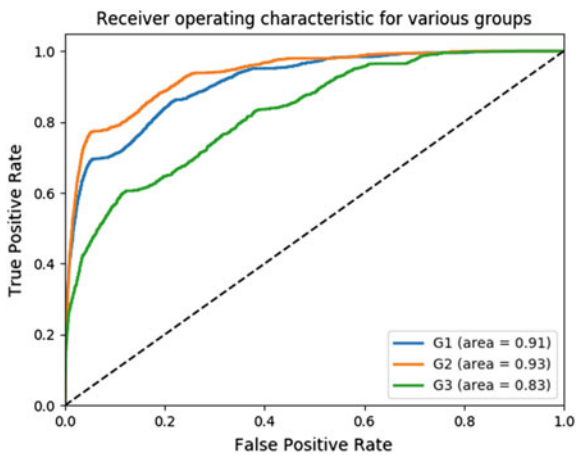
Receiver operating characteristic for various payload

- 0.1 bpp (area = 0.91)
- 0.2 bpp (area = 0.95)
- 0.25 bpp (area = 0.96)
- 0.3 bpp (area = 0.96)
- 0.4 bpp (area = 0.97)
- 0.5 bpp (area = 0.98)
- 0.75 bpp (area = 0.98)
- 1.0 bpp (area = 0.99)

Fig. 4.7 ROC plots of LRP + LDiP for various groups of algorithm detection

Receiver operating characteristic for various groups

- G1 (area = 0.91)
- G2 (area = 0.93)
- G3 (area = 0.83)

to 576. So, the dimensionality of the features obtained are reduced by use of two feature selection algorithms–GRASP-RFE (GR) and GRASP-BGWO (GB) using OAO (LR) classifier. The RFE feature selection method in scikit-learn 0.18.1 package [39] is applied as a dimensionality reducer. The default dimensionality reduction is half of the given features (288). The desired dimensions can be set by the user and are set from 100 to 500 in steps of 100 and the dimension where the best accuracy is obtained is reported. The improved results are tabulated in Table 4.6.

The results show that the GB selection process gives better results than the basic GRASP model. It is able to decrease the dimension by nearly 120 features yet increasing the accuracy to nearly 2% for all payloads. Comparing with GR, GB produces a minimum of 1% increase in accuracy for all payloads and additionally enjoys dynamic feature selection. And most of the results saturate after 20th

Table 4.4 Confusion matrix for algorithm detection of various groups using optimally concatenated LRP + LDiP features in 0.1 bpp payload

Group	Class wise accuracy (%)						Overall accuracy (%)
	Cover	LSBR	LSBM	LSBMR	LSBR2	LSBRmod5	
G1	82	57.2	45.2	54.6	96.6	88.8	70.73
G2	84	58.8	60.4	–	96.4	88.8	77.68
G3	85.8	56.2	47.2	56.2	–	–	61.35

Table 4.5 Algorithm detection using LRP + LDiP with various classifiers and payloads

Classifier	Detection accuracy in percentage for given payload in (bpp)							
	0.1	0.2	0.25	0.3	0.4	0.5	0.75	1.0
Naive Bayes	24.23	34.97	37.6	40.03	42.43	43.83	47.4	50.53
KNN	24.13	39.77	41.63	44.07	47.9	51.1	56.97	61.67
Linear SVM	32.03	65.53	75.7	65.53	70.43	79.57	86.13	92.5
Decision tree	37.0	49.73	52.9	55.43	62.4	65.77	71.5	74.13
Logistic Regression	65.1	80.13	83.83	87.03	89.97	91.47	94.4	**96.07**
Random forest	36.9	52.17	58.27	60.63	64.0	67.73	74.5	77.2
Extra trees	37.33	52.63	54.5	59.0	63.17	66.0	72.5	76.9
AdaBoost	38.77	49.5	54.0	53.97	58.27	60.8	60.4	67.73
Bagging	40.7	56.47	59.7	62.93	67.0	70.9	77.1	81.1
Gradient boosting	48.17	62.97	66.6	71.27	75.23	78.33	85.0	88.1
OAA (LR)	62.07	76.63	82.67	85.2	87.23	90.1	93.5	95.23
OAO (LR)	**70.73**	**82.93**	**86.3**	**88.37**	**90.57**	**91.87**	**94.5**	95.67

iteration of BGWO, thus making employment of bio-inspired algorithm better than any other Meta heuristic method in terms of both time and complexity. The obtained results reinstate the toughness of identification of algorithms in stego images and effectiveness of employment of a bio-inspired algorithm in selecting features for universal algorithm steganalysis.

4.5.5   Comparison with Existing Works

There is a scarcity of literature on identification of algorithms in steganalysis. Further difficulty, is finding literature that works with same stego images of same domain and algorithm (most of the literature on algorithm identification is on JPEG images). So, two types of comparisons are done. First, is the comparison with only existing literature [14] for multi-class classification of spatial LSB algorithms as

Table 4.6 Optimisation of features for LSB Algorithm steganalysis for various payloads using GR and GB

Payload in bpp	Optimisation of features by					
	GRASP		GR		GB	
	Dim	Overall accuracy (%)	Dim	Overall accuracy (%)	Dim	Overall accuracy (%)
0.10	576	70.73	500	70.53	452	**71.6**
0.20	576	82.93	400	83.23	474	**83.9**
0.25	576	86.3	400	85.23	460	**87.23**
0.30	576	88.37	300	88.43	443	**89.3**
0.40	576	90.57	400	90.87	439	**92.0**
0.50	576	91.87	500	91.0	443	**93.2**
0.75	576	94.5	400	94.2	464	**95.2**
1.00	576	95.67	300	95.33	321	**96.43**

such, which employs Subtractive Pixel Adjacency Matrix (SPAM) feature set with their LR classifier. Here, the multi-class classification is done only for a group of LSB variant algorithms—LSBR, LSBM, LSBR2 and LSBRmod5 for a payload of 0.5 bpp to yield an accuracy of 82.3%. Clearly, even a single proposed individual feature *90mLDiP* achieves an accuracy (89.68%) greater than the literature with lesser dimension of 256 compared to 686 of SPAM.

Since the latter comparison method is not complete, a second comparison is done by employing the universal state-of-the-art passive steganalysers for algorithms detection. To achieve this, SPAM [2], Spatial Rich Model (SRM) [16] and Projected SRM (PSRM) [17] features are extracted from our database are used for classification by the same OAO classifier. SPAM features were proposed for Markov chain based steganalysis of spatial domain algorithms, particularly for LSBM. Here the spatial pixel differences between adjacent neighbours of first and second order Markov chains were found and the probability transition matrix of the differences formed the 686 features of SPAM.

The SRM features were formed with the strategy of assembling various diverse noise sub models from various linear and non-linear filters. These noise sub models were formed from the joint PDF of neighbours in quantised noise residuals. This led to a huge dimensional feature (34,671) which could steganalyse both non adaptive scheme and content adaptive steganographic schemes. Later Holub and Fridrich [17] proposed another strategy of statistical representation of diverse noise models other than the joint PDF of neighbours. They suggested the projection of the residuals into a set of random vectors and called it the PSRM feature. These representations were advantageous than the co-occurrence matrix because they were able to capture dependencies over a large number of neighbourhood pixels, flexibly adjust the trade-off between accuracy and dimensionality and select random neighbourhood sizes to provide better, diverged and discriminant features. Though PSRM is a more agile model than SRM, the feature extraction time

Table 4.7 Comparison table for algorithm steganalysis for various payloads using LRP + LDiP GB method

Payload in bpp	Detection accuracy in %					
	Proposed method (LRP + LDiP-GB)		SPAM (Dim 686)	PSRM (Dim 12,870)	SRM (Dim 34,671)	
	Dim	Overall accuracy (%)	Overall accuracy (%)	Overall accuracy (%)	Overall accuracy (%)	
0.10	452	**71.6**	56.2	63.37	69.0	
0.20	474	**83.9**	72.43	80.53	82.47	
0.25	460	**87.23**	77.53	85.17	85.33	
0.30	443	**89.3**	81.33	87.33	88.5	
0.40	439	**92.0**	85.43	90.53	91.4	
0.50	443	**93.2**	88.6	93.2	94.03	
0.75	464	**95.2**	92.3	96.2	96.1	
1.00	321	**96.43**	94.27	97.63	97.27	

complexity of the PSRM model (approximately 672 s for a single image, SRM–5, SPAM–1 and LRP–0.3 s) makes steganalysis using PSRM highly difficult.

Table 4.7 shows the results of this comparison. The proposed method is better than all the existing methods for low volume payloads. In a low volume payload of 0.1 bpp, 71.6% accuracy is reached with a feature dimension of just 452. The proposed features excel SPAM, the designated steganalyser for traditional LSB steganalysis in all payloads. It also surpasses the PSRM and SRM features with at least 2% more accuracy with a diminutive feature nearly 34,000 less features in payloads less than 0.5 bpp. However, for high volume payloads, SRM and PSRM are able to achieve less than 1% increase at the cost of very huge dimensionality. Thus, the proposed features along with the efficient proposed optimisation technique proves to a boon to the steganalysis of algorithm in spatial LSB stego images.

4.5.6   Algorithm Detection in Content Adaptive Algorithms

From the previous sections, it can be seen that algorithm detection is a tough task with the low volume payload on closely related algorithms. In case of content adaptive steganalysis, it tends to be lot tougher with more similarity among very closely related content adaptive algorithms and very low embedding rates. As far as the authors' knowledge, there exists no literature for detecting algorithm among content adaptive stego images. Stego images from three content adaptive algorithms–Highly Undetectable steGanOgraphy (HUGO), Wavelet Obtained Weights (WOW) and Spatial UNIversalWAvelet Relative Distortion (S-Uniward or SW) are considered. All these algorithms are LSBM and content adaptive algorithms and in addition, WOW and SW obtain the embedding distortion from the same domain (wavelets). The stego images are created using random 1000 images of Bossbase

v1.01 with a payload of 0.4 bpp. It is to be noted that though the embedding payload is of 0.4 bpp, the embedding change rates for an image are that of 0.0933, 0.0918 and 0.0703 (HUGO, WOW, SW) (very, very low) which makes steganalysis of content adaptive algorithms with even 0.4 bpp tougher. It is also difficult to identify algorithms with same change rate than with different ones (As seen from Fig. 4.7, G3 was the most difficult). The train-test ratio is maintained at 50:50 and the median of tenfold cross validation result with OAO (LR) classifier is reported.

The proposed features detect content adaptive algorithms with accuracies below 50%, which illustrates the difficulty of the task. The best accuracy is obtained by *sLRP-k15* which offers 35.55% accuracy which is at par with SPAM with twice smaller number of features. However, the SRM and PSRM are better than the individual feature. So, the LRP + LDiP model from previous experimentation is then tested for content adaptive algorithms and the results are tabulated in Table 4.8 along with the existing state-of-the-art steganalysers–SPAM, PSRM and SRM.

Clearly from Table 4.8, it can be inferred that the proposed hybrid optimisation and features perform excellently even in first level of optimisation (GRASP) than the agile PSRM feature model with just 576 features against 12,870. Also, further employment of bio-inspired BGWO helps to increase the performance by 3% with nearly 200 less features. The confusion matrix of the classification is given in Table 4.9. It can be observed that Cover images are better identified followed by HUGO and SW. The most difficult WOW images are the least identified and are mostly misclassified as Cover. A tougher problem of identification of algorithms in content adaptive scenario is thus addressed by a universal feature common to all type of spatial LSB steganographic algorithm and whose performance is improved by the proposed novel hybrid optimisation technique—GRASP-BGWO.

Table 4.8 Detection of content adaptive algorithms for 0.4 bpp payload

Features (dimension)	Accuracy in %
SPAM (686)	35.7
SRM (34,671)	46.5
PSRM (12,870)	44.1
Individual feature mLRP-k15 (256)	35.55
LRP + LDiP-GRASP (576)	45.55
LRP + LDiP-GB (402)	**47.9**

Table 4.9 Confusion matrix for content adaptive algorithm detection using optimal LRP + LDiP–GB features

Embedding algorithm	Classified as			
	Cover	HUGO	WOW	SW
Cover	307	71	54	68
HUGO	76	279	55	90
WOW	157	86	159	98
SW	98	108	80	213

4.6   Conclusion

A low dimensional local steganalytic feature, which is sensitive to the embedding algorithm and that change considerably with payloads, is presented for LSB variant algorithm detection. It was observed that the algorithm detection is highly dependent on the training algorithms, payloads and features. The proposed model excels all existing state-of-the-art steganalysers even in low volume payload. The universal nature of the feature is further established in detecting steganographic algorithms of content adaptive nature. Additionally, the proposed hybrid method of optimisation helps to improve performance by 12–13% with a minimum of 400 features for maximum 6 class classification in spatial LSB images. Thereby, a new low dimensional feature selection using hybrid GRASP-BGWO optimisation is proposed using novel local descriptors for effective universal algorithm steganalysis of spatial LSB images. The future scope is to scale the existing feature models along with the proposed models for much larger number of steganographic algorithms.

References

1. Johnson, N.F.: Steganography software. Available at http://www.jjtc.com/Steganography/tools.html (2012). Accessed 27 Aug 2015
2. Pevny, T., Bas, P., Fridrich, J.: Steganalysis by subtractive pixel adjacency matrix. IEEE Trans. Inf. Forensics Secur. **5**(2), 215–224 (2010)
3. Cogranne, R., Retraint, F.: An asymptotically uniformly most powerful test for LSB matching detection. IEEE Trans. Inf. Forensics Secur. **8**(3), 464–476 (2013)
4. Fridrich, J., Kodovskỳ, J.: Steganalysis of LSB replacement using parity-aware features. In: Proceedings of Fourteenth International Conference on Information Hiding, pp. 31–45. Springer, Heidelberg (2013)
5. Veena, S.T., Arivazhagan, S.: Local descriptor based steganalysis of spatial LSB variant stego images. Presented at the TEQIP II International Conference on Computation, Communication and Innovation, pp. 54–57. ACCET, Karaikudi (2016)
6. Farhat, F., Ghaemmaghami, S.: Towards blind detection of low-rate spatial embedding in image steganalysis. IET Image Process. **9**(1), 31–42 (2015)
7. Yu, J., Li, F., Cheng, H., Zhang, X.: Spatial steganalysis using contrast of residuals. IEEE Signal Process. Lett. **23**(7), 989–992 (2016)
8. Swaminathan, A., Wu, M., Liu, K.J.R.: Digital image forensics via intrinsic fingerprints. IEEE Trans. Inf. Forensics Sec. **3**(1), 101–117 (2008)
9. Bell, G., Lee, Y.-K.: A method for automatic identification of signatures of steganography software. IEEE Trans. Inf. Forensics Sec. **5**(2), 354–358 (2010)
10. Pevnỳ, T., Fridrich, J.: Towards multi-class blind steganalyzer for JPEG images. In: Digital Watermarking, pp. 39–53. Springer, Heidelberg (2005)

11. Pevny, T., Fridrich, J.: Determining the stego algorithm for JPEG images. IEE Proc. Inf. Sec. **153**(3), 77 (2006)
12. Pevny, T., Fridrich, J.: Merging markov and dct features for multi-class jpeg steganalysis. In: Proceedings of SPIE: Security, Steganography, and Watermarking of Multimedia Contents IX, vol. 6505, pp. 650503–650513 (2007)
13. Dong, J., Wang, W., Tan, T.: Multi-class blind steganalysis based on image run-length analysis. In: Ho, A.T.S., Shi, Y.Q., Kim, H.J., Barni, M. (eds.) Proceedings on 8th International Workshop on Digital Watermarking IWDW 2009, Guildford, UK, August 24–26, pp. 199–210. Springer, Berlin (2009)
14. Lubenko, I., Ker, A.D.: Steganalysis using logistic regression. In: IS&T/SPIE Electronic Imaging, p. 78800 K (2011)
15. Zhu, J., Guan, Q., Zhao, X.: Multi-class JPEG image steganalysis by ensemble linear svm classifier. In: Shi, Y.-Q., Kim, H.J., Pérez-González, F., Yang, C.-N. (eds.) 13th International Workshop on Digital-Forensics and Watermarking (IWDW 2014), Taipei, Taiwan, 1–4 Oct 2014. Revised selected papers, pp. 470–484. Springer International Publishing, Cham (2015)
16. Fridrich, J., Kodovsky, J.: Rich models for steganalysis of digital images. IEEE Trans. Inf. Forensics Sec. **7**(3), 868–882 (2012)
17. Holub, V., Fridrich, J.: Random projections of residuals for digital image steganalysis. IEEE Trans. Inf. Forensics Sec. **8**(12), 1996–2006 (2013)
18. Shi, Y.Q., Sutthiwan, P., Chen, L.: Textural features for steganalysis. In: Information Hiding, pp. 63–77 (2013)
19. Lyu, S., Farid, H.: Steganalysis using higher-order image statistics. IEEE Trans. Inf. Forensics Secur. **1**(1), 111–119 (2006)
20. Veena, S.T., Arivazhagan, S.: Quantitative steganalysis of spatial LSB based stego images using reduced instances and features. Pattern Recognition Letters. Available online on Aug 2017. https://doi.org/10.1016/j.patrec.2017.08.016
21. Ker, A.D.: Steganalysis of embedding in two least-significant bits. IEEE Trans. Inf. Forensics Secur. **2**(1), 46–54 (2007)
22. Sharp, T.: An implementation of key-based digital signal steganography. In: International Workshop on Information Hiding, pp. 13–26. Springer, Berlin (2001)
23. Mielikainen, J.: LSB matching revisited. IEEE Signal Process. Lett. **13**(5), 285–287 (2006)
24. Xu, W.-L., Chang, C.-C., Chen, T.-S., Wang, L.-M.: An improved least-significant-bit substitution method using the modulo three strategy. Displays **42**, 36–42 (2016)
25. Lafferty, P., Ahmed, F.: Texture-based steganalysis: results for color images. In: Proceedings on SPIE 5561, Mathematics of Data/Image Coding, Compression, and Encryption VII, with Applications, pp. 145–151 (2004)
26. Gui, X., Li, X., Yang, B.: Steganalysis of LSB matching based on local binary patterns, pp. 475–480 (2014)
27. Ker, A.D., Böhme, R.: Revisiting weighted stego-image steganalysis. Electron. Imaging **2008**, 681905 (2008)
28. Fridrich, J., Kodovskỳ, J., Holub, V., Goljan, M.: Steganalysis of content-adaptive steganography in spatial domain. In: Information Hiding, pp. 102–117. Springer, Berlin (2011)
29. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. Pattern Recogn. **29**(1), 51–59 (1996)
30. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
31. Lu, J., Liu, F., Luo, X.: Selection of image features for steganalysis based on the fisher criterion. Digit. Invest. **11**(1), 57–66 (2014)
32. Akhavan, S., Akhaee, M.A., Sarreshtedari, S.: Images steganalysis using GARCH model for feature selection. Sig. Process. Image Commun. **39**(Part A), 75–83 (2015)
33. Mohammadi, F.G., Abadeh, M.S.: Image steganalysis using a bee colony based feature selection algorithm. Eng. Appl. Artif. Intell. **31**, 35–43 (2014)

34. Ben Brahim, A., Limam, M.: A hybrid feature selection method based on instance learning and cooperative subset search. Pattern Recogn. Lett. **69**, 28–34 (2016)
35. Moradi, P., Gholampour, M.: A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy. Appl. Soft Comput. **43**, 117–130 (2016)
36. Mirjalili, S., Mirjalili, S.M., Lewis, A.: Grey wolf optimizer. Adv. Eng. Softw. **69**, 46–61 (2014)
37. Emary, E., Zawbaa, H.M., Hassanien, A.E.: Binary grey wolf optimization approaches for feature selection. Neurocomputing **172**, 371–381 (2016)
38. Filler, T., Pevný, T., Craver, S., Ker, A. (eds.) Information Hiding, vol. 6958. Springer, Berlin (2011)
39. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830