Moderate embed cross validated and feature reduced Steganalysis using principal component analysis in spatial and transform domain with Support Vector Machine and Support Vector Ma...

Article *in* Multimedia Tools and Applications · August 2022

DOI: 10.1007/s11042-022-13638-w

3 authors:

Deepa D Shankar
Abudhabi University

Nesma Khalil
University

Adresya Suresh Azhakath
Technical University of Denmark

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-022-13638-w

Moderate embed cross validated and feature reduced Steganalysis using principal component analysis in spatial and transform domain with Support Vector Machine and Support Vector Machine-Particle Swarm Optimization

Deepa D. Shankar[1] · Nesma Khalil[2] · Adresya Suresh Azhakath[3]

Received: 17 March 2021 / Revised: 11 April 2022 / Accepted: 1 August 2022

Abstract
The fast evolution of Information and Digital technology had given way for internet to be an effective medium for communication. This has also paved way for data exploitation. Therefore, users must protect their data from misuse. This led to the emergence of security framework like Information Hiding. Steganography and Steganalysis are of the two primary techniques in the field of Information Hiding. Steganography is the science of concealing confidential information, while steganalysis is the art of detecting the existence of that information. The primary goal of this research is to address the general concept of steganalysis, and various breaches associated with it. It involves a blind statistical steganalysis technique that is led in Joint Photographic Experts Group (JPEG) text embedded images by extracting features that illustrate an alteration during an embedding. The images used as embedding medium are uncalibrated and the percentage of the embedding used in this study is 50%. The text embedding is done using various steganographic schemes in the spatial and transform domain. The steganographic schemes considered are Least Significant Bit (LSB) Matching, Least Significant Bit (LSB) Replacement, Pixel Value Differencing and F5. After steganographic embedding of the data, the first order, second order, extended Discrete Cosine Transform (DCT) and

✉ Deepa D. Shankar
sudee99@gmail.com

Nesma Khalil
nesma.khalil@outlook.com

Adresya Suresh Azhakath
adresya.azhakath@gmail.com

[1] Department of Applied Sciences, Abu Dhabi University, Abu Dhabi, UAE

[2] Department of Mathematics, Abu Dhabi University, Abu Dhabi, UAE

[3] Department of Health Technology, Danmarks Teknikse Universitet, Lyngby, Denmark

Published online: 20 August 2022

Markov features are extracted. Then, Principal Component Analysis (PCA) is used as a system for feature dimensionality reduction. Furthermore, the technique of machine learning is incorporated by means of a classifier to identify the stego image and cover image. Support Vector Machine (SVM) and Support Vector Machine with Particle Swarm Optimization (SVM-PSO) are the classifiers examined in this paper for a comparative study. Moreover, the concept of cross-validation is also incorporated in this work. Six dissimilar kernel functions and four diverse samplings are used during classification to check on the effectiveness of the kernels and sampling in classification.

Keywords Steganalysis · LSB matching · LSB replacement · PVD · F5 · Support vector machine · Support vector machine-particle swarm optimization · Kernel · Sampling

1 Introduction

The data available on the internet is quite vast and diverse. The data may comprise of personal details of people from their bank transaction, hospital records to social security number. Hence, the possibility of a data leak or loss, intentionally or otherwise can pose a problem of data breach [18]. Content protection mechanisms were deemed mandatory to create a commercial infrastructure globally [33]. The search for a solution led to the emergence of the field of steganography and hence steganalysis. The technique of steganography enables a message to be hidden in a medium whereas steganalysis help in the identification of the presence of the message. Steganography system is a powerful scheme because it keeps the message undetectable as it is difficult for the user to recognize the presence of the secret message [5]. Unlike cryptography technique, which enables the user to fully decrypt the hidden message using a secret key, Steganography technique does not allow unwarranted user to detect the presence of the hidden message. Thus, using both techniques can provide an extra layer of protection over the other [1].

The term steganography in general would mean ‘secret writing’ [21]. Steganography follows the method of inserting a message into another data so that an unauthorized user cannot recognize the existence of the message. The embedding will be parametrized by a key, and it would be impossible to determine the presence of the data without having the key [34]. The data used for embedding the messages can be images, audios, or videos.

With the advent of technology, the steganographic scheme led to the creation of the steganalysis, which is the technique of identifying any covert communication [61]. Various methods of steganalysis are prevalent in the field of research, but the two fundamental types of steganalysis are targeted steganalysis and blind steganalysis. Targeted steganalysis strategy is used to evaluate a particular embedding process when the steganographic technique is known. This would give excellent results when tested for that steganographic method but would fail in other methods [76]. On the other hand, blind steganalysis is an approach capable of detecting the presence of a secret message even when the steganographic schemes are unknown since it does not require prior information about the embedding algorithm, but it will extract statistical values that may change when an embedding occurs. Thus, blind steganalysis can work for all image formats and embedding techniques. This study employs the use of statistical blind steganalysis and thereby position the steganalyst as passive warden. To be specific, the accuracy measurement of detection of secret messages on images is the main goal of this research. The image format considered for the work in this paper is JPEG format since this image standard is a popular image transmission medium over the internet and hence the work is of relevance in the real-world scenario [32].

The quality of steganography depends on the factor of imperceptibility [2]. Hence, many steganographic algorithms strive to hide the messages by maintaining the imperceptibility. However, the steganography is easily subject to statistical analysis [31]. Steganography can be classified into two techniques; spatial domain and transform domain. In Spatial domain techniques, the encrypted message is hidden inside the image by directly modifying the pixel values of the image, whereas in transform domain techniques, the secret message is embedded in the transform domain coefficients of the image. Therefore, transform domain offers more complex techniques for steganalysis than the spatial domain, because it is resistant to geometric attacks, and it is hard to discover the existence of a message through statistical steganalysis [31]. Considering the above setbacks of steganographic schemes in different domains, this steganalytic work considers algorithms from both spatial and transform domain to have an idea on how effectively the break of steganography can be done in different domains. The transform domain used in this work is Discrete Cosine Transform (DCT). The Blind steganalysis can use the method of supervised learning technique as a means of classification. The supervised learning technique includes two phases-training the data and testing the model. The training phase helps the classifier to distinguish between the original image (cover image) and the image with the hidden message (stego image). This is done when the classifier updates its rules based on prediction and ground truth [4]. In the testing phase, a new image is input to the trained classifier and then detected either as a cover or stego. The supervising learning technique is considered in this study.

This research aims at detecting the hidden message in an embedded message with a certain amount of accuracy, keeping in mind the computational complexity and memory usage. The work applies only to Joint Photographic Experts Group (JPEG) formats with an image size of 256X 256. Each final evaluation is done in blocks of 8X8. Embedding capacities are varied from very small to very heavy embedding to check the accuracy of the classifier. Combination of interblock and intrablock features are used for extraction, to eliminate the drawbacks of each one put separately. Dimensionality reduction and validation scenarios are considered before the feature set is fed into the classifier. Standard image datasets are considered to have a better classification, since the research datasets are "clean" without any changes or noise. The paper considers a classifier and its Particle Swarm Optimization (PSO) variant to determine how each would produce a result under the same conditions. The classification conditions are altered by using different kernel functions and sampling.

The major contribution of the proposed method is as follows:

- The concept of block-based steganalysis has been used for the research. In the block-based approach analysis each image is in terms of 8×8 blocks. Therefore, the result will be more accurate since block-wise analysis is done on a single image as compared to one frame-based analysis.
- JPEG image format is considered for the study. This is because this format is widely used to store or transmit digital images over the internet due to the high compression ratio.
- The study considers two datasets for the work- INRIA holiday dataset for training and UCID dataset for testing.
- Message embedding is done using text.
- The embedding percentage used here is 50%.

- For embedding the message, four different steganographic algorithms such as LSB Replacement, LSB Matching, Pixel Value Differencing and F5 had been used.
- The next step is to extract relevant features from the image. The first order, second order, extended Discrete Cosine Transform (DCT) and markovian features are extracted. The research focuses on the combination feature set to eliminate the intrablock and interblock dependencies.
- After the extraction, redundant features are removed using the technique of dimensionality reduction by means of Principal Component Analysis (PCA).
- Validation of features are done using 10- fold cross validation techniques.
- Supervisory machine learning algorithm along with its optimization counterpart is used for a comparative study in the classification. Classification comparison of the above is done using Support Vector Machine and Support Vector Machine with Particle Swarm Optimization.
- Six different kernels namely radial, dot, polynomial, multiquadratic epanechnikov and ANOVA is used for each classifier. Linear, shuffle, stratified and automatic sampling is done on the data during the classification.
- Finally, classification measurements are done in terms of accuracy and a comparative study is done with each scenario.

The paper is organized as follows. Section 2 discusses the related work. Section 3 focuses on the process, methodology (Section 3.1), dataset (Section 3.2), block-based evaluation (Section 3.3), algorithm (Section 3.4), feature selection (Section 3.5), Extraction of features (Section 3.6), supervised learning (Section 3.7), testing and training set (Section 3.8), classification (Section 3.9), cross validation (section 3.10), dimensionality reduction (section 3.11), kernels (section 3.12) and sampling (section 3.13). Section 4 explains the experimental results and discussion. The research is finally concluded in Section 5.

2 Related work

The main idea of steganalysis is to distinguish between cover images and stego images [53] by defining the medium and its format. Hence, JPEG format is the most common medium used in the steganographic technique due to its low complexity of steganalysis and good performance. Therefore, JPEG format has been considered in this study. Steganalysis can be broadly divided into Targeted and Blind. The targeted steganalysis would target a particular algorithm [54]. On the other end, the blind steganalysis do not have a clue of the steganographic algorithm and hence use the help of features to detect the existence of a message [23].

The block-based steganalytic system; which extract features from the small blocks, yields better result than frame-based steganalytic systems; which extracts features from the entire image [14]. SeongHo Cho et al. proposed a novel system of steganalysis using multi classifier in a block-based system [12]. Kodovsky et al. also formulated a block based steganalytic system to detect YASS [37]. Due to the advantages of the block-based system highlighted in the above papers, the scheme is considered for analysis in this work.

Different steganographic techniques under spatial and transform domain are involved in this research. Comparative studies of small embedding in spatial and transform domain have been done by Adresya et al. [60] and by Deepa et al. [26] which point the better classification of each domain under different combinations of kernels and embedding percentages. Least

Significant Bit (LSB) techniques and Pixel Value Differencing (PVD) is under the spatial domain and F5 is under transform domain. Since the LSB embedding is done in spatial domain, it requires large dimensions to arrive at a conclusion. Zhang et al. introduced a targeted algorithm to detect LSB steganography [79]. However, Veena et al. suggested an algorithm with small parameters by merging local and global features [69]. LSB techniques can be sequential or random. The sequential embedding technique inserts the secret message in sequence from the LSB of the first pixel to the last bit of the message [35], while the random embedding technique incorporates the scattering of the message bits to confuse the steganalysts [78]. Mondal et al. proposed a technique where the cover image and the secret message should be divided into smaller segments. For this, a pseudo random sequence is used to embed the message into the small image segments. Wu and Hwang suggested an improved LSB substitution method to get an enhanced result [73]. Wu and Tsai proposed the concept of Pixel Value Differencing (PVD) where the images are divided into blocks and consecutive pixel differencing is calculated. More data can be embedded in edged areas than in the smooth areas since the difference in the edged areas is greater and hence imperceptible to human vision [15]. Swain proposed an adaptive PVD algorithm in both the horizontal and vertical direction [65]. F5 is an algorithm used in the transform domain. The embedding is performed in a randomized fashion throughout the DCT coefficients. The embedding is done by decreasing the coefficients by 1 [72].

Features are vital in terms of determining the presence of the secret message [36]. Hence, the feature selection is a crucial factor in steganalysis [46]. Pevny suggests CC-PEV features, which are a combination of Markov, First Order and Second Order features, help eliminate the issues faced by interblock and intra block features separately [50]. A comparative study with feature reduction using Principal Component Analysis was done by Deepa [57] and a k-fold cross-validation has been proposed to evaluate the prediction and conduct a comprehensive estimation. Zhang suggested a stratified validation model for face recognition using Support Vector Machine (SVM) [80]. Li proposed a parallel SVM training system using cross-validation [42]. The cross-validation seems to give better accuracy rate of the model with 10-fold cross-validation [70]. PCA is used in this paper for dimensionality reduction when a huge data of unrelated variables is involved [27]. Xuan et al. utilized the concept of PCA on steganalysis of high-dimensional features that is derived from co-occurrence matrix [62]. On the other side, Ma et al. discussed the importance of PCA in feature reduction and they claimed that the performance is well defined in linear systems [45].

Traditional optimization algorithms may have adverse factors like misalignment and noise. This is because the traditional optimizations are fixed at local optima. Hence, a metaheuristic concept came into picture. The metaheuristic methods have been used with success in multiple fields including the scientific research and the whale optimization algorithm is one of them [43]. This optimization algorithm has a balance between exploration and exploitation phases. However, there is a tradeoff between the convergence speed and optimal solution. Whale Optimization Algorithm (WOA) would increase the complexity for high complex problems as well [47]. Another evolutionary algorithm can be Biogeography-based optimization (BBO) algorithm. This algorithm gives better results with medical images [11]. However, the drawback of the BBO algorithm is its inability to exploit solutions and the lack of provision to select the best member from each generation [3].

The Support Vector Machine (SVM) can be a classification technique used for binary classification. The two main principles of SVM are the optimal hyperplane and the kernels [75]. The main advantage of SVM is the kernel trick and the ability to scale to high dimensions. Hence the technique is used in different domains such as geology [30], economics [38] and in education as well [19]. The Support Vector Machine (SVM) and its optimized variant with Particle Swarm Optimization (SVM-PSO) were used in a comparative study by Deepa et al. [58] at 25% embedding rate. The embedding percentage makes a marked difference when the classifiers are being used. Moreover, a comparative study was performed with different types of kernels with SVM-PSO that resulted in a set of percentages that showcased the better classification under different steganographic embedding, the various classifiers and the kernels used in each combination [59]. Raj et al. suggested that the incorporation of PSO into SVM is to optimize the classifier by selecting the best parameters of SVM and automatically utilize the best features to discriminate various classes [51].

3 Process

Recent research had shown how the techniques of machine learning can be effectively used in the domain of steganalysis. The limited literature studies have also opened horizons in which machine learning can be a prime focus. Previous literature also shows that the statistical steganalysis had given promising results when image classification is done using machine learning. The paper aims to remove the pitfalls of previous work and enhance the detection percentage of the analysis. The work aims at moderate 50% text embedding and a comparative study of the classification is done using SVM and SVM-PSO.

3.1 Methodology

As mentioned earlier, the paper makes use of the JPEG images as a basis for steganalysis. The selection of the JPEG format is due to the high ratio of compression and visual image quality [6]. The lossy behavior of the format makes it suitable for data hiding [24]. Lossy compression is used with Huffman entropy coding to encode blocks [77].

A moderate embedding of 50% is used in the paper. Steganalysis is performed on images embedded using both spatial and transform domain algorithms. In this study, the images are divided into small 8×8 blocks and analysis is done in each block. A comparative study of the analysis is done using two supervised learning classifiers namely SVM and SVM-PSO. A block diagram of the architecture is shown in Fig. 1:

3.2 Dataset

The images required for the study were selected from two different datasets that are freely available online for research. The images datasets are INRIA holidays [56] dataset and UCID [55] dataset. These are standard datasets that are publicly available on the internet. Using the standard dataset will assist in validating, benchmarking, and comparing results with previous reports. This would emphasize the transparency of academic research. The INRIA Holidays dataset contains a huge database of high-resolution images with scenery and man-made pictures. There are 500 sets of images and the first one is the query image. UCID dataset is generally used for content retrieval research since it has the ground truth of the image database. In UCID dataset, the images are in the uncompressed form and can be used as a benchmark for compressed image retrieval techniques. The dataset details are given in Table 1.

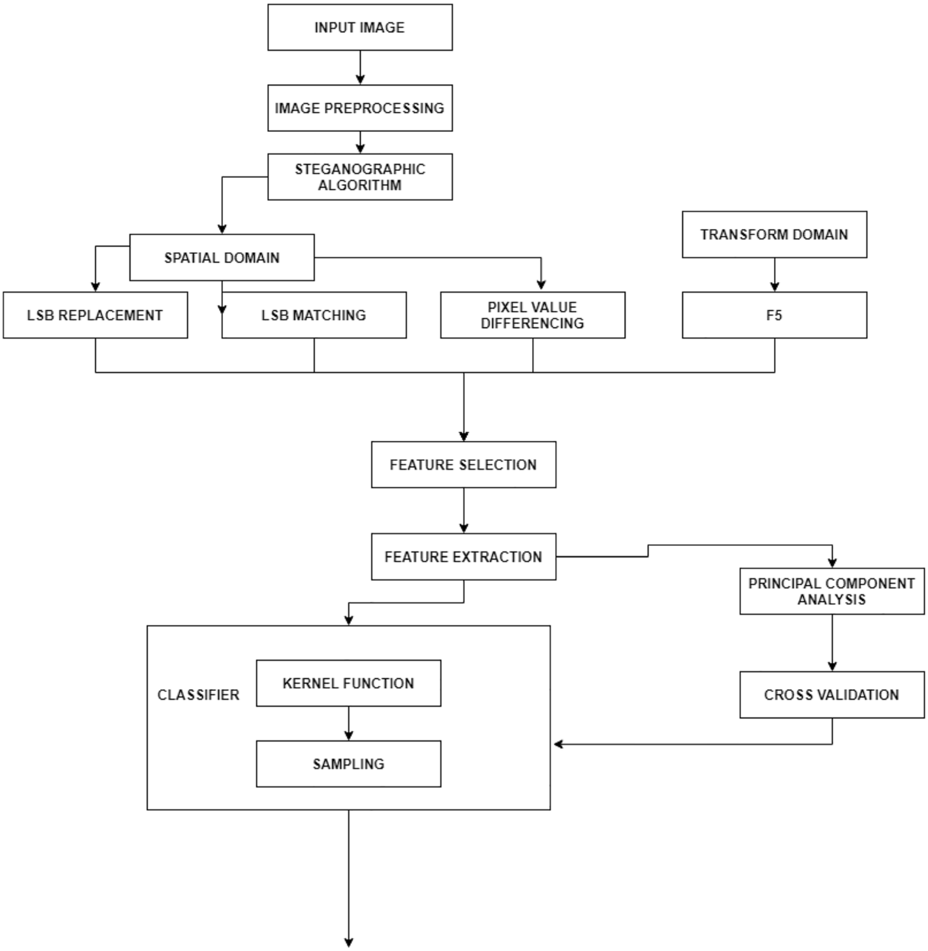

Fig. 1 Block diagram of the architecture

3.3 Block based evaluation

For the steganalysis, images are split into 8 × 8 blocks. Cho et al. proposed the blocked based approach where the image is divided into homogenous blocks. Through experiments, it was found that the block-based images with merged feature set provided great performance results. As per Cho et al. [13], the block-based approach can be used as a universal framework for all types of feature set. This approach would help a classifier even if a single unknown image is given as test data.

Table 1 Description of datasets used in the study

Datasets for training and testing	INRIA Holidays (training)-1500 images UCID (testing) -800 images
Format Standard Size	JPEG 256X256

3.4 Algorithms

The paper focuses on the analysis of images in both spatial and transform domain. The transform domain used in the paper is Discrete Cosine Transform (DCT). Three steganographic algorithms in spatial domain and one algorithm in transform domain were considered for analysis.

3.4.1 Least Significant Bit Replacement (LSBR)

LSB data embedding is a common steganographic technique of hiding secret messages into DCT coefficients. This is used in the spatial domain as well. The algorithm substitutes the LSB value of each pixel of the cover image with a single aspect of the hidden data. If the value of the LSB aligns with the value of the secret message, the algorithm will leave the pixel value unchanged. If not, the LSB will be altered by either expanding or decrementing even or odd pixels, respectively. The advantage of this embedding technique is its simplicity of implementation.

3.4.2 Least Significant Bit Matching (LSBM)

This steganographic algorithm is also known as ±1 embedding. Unlike LSB Matching, the algorithm tends to randomly add or subtract a single pixel value from the cover image that does not match with LSB value of the secret message bit. The advantage of this technique is that it has a good embedding ability, and that the modification will be almost imperceptible to the human eye. The algorithm produces a significant space embedding capacity of approximately 1 bit per pixel per color component.

3.4.3 Pixel Value Differencing (PVD)

In the Pixel Value Differencing (PVD) steganographic technique [66], the cover image is divided into several non-overlapping blocks consisting of two consecutive pixels. From each segment, the difference value, which can range from 0 to 255, is retrieved. If the difference value is low, this ensures that the block is sited in a smooth area so only a limited fraction of secret data can be embedded. Otherwise, it means that the block is in a textured area and thus more secret bits can be hidden. The quantization range table is designed with values ranging from 0 to 255 and the number of bits hidden depends upon the quantization range table.

3.4.4 F5

F5 technique is a transform domain embedding algorithm that provides a high embedding capacity without necessarily affecting the security [22]. To embed a secret message using F5 algorithm, instead of replacing the LSBs of the quantized DCT coefficients with the message bits, the absolute value of the non-zero DCT coefficients is reduced by one.

The embedding procedure of F5 steganography is as follows: [7].

1. Obtain the image
2. Measure the table of quantification directly linked to the quality factor and condense the image
3. Compute the estimated embedding capacity

4. Select a PRNG seed which would consider a random path to insert the message.
5. The message is segmented into blocks of x bits inserted randomly into 2^x-1 coefficient

3.5 Feature selection

Features play a prominent role in statistical steganalysis. The feature should be identified to perform better during the classification phase. The selected features should be relevant and need to project a change when embedding is done on the image. The features make use of three relevant features – First order, Second order and Markov features. The First Order features are intrablock features. The Second order and Markov features are interblock features. The combination of features was selected in the research to eliminate the pitfalls of both. The first order features are the global histogram, dual histogram, and individual histogram. The individual histogram shows distribution of DCT coefficients that varies for different modes. The histogram makes use of the 5 lower frequency points. The global histogram considers all DCT coefficients for the 8 × 8 block. Consider that the stego image is characterized by an array of DCT coefficient $d_k(i,j)$, where i and j are coefficients, and where k is the block. The global histogram of the whole of 64 k blocks can be signified by G_r where $r = X,\ldots,Y$, $X = min_{k,i,j}\ (d_k(i,j))$, $Y = max_{k,i,j}\ (d_k(i,j))$. Some steganographic algorithms preserve these statistics, hence dual histogram is generated to provide additional statistics. For a fixed coefficient value, the dual histogram [50] captures the distribution of the given coefficient value between different DCT modes. It is represented by the following formula.

$$g_{ij}^{d} = \sum_{k=1}^{B} x\left(d, d_{k(i,j)}\right) \tag{1}$$

where B is the block number.

When the DCT coefficients of the different blocks are independent, a certain embedding scheme that conserves the statistics of the first order shall be recorded. However, natural images exhibit correlation greater than 8 pixels. Hence, the features that capture the interblock dependency are used. Variation, Blockiness and Markovian features are the interdependency features [50]. Variation is represented by the following formula

$$V = \frac{\sum_{i,j=1}^{8} \sum_{k=1}^{|I_r|-1} \left| d_{Ir(k)(i,j)} - d_{Ir(k+1)(i,j)} \right| + \sum_{i,j=1}^{8} \sum_{k=1}^{|I_c|-1} \left| d_{Ic(k)(i,j)} - d_{Ic(k+1)(i,j)} \right|}{|I_r| + |I_c|} \tag{2}$$

where I_r and I_c are the index vector blocks when inspected by columns and rows. Thus, variation calculates distinctions in coefficients in just the same position in successive blocks, initially row-wise, and then column-wise. The embedding changes may increase the discontinuities along the 8 × 8 boundaries. Thus, Blockiness has been introduced into the features. Blockiness is represented by the following formula

$$B_\alpha = \frac{\sum_{i=1}^{\left|(M-1)/8\right|} \sum_{j=1}^{N} \left| x_{8i,j} - x_{8i+1,j} \right|^{n} + \sum_{i=1}^{\left|(N-1)/8\right|} \sum_{j=1}^{M} \left| x_{8i,j} - x_{8i+1,j} \right|^{n}}{N\left\lfloor (M-1)/8 \right\rfloor + M\left\lfloor (N-1)/8 \right\rfloor} \tag{3}$$

M and N are the dimensions of the image. Blockiness calculates the difference between the pixel values at boundaries of each block. It is calculated for both row and column and then summated. The co-occurrence matrix is used to determine the likelihood function of combinations of adjacent coefficients.. Cooccurrence can be represented by the following formula

$$C_{st} = \frac{\sum_{k=1}^{|I_r|-1}\sum_{i,j=1}^{8}\delta\left(s, d_{i,(k)}(i,j)\right)\delta\Big(t, d_{i,(k+1)}(i,j) + \sum_{k=1}^{|I_c|-1}\delta\left(s, d_{i,(k)}(i,j)\right)\delta\Big(t, d_{i,(k+1)}(i,j)}{|I_r| + |I_c|} \tag{4}$$

The Markov feature set [50] is generated by computing four difference matrices from a quantized 2D array of a JPEG image taken along four directions-horizontal, vertical, diagonal major, and diagonal minor directions. The Markovian features are calculated by the following formulas

$$F_h(j,k) = F(j,k) - F(j+1,k) \tag{5}$$

$$F_v(j,k) = F(j,k) - F(j,k+1) \tag{6}$$

$$F_d(j,k) = F(j,k) - F(j+1,k+1) \tag{7}$$

$$F_m(j,k) = F(j+1,k) - F(j,k+1) \tag{8}$$

where *F(j,k)* is a quantized DCT coefficient, $j \in [1, R_j - 1]$, $k \in [1, R_k - 1]$, R_j is the 2D array size of the JPEG in horizontal direction, R_k is the array size in vertical direction, F_h, F_v, F_d, F_m are the difference arrays in horizontal, vertical, major, and minor diagonals, respectively.

From this, the actual transition probability matrices are created.

$$M_h(j,k) = \frac{\sum_{x=1}^{N-2}\sum_{y=1}^{M}\partial_{F_h}(x,y), j\partial_{F_h}(x+1,y), v}{\sum_{x=1}^{N-1}\sum_{y=1}^{M}\partial_{F_h}(x,y), u} \tag{9}$$

$$M_v(j,k) = \frac{\sum_{x=1}^{N}\sum_{y=1}^{M-2}\partial_{F_v}(x,y), j\partial_{F_v}(x,y+1), v}{\sum_{x=1}^{N}\sum_{y=1}^{M-1}\partial_{F_v}(x,y), u} \tag{10}$$

$$M_d(j,k) = \frac{\sum_{x=1}^{N-2}\sum_{y=1}^{M-2}\partial_{F_d}(x,y), j\partial_{F_d}(x+1,y+1), v}{\sum_{x=1}^{N-1}\sum_{y=1}^{M-1}\partial_{F_d}(x,y), u} \tag{11}$$

$$M_m(j,k) = \frac{\sum_{x=1}^{N-2}\sum_{y=1}^{M-2}\partial_m(x+1,y), j\partial_m(x,y+1), v}{\sum_{x=1}^{N-1}\sum_{y=1}^{M-1}\partial_{F_m}(x,y), u} \tag{12}$$

3.6 Extraction of features

Features are extracted from images embedded by messages using the steganographic algorithms LSB Matching, LSB Replacement, PVD and F5. All extracted images are normalized to remove the redundant values. For the first order features, normalization is done in the range (−5..5). Hence, the global histogram has 11 functionalities. The dual histogram will be 11 functionalities for the 9 lower AC modes which would sum to 99 functionalities. The individual histogram has 11 functionalities in 5 DC modes which sum to 55 functionalities. The second order features would consider the central elements of the range [−2,2] X [−2,2]. The central elements are considered because it would hold the relevant information for steganalysis. This is attributed to the reason that the DCT coefficients are based on a generalized Gaussian distribution centered across zero. Hence the second order co-occurrence would have 25 functionalities, the variance will have 1 functionality and the blockiness would have 2 functionalities. The markov features also use the central part of the matrix [−4,4] in horizontal, vertical and the two diagonal parts. This would sum up to 324 functionalities. However, the average of the 4 matrices is used which would be calculated to 81 functionalities. The above mentioned functionals are tabulated in the Table 2.

3.7 Supervised learning

Machine learning was a very powerful medium of classification. The supervised learning in machine learning had given efficient results than its unsupervised counterpart. Supervised learning helped to have a mechanism to map the input samples with the output [8].The results were of better accuracy and reliability in supervised learning.

3.8 Testing and training sets

As mentioned earlier in the paper, 1479 images from INRIA Holidays database are used to train the classifier. The test dataset is a group of 800 images from UCID dataset. Different standard datasets are involved in the study to avoid extreme overfitting. The analysis will yield better result if the images are from a different dataset.

Table 2 Table of features to be extracted

Features	method	no. of functionalities
First order (INTRABLOCK)	Global Histogram	11
	Individual Histogram	55
	Dual Histogram	99
Second order (INTERBLOCK)	Co-occurrence	25
	Blockiness	2
	Variance	1
markov (INTERBLOCK)		81
total		274

3.9 Classification

Classification is performed after the features have been retrieved. The images are classified into two segments, based on their characteristics [29]. Thus, the images in this paper are classified either as a cover image or a stego image [9].

3.9.1 Support Vector Machine

The Support Vector Machine by Vapnik [68] is a pattern recognition classifier used for binary classification [44]. The classifier generates a hyperplane or a group of hyperplanes for classification. The aim of the classifier is to find out the optimal hyperplane that produces the lowest distance to the training samples. This would help to maximize the boundary of the training set of data. The margin can be increased in between the classes. Data points next to the decision-making surface are known as support vectors. If the margin is much less than close to the data point and the decision surface, then noise will be introduced into the system which restricts proper classification. Hence, the margin should be optimal for a proper classification. SVM shows excellent performance for nonlinear problems [20]. This classifier is great for high dimensional data because of its reduced sensitivity to dimensionality and robustness to noise [40]. The classification of SVM is as shown in the figure below. Figure 2.

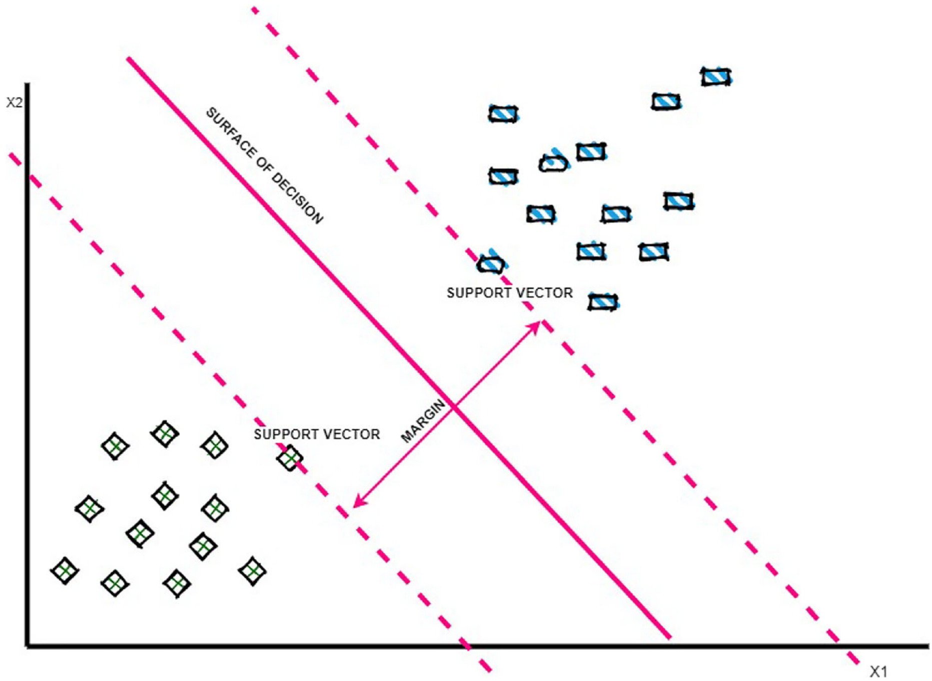

Fig. 2 Classification of Support Vector Machine

3.9.2 Support Vector Machine -Particle Swarm Optimization (SVM-PSO)

The Support Vector Machine with Particle Swarm Optimization (SVM-PSO) classifier is used in the paper to perform a comparative analysis with SVM. SVM-PSO was developed in 1995 [49] and has become increasingly popular due to its flexibility, ease of implementation and low computational requirement [16]. The PSO is an evolutionary model of computing based on swarm intelligence [41]. In PSO, the group of birds is called a particle, which will form a population in a D-dimensional feature space [25].

The algorithm for parameter selection of PSO in SVM [17] will be as follows:

1. Initialize the parameters. They can be maximum iteration number, population, etc. Determination of the initial velocity v_i^0 and location x_i^0 must be done. X = rand ($-x_{id\,max}$,x_{idmax}) will be determined next, where rand() is a random function in the range [0,1]. Set gen = 1.
2. A set of c and r is generated randomly. Each selected kernel function and parameters are considered as an individual SVM parameter. Thus, the evolution starts.
3. Calculate the new position v_i^0 and x_i^0 of the i^{th} particle. Input x_i into the SVM model for forecasting. The f_i, k-fold cross-validation, is then calculated.
4. The offspring generation is done. The best global value is generated from the equation given below

$$v_{id}^{t+1} = \omega v_{id}^{t} + c_1.r_1.\left(p_{id} - x_{id}^{t}\right) + c_2.r_2.\left(p_{gd} - x_{id}^{t}\right) \tag{13}$$

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \tag{14}$$

where $Vi = (v_{i1},v_{i2},v_{i3}....v_{iD})$ is the velocity of the i-th particle, $Pi = (p_{i1},p_{i2},p_{i3}....p_{iD})$ is the top location of this particle. The ideal position of swarm is $P_g = (p_{g1},p_{g2},p_{g3}....p_{gD})$. When the i-th particle is at the t-th iteration, x_{tid} and y_{tid} are the d-th position and velocity. c_1,c_2,r_1,r_2 are random numbers which have range of values from 0 to 1 and ώ is the inertial weight of PSO.

$$\text{Set } b = b + 1 \tag{15}$$

when b is equal to the maximum number of iterations, the circulation will be stopped, thus obtaining the optimal SVM parameters.

5. The PSO assists with optimization and thus increases accuracy when paired with SVM. PSO was considered in the research with SVM.

3.10 Cross validation

The cross-validation techniques are a general scheme to estimate the prediction error of any supervised learning classifier. Numerous preparation and testing are conducted. During the training process, the training set is divided into k equally populated folds. The training is done on k-1 folds and evaluated on the remaining fold. This is repeated k times, using every fold for error estimation.

In this work, k-fold cross-validation has been performed. Here, k stands for the number of validations and N is the number of samples in the x dataset. Hence, cross-validation is like bootstrapping. Ten-fold cross-validation is used in this work and hence 10 sets of data are used.

3.11 Dimensionality reduction using Principal Component Analysis

One of the main problems in steganalysis is reducing the dimensions and features of data. Dimensionality reduction can improve the differences between the cover image and stego image so that better accuracy can be obtained during classification [63]. One of the efficient methods for linear dimensionality reduction is Principal Component Analysis (PCA). In the framework of PCA, an eigenvector denotes the axis or the direction. The resultant of eigenvalue represents the variance with the eigenvector. Increase in the eigenvalue, the variance will increase along with the eigenvector. It may be inadequate to pick features as per the largest ranking of PCA's eigenvalue for steganalysis [48].

3.12 Kernels

The kernel function is used in the classifier to convert a non-linear decision surface to a linear surface. The robustness of the classifier depends on the parameter of the kernel functions [71]. Six different kernels are considered in the research here.

Dot kernel is the most common kernel in the classifier. Dot kernel is the product of internal variables j and k. The dot kernel is characterized as

$$z\,(j,k) = j^{*}k \tag{16}$$

where z is the dot kernel. The dot kernel is in vogue now, in relation to SVM [10].

The radial kernel is defined as

$$z(a,b) = \exp\left(-\gamma\sum_{j=1}^{p}\left(a_{ij}-b_{ij}\right)^{2}\right) \tag{17}$$

γ Tuning parameter that accounts for the decision boundary's smoothness and governs the model's variance.

z radial kernel, a and b are variables. This kernel is used when there is no knowledge of the data beforehand [39].

The polynomial kernel is used to match the data with the normalized training data [74]. The polynomial kernel is defined as

$$z\,(j,k) = \left(j^{*}\,k+1\right)\,l \tag{18}$$

where l is the degree of the polynomial, j and k are variables, and z is the polynomial kernel.

Multiquadric is a classic example of non-positive definite kernel [64].

The multiquadric kernel is defined as:

$$z\,(i,j) = \left(\|\,i-j\,\|2+k^{2}\right)\,0.5 \tag{19}$$

where k is a constant, z is the multiquadric kernel, and i and j are variables.

The Epanechnikov kernel can be defined as:

$$k(c) = \frac{3}{4}\left(1-c^2\right) for\ |c| \leq 1 \tag{20}$$

where k is the Epanechnikov kernel and c is variable [67].

ANOVA kernel is defined as:

$$k(a,b) = \sum_{k=1}^{n} exp\left(-\sigma\left(a^k-b^k\right)^2\right) \tag{21}$$

where k is the ANOVA kernel, a, b, and k are variables, n is the number of images.

ANOVA is known to perform better in multidimensional problems [28].

3.13 Sampling

Sampling determines the spatial resolution of an image and uses it to analyze and modify pixels. Four different samplings are considered for research in the paper. Stratified sampling is a type of sampling where the pixels are divided into smaller groups known as strata [52]. Then a sample is taken randomly from each group for the analysis. The entire image will be taken and missing important features will be less. Linear sampling is used to solve the problem of inverse diffraction in acoustics. The linear sampling method (LSM) is a basic and reliable way of image shaping. The shuffle sampling takes place by shuffling the pixels randomly and the pattern will be generated with good distribution in each dimension. This is also called a padding approach. Automatic sampling is the process of cataloguing the co-ordinate values automatically. This sampling converts the images into a digitalized format. The continuous data is sent to digitalize the image.

4 Experimental results and discussion

A comparative classification result analysis is completed by means of three steganographic schemes in the spatial domain namely LSB Matching, LSB Replacement, PVD and F5 in the DCT domain.

4.1 Results of classification of images using Support Vector Machine

As discussed earlier, the research is done using a moderate embedding percentage of 50 in the dataset. Different algorithms in spatial and transform domain had been used for the analysis.

When SVM was used as a classifier, the following results were available. From the result in Table 3, LSB replacement, LSB matching, PVD and F5 have mostly the same results for all kernels in linear sampling. The only change is seen in F5 algorithm for the dot and ANOVA kernels. It was found from the observations and evaluation in Fig. 3 that the results for the spatial domain schemes are the same. For linear sampling, the classification was very low. A better classification is specified by ANOVA kernel in shuffled sampling for the transform domain steganographic scheme F5.

Table 3 Classification results of classification of JPEG images using Support Vector Machine as the classifier

		Linear	Shuffle	Stratified	Automatic
LSB Replacement					
	Dot	15.91	**54.73**	52.01	52.01
	Radial	15.91	26.02	26.45	26.45
	Polynomial	15.91	53.91	52.79	52.79
	Multiquadric	15.91	49.62	50	50
	Epanechnikov	15.91	26.89	27.12	27.12
	ANOVA	15.91	52.82	49.03	49.03
LSB Matching	Dot	15.91	**54.73**	52.86	52.86
	Radial	15.91	26.02	26.52	26.52
	Polynomial	15.91	53.91	52.86	52.86
	Multiquadric	15.91	49.62	50	50
	Epanechnikov	15.91	26.69	27.26	27.26
	ANOVA	15.91	52.68	49.44	49.44
PVD	Dot	15.91	**54.73**	52.86	52.86
	Radial	15.91	26.02	26.52	26.52
	Polynomial	15.91	53.91	52.86	52.86
	Multiquadric	15.91	49.62	50	50
	Epanechnikov	15.91	26.69	27.26	27.26
	ANOVA	15.91	52.68	49.44	49.44
F5	Dot	79.23	96.95	96.95	96.95
	Radial	15.91	49.07	50.37	50.37
	Polynomial	15.91	96.28	96.88	96.88
	Multiquadric	15.91	48.62	50	50
	Epanechnikov	15.91	49.07	50.37	50.37
	ANOVA	81.01	**97.84**	97.69	97.69

4.2 Results of classification of images using Support Vector Machine-Particle Swarm Optimisation

Unlike previous results, linear sampling in Table 4 has different classification rates for different kernels and schemes. The ANOVA kernel gives a better classification except in linear sampling.

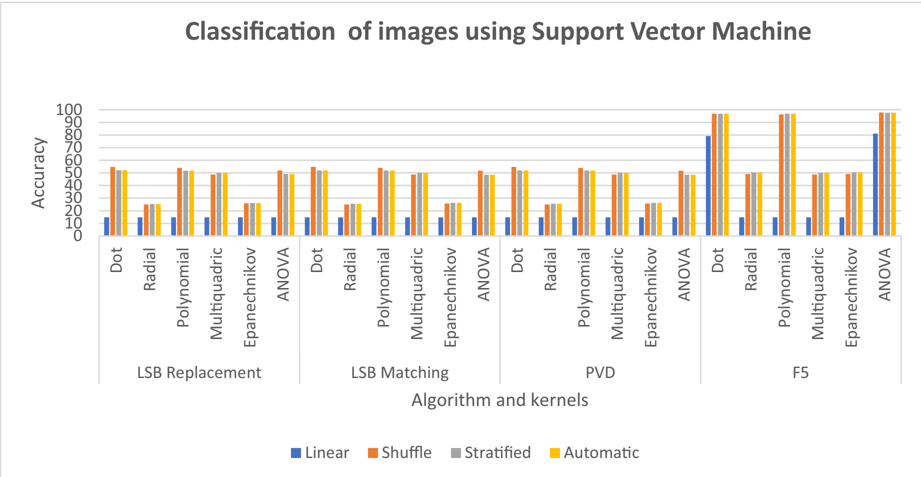

Fig. 3 Result of classification of images using Support Vector Machine for 50%

Table 4 Results of classification of JPEG images using Support Vector Machine-Particle Swarm Optimisation as the classifier

LSB Replacement		Linear	Shuffle	Stratified	Automatic
	Dot	56.62	49.73	50	50
	Radial	20.92	36.91	35.93	35.93
	Polynomial	**65.08**	52.34	51.18	51.18
	Multiquadric	61.07	51.97	51.26	51.26
	Epanechnikov	16.23	33.49	32.8	32.8
	ANOVA	23.38	53.54	51.49	51.49
LSB Matching	Dot	55.62	48.62	50	50
	Radial	21.07	35.74	35.04	35.04
	Polynomial	**65.08**	51.23	50.07	50.07
	Multiquadric	63.18	50.97	50.26	50.26
	Epanechnikov	16.19	32.91	31.29	31.29
	ANOVA	23.38	53.72	51.41	51.41
PVD	Dot	55.73	48.73	50	50
	Radial	21.4	35.63	34.63	34.63
	Polynomial	**65.18**	51.34	50.18	50.18
	Multiquadric	63.25	50.97	50.33	50.33
	Epanechnikov	16.16	33.21	32.51	32.62
	ANOVA	23.16	53.61	54.32	54.32
F5	Dot	56.22	48.62	50.22	50.22
	Radial	44.83	55.03	50.37	50.37
	Polynomial	65.3	49.22	50.37	50.37
	Multiquadric	63.59	51.23	50.15	50.15
	Epanechnikov	19.58	55.1	53.12	53.12
	ANOVA	52.12	92.63	**92.86**	**92.86**

The overall results in Fig. 4 are better than the classification results for uncalibrated images using SVM. For LSB replacement, ANOVA gave the best result with shuffled sampling. On the contrary, PVD gave better results with polynomial kernels on shuffled sampling, followed by stratified and automatic sampling. The better classification rate given in F5 algorithms is for ANOVA kernel with shuffled sampling.

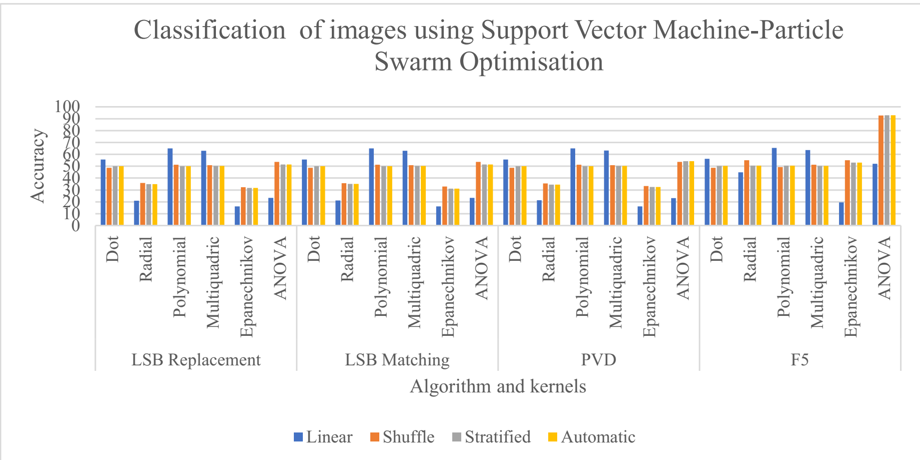

Fig. 4 Result of classification of images using Support Vector Machine-Particle Swarm Optimisation for 50%

Table 5 Classification of uncalibrated images with cross-validation using Support Vector Machine for 50%

		Linear	Shuffle	Stratified	Automatic
LSB Replacement					
	Dot	10.69	54.76	**54.97**	**54.97**
	Radial	5.96	11.41	11.05	11.05
	Polynomial	10.99	53.42	53.78	53.78
	Multiquadric	10.01	48.12	49.94	49.94
	Epanechnikov	6.13	12.45	12.18	12.18
	ANOVA	9.46	46.43	48.11	48.11
LSB Matching	Dot	10.48	54.79	**54.85**	**54.85**
	Radial	5.99	11.44	11.05	11.05
	Polynomial	11.02	53.6	53.72	53.72
	Multiquadric	10.01	48.12	49.94	49.94
	Epanechnikov	6.13	12.72	12.36	12.36
	ANOVA	8.31	47.95	48.99	48.99
PVD	Dot	10.57	54.7	**54.85**	**54.85**
	Radial	5.95	11.41	11.02	11.02
	Polynomial	10.98	53.39	53.81	53.81
	Multiquadric	10	48.12	49.94	49.94
	Epanechnikov	6.13	12.54	12.39	12.39
	ANOVA	8.04	46.93	49.05	49.05
F5	Dot	93.09	98.09	98.06	98.06
	Radial	10	48.57	50.3	50.3
	Polynomial	92.38	97.23	97.08	97.08
	Multiquadric	10	48.12	49.94	49.94
	Epanechnikov	10	48.57	50.3	50.3
	ANOVA	92.97	**98.21**	98.03	98.03

4.3 Results of classification of uncalibrated images with cross-validation using Support Vector Machine

LSB replacement in Table 5 has radial and Epanechnikov kernels giving low classification for all steganographic schemes. ANOVA give a better classification rate with different sampling.

The spatial domain steganographic algorithm showed the same pattern of classification for all the kernels and sampling as in Fig. 5. The dot kernel gives better classification for the

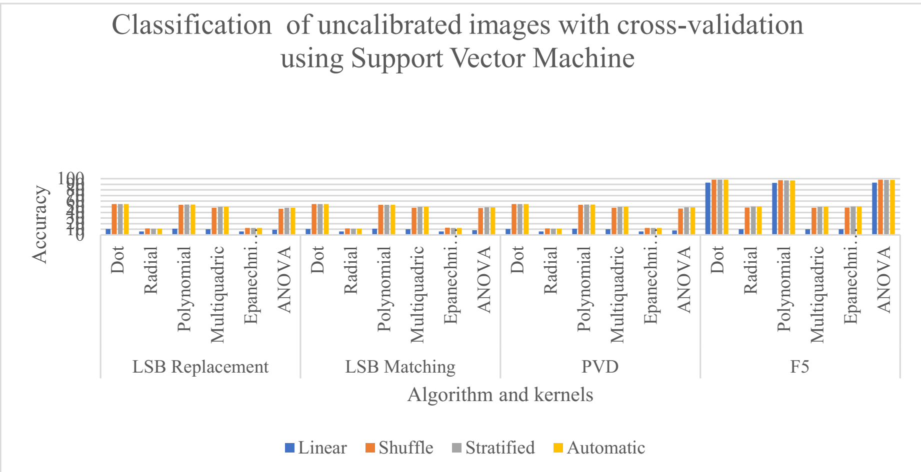

Fig. 5 Result of classification of uncalibrated images with cross-validation using Support Vector Machine for 50%

spatial domain schemes for its stratified and automatic sampling. The F5 algorithm works differently. The radial and Epanechnikov kernels give low classification rate for linear sampling, but better classification rate for the other samplings. The excellent classification is given by ANOVA in its shuffled sampling.

4.4 Results of classification of uncalibrated images with cross-validation using Support Vector Machine-Particle Swarm Optimisation

In LSB replacement in Table 6, the radial and Epanechnikov kernels generally give low classification ranging from for all sampling. The ANOVA kernel in F5 has the classification ranging from 85.5% to 93.51% with different sampling.

The spatial domain steganographic schemes work in a similar pattern as in Fig. 6. The radial and Epanechnikov kernels give low classification rate for all sampling. The ANOVA gives a classification rate range from 51% to 55% with its shuffled, stratified, and automatic sampling. Meanwhile, the F5 algorithm gives a better classification rate for the stratified and automatic sampling.

4.5 Results of classification of uncalibrated images with Principal Component Analysis using Support Vector Machine

The linear sampling in Table 7 for almost all schemes and kernels has a low classification. The multiquadratic and ANOVA kernel have a moderate classification percentage.

Table 6 Classification of uncalibrated images with cross-validation using Support Vector Machine-Particle Swarm Optimisation for 50%

LSB Replacement		Linear	Shuffle	Stratified	Automatic
	Dot	22.12	49.85	49.97	49.97
	Radial	18.79	28.02	28.5	28.5
	Polynomial	28.25	49.58	50.6	50.6
	Multiquadric	**69.74**	51.91	51.91	51.91
	Epanechnikov	11.43	22.16	22.34	22.34
	ANOVA	51.18	53.54	53.34	53.34
LSB Matching	Dot	31.61	48.87	50.06	50.06
	Radial	18.19	28.08	28.56	28.56
	Polynomial	41.9	50.24	50.63	50.63
	Multiquadric	**69.74**	51.97	52.03	52.03
	Epanechnikov	10.93	22.42	22.78	22.78
	ANOVA	16.19	53.66	58.27	58.27
PVD	Dot	42.05	49.14	50.54	50.54
	Radial	18.94	27.4	28.59	28.59
	Polynomial	41.54	49.82	50.39	50.39
	Multiquadric	**70.4**	51.85	51.7	51.7
	Epanechnikov	11.76	22.48	21.77	21.77
	ANOVA	36.54	54.56	53.45	53.45
F5	Dot	20.51	49.05	51.07	51.07
	Radial	45.89	55	54.47	54.47
	Polynomial	34.55	49.46	50.42	50.42
	Multiquadric	67.18	51.43	50.95	50.95
	Epanechnikov	31.23	56.2	56.46	56.46
	ANOVA	85.5	92.64	**93.51**	**93.51**

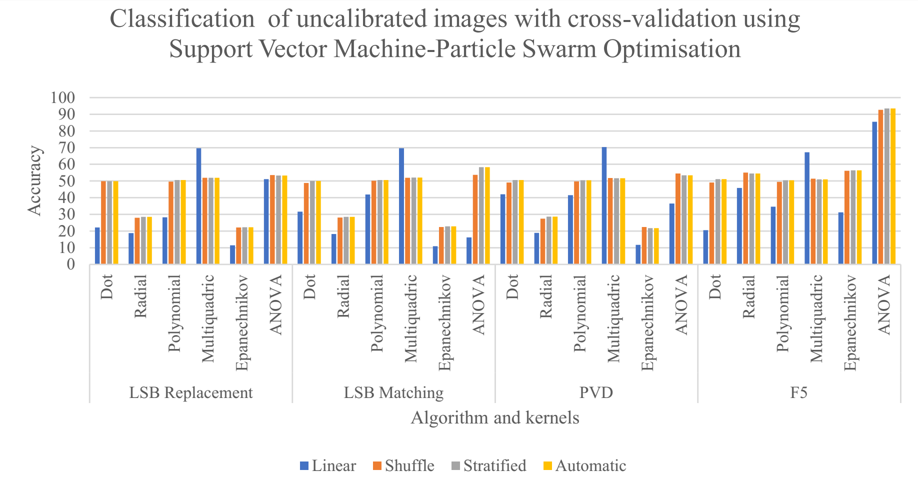

Fig. 6 Result of classification of uncalibrated images with cross-validation using Support Vector Machine-Particle Swarm Optimisation for 50%

The dot and ANOVA kernels have an excellent rate of percentage. LSB replacement and LSB matching follow the same pattern as in Fig. 7. The best classification is found for ANOVA with shuffled sampling.

Table 7 Classification of uncalibrated images with Principal Component Analysis using Support Vector Machine for 50%

LSB Replacem*ent*		Linear	Shuffle	Stratified	Automatic
	Dot	14.89	**54.73**	51.26	51.26
	Radial	14.89	24.94	26.41	26.41
	Polynomial	14.89	52.94	51.12	51.12
	Multiquadric	14.89	48.62	51.12	51.12
	Epanechnikov	14.89	25.91	26.34	26.34
	ANOVA	14.89	45.35	44.87	44.87
LSB Matching	Dot	14.89	**54.5**	51.19	51.19
	Radial	14.89	24.87	26.49	26.49
	Polynomial	14.89	53.02	51.12	51.12
	Multiquadric	14.89	48.62	50	50
	Epanechnikov	14.89	25.99	26.34	26.34
	ANOVA	14.89	39.61	40.33	40.33
PVD	Dot	14.89	51.04	**54.65**	**54.65**
	Radial	14.89	25.02	25.45	25.45
	Polynomial	14.89	52.79	51.19	51.19
	Multiquadric	14.89	48.92	50	50
	Epanechnikov	14.89	25.84	26.41	26.41
	ANOVA	14.89	44.42	40.55	40.55
F5	Dot	76.47	95.23	96.21	96.21
	Radial	14.89	49.59	52.01	52.01
	Polynomial	19.58	67.54	70.01	70.01
	Multiquadric	14.89	48.62	50	50
	Epanechnikov	14.89	49.07	50.37	50.37
	ANOVA	79.6	**96.35**	95.98	95.98

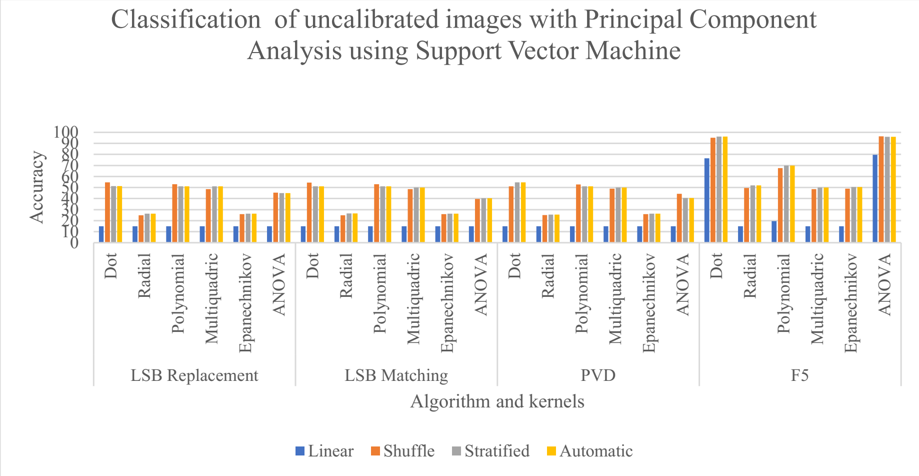

Fig. 7 Result of classification of uncalibrated images with Principal Component Analysis using Support Vector Machine for 50%

4.6 Results of classification of uncalibrated images with Principal Component Analysis using Support Vector Machine-Particle Swarm Optimisation

In LSB replacement for Table 8, the radial and Epanechnikov kernels have a low classification The ANOVA, polynomial, and dot kernels have a classification ranging from 86.31 to 92.56.

Table 8 Classification of uncalibrated images with Principal Component Analysis using Support Vector Machine-Particle Swarm Optimisation for 50%

LSB Replacement		Linear	Shuffle	Stratified	Automatic
	Dot	14.4	54.12	52.88	52.88
	Radial	19.86	21.35	21.53	21.53
	Polynomial	32.77	54.92	51.39	51.39
	Multiquadric	**79.05**	45.78	47.62	47.62
	Epanechnikov	19.86	22.24	22.52	22.52
	ANOVA	33.96	43.1	39.29	39.29
LSB Matching	Dot	16.09	55.11	52.88	52.88
	Radial	19.86	21.35	21.43	21.43
	Polynomial	32.47	54.92	50.99	50.99
	Multiquadric	**79.15**	45.68	47.92	47.92
	Epanechnikov	19.86	22.24	22.72	22.72
	ANOVA	34.66	42.2	43.45	43.45
PVD	Dot	14.9	55.01	53.77	53.77
	Radial	19.86	21.35	21.53	21.53
	Polynomial	32.97	54.92	53.77	53.77
	Multiquadric	**78.65**	45.88	47.82	47.82
	Epanechnikov	19.86	22.14	22.52	22.52
	ANOVA	34.36	42.01	41.47	41.47
F5	Dot	62.66	91.96	**92.56**	**92.56**
	Radial	19.86	51.44	50.69	50.69
	Polynomial	71.5	90.76	90.87	90.87
	Multiquadric	38.43	7.85	7.54	7.54
	Epanechnikov	19.86	51.44	50.5	50.5
	ANOVA	40.52	87.29	86.31	86.31

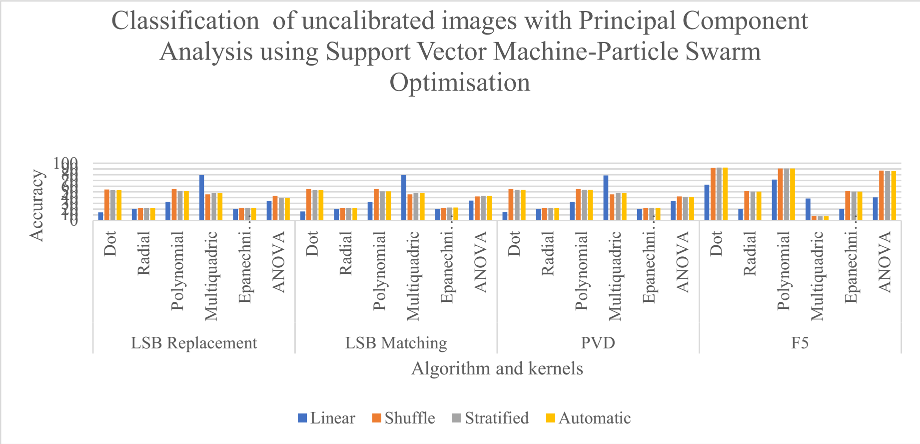

Fig. 8 Result of classification of uncalibrated images with Principal Component Analysis using Support Vector Machine-Particle Swarm Optimisation for 50%

The spatial domain algorithms work in a similar pattern in Fig. 8. The radial and Epanechnikov kernels have comparatively low classification results. The polynomial kernel shows relatively better results. However, the high results are shown by the multiquadratic kernel in linear sampling. The results with F5 algorithm are different from its spatial steganographic counterparts. The dot, polynomial, and ANOVA kernels give better results with F5. However, excellent results are given by the dot kernel in stratified and automatic sampling.

4.7 Results of classification of uncalibrated images with cross-validation and Principal Component Analysis using Support Vector Machine

For linear sampling in Table 9, most of the kernels in all the schemes give a low rate of classification. The F5 algorithm has a good classification for dot and ANOVA kernels for all samplings.

The spatial domain schemes follow the same pattern in the classification percentages as in Fig. 9. The linear sampling generally gives a low classification rate. The radial and Epanechnikov kernels are exhibiting continuous low classification rates for all sampling and all kernels. The best classification is, however, given by ANOVA for its shuffled sampling.

4.8 Results of classification of uncalibrated images with cross-validation and Principal Component Analysis using Support Vector Machine-Particle Swarm Optimisation

For the spatial domain schemes in Table 10, the radial and Epanechnikov kernels have low classification The multiquadratic for all the schemes have moderate classification percentages.

The good classification is given by the dot kernel. In LSB replacement, LSB matching, and PVD, the radial and Epanechnikov kernels give a low classification for all its sampling as in Fig. 10. The ANOVA kernel has increased classification, but not much to be considered. However, the excellent classification is given by the dot kernel in its shuffled sampling.

From the above result it can be concluded that the algorithm in the transform domain give a better classification percentage, mostly with the stratified and automatic sampling. The spatial

Table 9 Classification of uncalibrated images with cross-validation and Principal Component Analysis using Support Vector Machine

LSB Replacement		Linear	Shuffle	Stratified	Automatic
	Dot	12.81	54.23	**54.85**	**54.85**
	Radial	5.96	11.29	11.08	11.08
	Polynomial	11.52	52.62	52.8	52.8
	Multiquadric	10.01	48.12	49.94	49.94
	Epanechnikov	6.19	12.75	12.48	12.48
	ANOVA	7.44	34.99	35.94	35.94
LSB Matching	Dot	12.33	53.87	**54.73**	**54.73**
	Radial	5.96	11.32	11.14	11.14
	Polynomial	11.52	52.53	52.98	52.98
	Multiquadric	10.01	48.12	49.94	49.94
	Epanechnikov	6.19	12.72	12.75	12.75
	ANOVA	7.65	35.32	38.45	38.45
PVD	Dot	12.6	53.96	**54.79**	**54.79**
	Radial	5.96	11.35	11.11	11.11
	Polynomial	11.52	52.62	52.86	52.86
	Multiquadric	10.01	48.12	49.94	49.94
	Epanechnikov	6.13	12.66	12.57	12.57
	ANOVA	7.71	36.6	36.15	36.15
F5	Dot	92.47	96.69	96.69	96.69
	Radial	10.01	48.57	50.48	50.48
	Polynomial	57.59	79.15	79.33	79.33
	Multiquadric	10.01	48.12	79.33	79.33
	Epanechnikov	10.01	48.51	50.3	50.3
	ANOVA	91.75	**96.99**	96.96	96.96

domain algorithms do give good results with various sampling. However, the transform domain algorithm gives better results.

The moderate embedding scenario can be compared as a real-life scenario since the embedding usually involves images or text messages that need to be communicated covertly. If an image needs to be transmitted, it would utmost embed up to 50% of the cover image. As

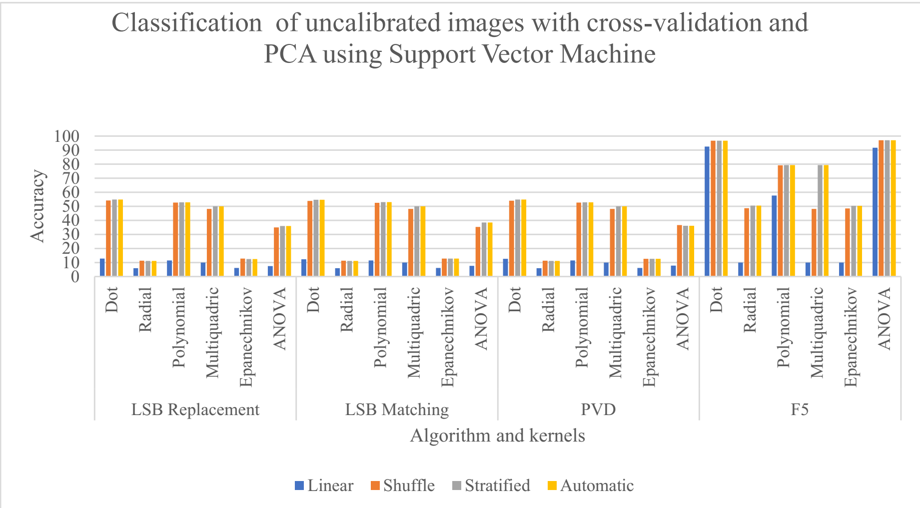

Fig. 9 Result of classification of uncalibrated images with cross-validation and Principal Component Analysis using Support Vector Machine for 50%

Table 10 Classification of uncalibrated images with cross-validation and Principal Component Analysis using Support Vector Machine-Particle Swarm Optimisation for 50%

LSB Replacement		Linear	Shuffle	Stratified	Automatic
	Dot	53.66	54.44	**54.76**	**54.76**
	Radial	5.98	11.11	11.38	11.38
	Polynomial	48.32	53.69	54.73	54.73
	Multiquadric	45.27	45.86	46.07	46.07
	Epanechnikov	6.19	11.97	12.33	12.33
	ANOVA	22.71	39.4	40.98	40.98
LSB Matching	Dot	53.93	54.26	**55.06**	**55.06**
	Radial	5.99	11.14	11.38	11.38
	Polynomial	49.79	54.71	53.81	53.81
	Multiquadric	45.21	45.8	45.89	45.89
	Epanechnikov	6.16	12	12.3	12.3
	ANOVA	19.86	39.73	40.74	40.74
PVD	Dot	53.57	**54.53**	54.2	54.2
	Radial	5.99	11.29	11.47	11.47
	Polynomial	50.39	53.43	54.41	54.41
	Multiquadric	45.12	45.89	45.98	45.98
	Epanechnikov	6.16	12.06	12.42	12.42
	ANOVA	20.49	40.44	40.74	40.74
F5	Dot	91.54	**93.12**	93.03	93.03
	Radial	10.01	49.37	50.98	50.98
	Polynomial	55.84	73.65	65.96	65.96
	Multiquadric	6.49	6.97	7.09	7.09
	Epanechnikov	10.01	49.25	50.6	50.6
	ANOVA	78.76	86.6	87.31	87.31

per the research, the transform domain algorithm can be used to get better results in the identification of a hidden message than a spatial algorithm. The choice of sampling would further benefit the classification percentages.

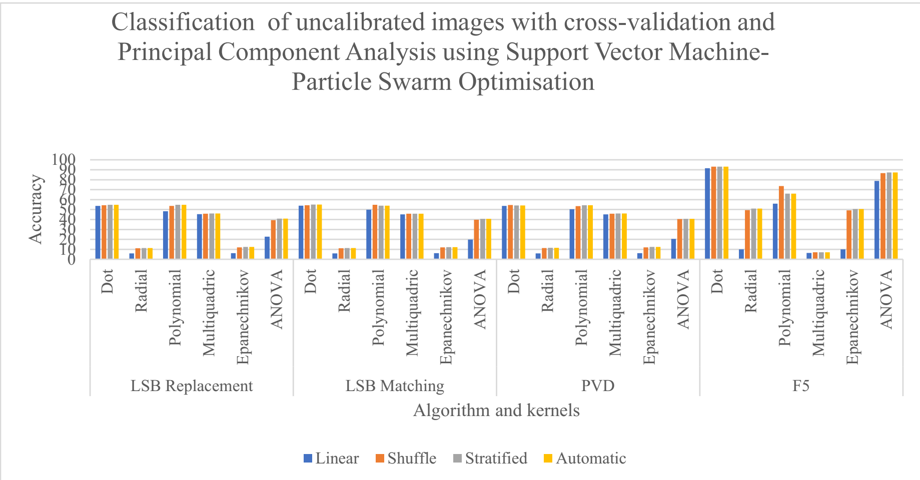

Fig. 10 Result of classification of uncalibrated images with cross-validation and Principal Component Analysis using Support Vector Machine-Particle Swarm Optimization for 50%

5 Conclusion and future scope

In this article, statistical steganalysis is used in JPEG images with 50% embedding. Various characteristics susceptible to embedding are used for the study. The concept of cross-validation had been incorporated in the research. Principal Component Analysis had been used for dimensionality reduction. First order, second order and features from Markov are used. The steganographic patterns used are the spatial domain with LSB Substitution, LSB Matching and PVD and the Discrete Cosine Transform domain with F5. For systematic review, SVM and SVM-PSO were developed to examine the picture as a cover or a stego. The training images are a compilation of 1500 images from the INRIA Holiday dataset and 800 images from the UCID image dataset are gathered for evaluation. Six separate kernels and four types of samplings are used. The findings prove that embedding using the transform domain steganographic scheme provides a different degree of classification than steganographic strategies in the spatial domain. For SVM, the ANOVA kernel provide a great performance with shuffled sampling for F5. With SVM-PSO, for F5 approximation, the ANOVA kernel with stratified sampling shows a better result. Supplementary process can be conducted with different spatial and transform domain algorithms. In addition, a comparative analysis can be completed to examine the consequences of different percentage of embedding during the classification. As a future scope, different formats of the transmission medium may be used to check on the effect of the various medium during embedding and subsequent classification. Recent deep learning classification techniques can also be considered for the classification in the future research. As for statistical extraction of features, a set of hybrid feature, based on their quality of identification, can be used. For high dimensional features, the concept of dimensionality reduction can also be introduced.

Data Availability Data and material can be made available upon request.

Declarations

Ethics approval The paper follows the journal ethics.

References

1. Aditit S (2016) Security and Information Hiding based on DNA steganography. Int J Comput Sci Mobile Comput 5(3):827–832
2. Al-Omari ZY, Al-Taani AT (2017) Secure LSB steganography for colored images using character-color mapping. In: 2017 8th International Conference on Information and Communication Systems (ICICS). IEEE, pp 104–110
3. Ammu PK, Sivakumar KC, Rejimoan R (2013) Biogeography-based optimization-a survey. Int J Electron Comput Sci Eng 2(1):154–160
4. Avcibas I, Memon N, Sankur B (2003) Steganalysis using image quality metrics. IEEE Tran Image Process 12(2):221–229
5. Bas TDP (2018) Natural steganography in JPEG Compressed images. In: Proc IS&T, Electronic Imaging, Media Watermarking, Security, and Forensics, San Francisco
6. Bhasin V, Bedi P (2013) Steganalysis for JPEG images using extreme learning machine. In: 2013 IEEE International Conference on Systems, Man, and Cybernetics. IEEE, pp 1361–1366
7. Briffa JA, Schaathun HG, Wahab AWA (2009) Has F5 really been broken?
8. Castelli (2019) Supervised learning: classification. Encycl Bioinforma Comput Bio: 342–349

9. Chaeikar SS, Ahmadi A (2019) Ensemble SW image steganalysis: a low dimension method for LSBR detection. Signal Process Image Commun 70:233–245. https://doi.org/10.1016/j.image.2018.10.004
10. Chen GY, Bhattacharya P (2006) Function dot product kernels for support vector machine. In: 18th International Conference on Pattern Recognition (ICPR'06). IEEE, vol 2, pp 614–617. https://doi.org/10.1109/ICPR.2006.586
11. Chen Y, He F, Li H, Zhang D, Wu Y (2020) A full migration BBO algorithm with enhanced population quality bounds for multimodal biomedical image registration. Appl Soft Comput 93:106335. https://doi.org/10.1016/j.asoc.2020.106335
12. Cho S Wang J Kuo C-CJ, Cha B-H (2010) Block-based image steganalysis for a multi-classifier. In: IEEE International Conference on Multimedia and Expo
13. Cho S et al (2010) Block-based image steganalysis: Algorithm and performance evaluation. J Vis Commun Image Represent 24:1679–1682. https://doi.org/10.1109/ISCAS.2010.5537499
14. Cho S, Cha BH, Wang J, Kuo CCJ (2011) Performance study on block-based image steganalysis. In: 2011 IEEE International Symposium of Circuits and Systems (ISCAS). IEEE, pp 2649–2652. https://doi.org/10.1109/ISCAS.2011.5938149
15. Da-Chun TW-HW (2003) A steganographic method for images by pixel value differencing. Pattern RecognLett 24:1613–1626
16. Demidova ENL (2016) The SVM classifier based on the modified particle swarm optimization. Int J Adv Comput Sci Appl
17. Du Jinglin et al (2017) A Prediction of Precipitation Data Based on Support Vector Machine and Particle Swarm Optimization (PSO-SVM) Algorithms. Algorithms 10:57. https://doi.org/10.3390/a10020057
18. Edwards B, Hofmeyr S, Forrest S (2016) Hype and heavy tails: a closer look at data breaches. J Cybersecurity 2(1):3–14
19. Fahmi S, Purnamawati L, Shidik GF, Muljono M, Fanani AZ (2020) Sentiment analysis of student review in learning management system based on sastrawi stemmer and SVM-PSO. In: 2020 International Seminar on Application for Technology of Information and Communication (iSemantic). IEEE, pp 643–648.https://doi.org/10.1109/iSemantic50169.2020.9234291
20. Foody GM, Mathur A (2004) A relative evaluation of multiclass image classification by support vector machines. IEEE Trans Geosci Remote Sens
21. Fridrich J (2009) Steganography in digital media: principles, algorithms, and applications. Cambridge University Press
22. Fridrich J, Goljan M, Hogea D (2002) Steganalysis of JPEG images: Breaking the F5 algorithm. In International Workshop on Information Hiding. Springer, Berlin, Heidelberg, pp 310–323. https://doi.org/10.1007/3-540-36415-3_20
23. Fridrich J, Goljan M, Hogea D, Soukal D (2003) Quantitative steganalysis of digital images: estimating the secret message length. Multimedia Syst 9:288–302. https://doi.org/10.1007/s00530-003-0100-9
24. Fridrich J, Tomas P, Kodovsky J (2007) Statistically undetectable JPEG steganography: Dead ends challenges, and opportunities. In: Proc of ACM Workshop on Multimedia and Security, pp 3–14. https://doi.org/10.1145/1288869.1288872
25. Garcia NPJ et al (2016) A hybrid PSO optimized SVM−based model for predicting a successful growth style of the Spirulina platensis from raceway experiments data. Elsevier J Computational App Math, pp 293–303
26. Gireeshan MG, Shankar DD, Azhakath AS (2021) Feature reduced blind steganalysis using DCT and spatial transform on JPEG images with and without cross validation using ensemble classifiers. J Ambient Intell Humanized Comput
27. Han MJ (2012) Data mining: concepts and techniques. Elsevier
28. Hofmann T, Scholkopt B, Smola AJ (2008) Kernel Methods in Machine Learning. Ann Stat 36(3):1171–1220
29. Hou X et al (2017) Combating highly imbalanced steganalysis with small training samples using feature selection. J Vis Commun Image Represent: 49. https://doi.org/10.1016/j.jvcir.2017.09.016
30. Huang Y, Zhao L (2018) Review on landslide susceptibility mapping using support vector machines. Catena 165:520–529
31. Hussain M, Wahab AWA, Idris YIB, Ho AT, Jung KH (2018) Image steganography in spatial domain: a survey. Signal Processi Image Commun 65:46-66
32. Jin Z, Feng G, Ren, Y, Zhang X (2020) Feature extraction optimization of JPEG steganalysis based on residual images. Signal Process 170:107455
33. Kang JS, You Y, Sung MY (2007) Steganography using block-based adaptive threshold. In: 2007 22nd international symposium on computer and information sciences. IEEE, pp 1–7
34. Kaur S, Bansal S, Bansal RK (2014) Steganography and classification of image steganography techniques. In: International Conference on Computing for Sustainable Global Development (INDIACom)

35. Ker AD (2007) Steganalysis of embedding in two least-significant bits. IEEE Trans Inform Forensics Secur 2(1):46–54
36. Ker AD, Bas P, Böhme R, Cogranne R, Craver S, Filler T, Pevný T (2013) Moving steganography and steganalysis from the laboratory into the real world. In: Proceedings of the first ACM workshop on Information hiding and multimedia security, pp 45–58
37. Kodovsky J, Penvy T, Fridrich J (2010) Modern steganalysis can detect YASS. Proceedings of SPIE - The Int Soc Opt Eng 7541:754102. https://doi.org/10.1117/12.838768
38. Kouziokas GN (2020) A new W-SVM kernel combining PSO-neural network transformed vector and Bayesian optimized SVM in GDP forecasting. Eng Appl Artif Intell 92:103650. https://doi.org/10.1016/j.engappai.2020.103650
39. Kuo BC, Ho HH, Li CH, Hung CC, Taur JS (2013) A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification. IEEE J Sel Top Appl Earth Obs Remote Sens 7(1):317–326
40. Lanckriet GR, De Bie T, Cristianini N, Jordan MI, Noble WS (2004) A statistical framework for genomic data fusion. Bioinformatics 20(16):2626–2635. https://doi.org/10.1093/bioinformatics/bth294
41. Li X, Wang J (2007) A steganographic method based upon JPEG and particle swarm optimization algorithm. Inf Sci 177:3099–3109
42. Li Q et al (2013) Parallel multitask cross validation for Support Vector Machine using GPU. J Parallel Distrib Comput: 293–302
43. Luo J, He F, Yong J (2020) An efficient and robust bat algorithm with fusion of opposition-based learning and whale optimization algorithm. Intell Data Anal 24:581–606. https://doi.org/10.3233/IDA-194641
44. Lyu S, Farid H (2003) Detecting hidden messages using higher-order statistics and support vector machines. Inf Hiding: 2578
45. Ma XY (2018) Selection of rich model Steganalysis features based on decision rough set α-positive region reduction. IEEE transactions on circuits and Systems for Video Technology
46. Miche Y et al (2007) Extracting relevant features of steganographic schemes by feature selection techniques. Third Wavila Challenge, Saint-Malo, France
47. Mohammed HM, Umar SU, Rashid TA (2019) A systematic and meta-analysis survey of whale optimization algorithm. Comput Intell Neurosci:25. https://doi.org/10.1155/2019/8718571
48. Mudrov M (2005) Principal component analysis in image processing for image compression
49. Networks RJ (1995) Particle swarm optimization. In: Proceedings of IEEE international conference on neural Networks
50. Pevny T, Fridrich J (2007) Merging Markov and DCT features for multi-class JPEG steganalysis. In Security, steganography, and watermarking of multimedia contents IX. SPIE, vol 6505, pp 28–40.https://doi.org/10.1117/12.696774
51. Raj, S, Ray, KC (2017) ECG signal analysis using DCT-based DOST and PSO optimized SVM. IEEE Trans Instrume Meas: 1–9. https://doi.org/10.1109/TIM.2016.2642758
52. Rezk E, Awan Z, Islam F, Jaoua A, Al Maadeed S, Zhang N, Rajpoot N (2017) Conceptual data sampling for breast cancer histology image classification. Comput Biol Med 89:59–67
53. Sajedi H (2016) Steganalysis based on steganography pattern discovery. J Inf Secur Appl 30:3–14
54. Schaathun H (2012) Machine learning in image Steganalysis, John Wiley and Sons
55. Schaefer G, StitchM (2003) UCID: An uncompressed color image database. In: Storage and retrieval methods and applications for multimedia 2004. SPIE, vol 5307, pp 472–480.https://doi.org/10.1117/12.525375
56. Schmid HJMDC (2008) Hamming embedding and weak geometry consistency for large scale image search - extended version
57. Shankar DD (2020) Impact of features selected by principal component analysis in feature based steganalysis in calibrated and non-calibrated images. Int J Psychosoc Rehabil 6(24):4226–4243
58. Shankar DAA (2020) Minor blind feature based Steganalysis for calibrated JPEG images with cross validation and classification using SVM and SVM-PSO. Multimed Tools Appl:4073–4092
59. Shankar DD, Azhakath AS (2019) Steganalysis of minor embedded JPEG image in transform and spatial domain system using SVM-PSO. IEEE International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), Dubai, United Arab Emirates
60. Shankar DD, Azhakath AS (2021) Small embed cross-validated jpeg steganalysis in spatial and transform domain using SVM. Advances in Machine Learning and Computational Intelligence. Algorithms for Intel systemsligent
61. Shinder L, Cross M (Eds) (2008) Chapter 12 - Understanding cybercrime prevention. In: Scene of the Cybercrime (2nd Edn) Syngress, pp 505–554. https://doi.org/10.1016/B978-1-59749-276-8.00012-1
62. Shlens J (2014) A tutorial on principal component analysis

63. Silva CCD (2017) Principal component analysis (PCA) as a statistical tool for identifying key indicators of nucear power plant cable insulation degradation. Iowa State University
64. Souza R (2010) Kernel functions for machine learning applications
65. Swain G (2016) Adaptive pixel value differencing steganography using both vertical and horizontal edges. Multimed Tools Appl 75:13541–13556. https://doi.org/10.1007/s11042-015-2937-2
66. Tseng HW, Leng HS (2013) A steganographic method based on pixel-value differencing and the perfect square number. J Appl Math 2013. https://doi.org/10.1155/2013/189706
67. Utkin LV, Chekh AI, Zhuk YA (2016) Binary classification SVM-based algorithms with interval-valued training data using triangular and Epanechnikov kernels. Neural Netw 80:53–66. https://doi.org/10.1016/j.neunet.2016.04.005
68. Vapnik VN (1999) An overview of statistical learning theory. IEEE Trans Neural Netw 10(5):988–999
69. Veena STAS (2018) Quantitative steganalysis of spatial LSB based stego images using reduced instances and features. Pattern Recogn Lett 105:39–49
70. Verma G, Verma H (2019) Predicting breast cancer using linear kernel support vector machine. In: Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE). https://doi.org/10.2139/ssrn.3350254
71. Villa A et al (2008) Gradient Optimization for multiple kernel's parameters in support vector machines classification 4:224–227. https://doi.org/10.1109/IGARSS.2008.4779698
72. Westfeld A (2001) F5—a steganographic algorithm. In International workshop on information hiding. Springer, Berlin, Heidelberg, pp 289–302
73. Wu NI, Hwang MS (2017) A novel LSB data hiding scheme with the lowest distortion. Imaging Sci J 65(6): 371–3788. https://doi.org/10.1080/13682199.2017.1355089
74. Yang C et al (2014) Evaluating unsupervised and supervised image classification methods for mapping cotton root rot. Precision Agriculture: 16. https://doi.org/10.1007/s11119
75. Yao XL, LG T, Dai GC (2008) Landslide susceptibility mapping based on Support Vector Machine: A case study on natural slopes of Hong Kong, China. Geomorphology 101:572–582. https://doi.org/10.1016/j.geomorph.2008.02.011
76. Yedroudj M (2019) Steganalysis and steganography by deep learning (Doctoral dissertation, Montpellier University)
77. Yi-Fei T, Tan W-N, Guo X (2013) Integrated lossy and lossless compression with lsb insertion technique in steganography. Proc SPIE: 8878. https://doi.org/10.1117/12.2031061
78. Yu L et al (2010) Improved adaptive LSB steganography based on chaos and genetic algorithm. EURASIP J Adv Signal Process: 876946. https://doi.org/10.1155/2010/876946
79. Zhang J, Cox IJ, Doerr G (2007) Steganalysis for LSB matching in images with high-frequency noise. IEEE 9th Workshop on Multimedia Signal Processing
80. Zhang Y-D et al (2016) facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation. IEEE