Improve steganography system using agents software based on statistical and classification technique

Estabraq Hussein Jasim Halboos, Abbas M. Albakry
Department of Computer Science, Informatics Institute for Postgraduate Studies, Iraq Commission for Computer and Informatics, Baghdad, Iraq

Article Info

Article history:

Received Aug 8, 2022
Revised Oct 6, 2022
Accepted Oct 24, 2022

Keywords:

Agent software
Imperceptibility
Least significant bit
Steganography
SVM classifier

ABSTRACT

In digital communications, information security is a paramount necessity. In the hiding algorithm, there are three basic parameters: security, capacity, and imperceptibility. Therefore, there are many ways to design the steganography algorithm, such as least significant bit (LSB), discrete wave transformation (DWT), and discrete cosine transform (DCT). The aim of this paper is to improve agent software design based on a steganography system. It proposed an agent system based on a support vector machine (SVM) classifier to hide a secret message in a certain cover image. The common dataset for steganography uses 80% training and 20% testing to get accurate results. Developing an agent system depends on six statistical parameters such as energy, standard deviation, histogram, variance, mean, and entropy. This resulted in features classified by the SVM classifier to predict the best cover image to be nominated for embedding. Worthy results were obtained in terms of imperceptibility, attack, and cover image prediction by statistical issues.

Corresponding Author:

Estabraq Hussein Jasim Halboos
Department of Computer Science, Informatics Institute for Postgraduate Studies
Iraq Commission for Computer and Informatics
Street AL-Nidal, 00964, Baghdad, Iraq
Email: ms202030596@iips.icci.edu.iq

1. INTRODUCTION

One of the important things in societies is information security. Previously, to hide messages at that time known encryption methods were used. The most sensitive data needs protection, therefore people uses several methods with different contexts, and two methods stand out here: encryption and steganography [1]. In the field of the internet and digital security, many people and hackers try to deceive technology for the purpose of obtaining data or personal files. On the other hand, there are those who are trying hard to develop techniques to maintain information security [2]. This is more than competition, research and development, so we see that many institutions, universities, and bodies are interested in hiding their security of information and its robustness as well, hence the need for information hiding and encryption techniques to protect that information and to solve problems related to safety and accuracy.

In this regard, the purpose of this research is to find and develop an approach to digital security for systems that need to protect copyright as well as valuable information. The work on steganography here uses the techniques of varying intensity to embed in the least significant bit (LSB) position of the pixels in certain image. The internet is the backbone of all services, so the use has begun to increase with the passage of time, and the devices associated with it have made the systems supporting it very complex, large and indefinite in size. These systems accommodate an unlimited number of users as they are heterogeneous systems and include continuous developments to keep pace with technology, so the independence of these programs is required in order to adapt to technological changes [3].

The special features of software agents are as follows: independence, meeting modern necessities, adaptability, and covering all computing processes. Computing in all organizations is connected to networks, and these complex systems are modeled to suit the work required. Through research, software agents have been developed into active areas and so on, and thus received a lot of attention, especially in the technology field [4]. Testing software agents is a difficult task due to the independence and distribution of these systems. Since the work here is open work and a certain context must be provided for that, and understanding the context is not easy, the operability that is accomplished is not normal and needs coordination with peers. the time. So the difficulty here lies in testing the system, which is predetermined on the basis of infinite work. The test is done by sending messages and tracing their trail, not through calling. Because the work of software is parallel, the tracking is not easy [5]. The agents correct the work automatically with the help of nearby areas and their work as a unit. One integrated supports the other for this reason often the tests are different for the same source because of the change in the smart technology of the method. Software agents can be defined as computational programs that have a set of features such as a sense of the interactive environment and changing variables, proactiveness in making decisions, and the independence of agents and beneficiaries, such as choosing certain actions within certain situations to reach the desired goals [6]. Knowledge sharing is also one of the goals to be achieved when working with software agents and competition is one of the options available here.

One of the main factors for hiding information is security, which has become one of the basics nowadays, especially after the massive perception in internet and communication technology. There are three systems that are the basis for information security, which are encryption, watermark, and steganography [7]. The work environment in this research is digital images, so choosing steganography is feasible and beneficial for the future of data security. Maintaining information security sometimes is not enough, so hiding that information may be more feasible at this stage [8]. There are several types of files that can be hidden in digital images such as text, video, audio and network protocol. Here, secret messages are absolutely necessary and it is very necessary to find new ways to hide them inside the picture. The cover image can contain a certain amount of invisible data. Because of the weaknesses in encryption algorithms, the science of hiding data has gained a lot of interest, with encryption concerned with the security of information [9]. Invisible messages can be exposed to attacks to a lesser extent than other methods of data hiding such as encryption and watermark, the process of hiding information depends on adding information to a master key. It is very important to choose the appropriate cover for the image in which we will hide the data, which needs an artificial intelligence process, and this is what the proposed research aims at

2. STEGANOGRAPHY

Steganography is a branch of information security that contains several ways to hide information digitally, as it is the main structure in creating messages. Figure 1 shows the data hiding classification quoted from [10]. steganography is an important category in the field of message hiding. The other classification is that steganography is strong or fragile. The watermark cannot remove information from it unless part of the image information is damaged, and this information is often. It is subject to change because it is among the least important area in the image, or the so-called LSB [11]. Among the security information that has a hidden character, including fingerprints, and Iris of our most interest in the technique of hiding information, which is the focus of our research is to include a message in the cover of the image, which is considered the main carrier in information transmission technologies. In this regard have to declare classification terms such as:

- Text semarams refer to information hiding via different methods for data presentation.
- Jargon code used when understanding the language by two parities agreement of using different ways.
- Covert cipher: representing when different payload capacity hiding into collection area referred by user.
- Covert grille: used when template cover entire image to enable some features under certain condition.

The steganography consists of several basic components, which are considered essential components, starting with the original image before embedding it, which is called the cover image, and the text to be included within the image is called the secret message, which is supposed to be transferred to the binary system [12]. The embedding process, which reflects the proposed method, is stored in the key called the stego key. At the other end of the network, the recipient receives stego image, i.e. the image loaded with the secret, and extracts the secret from it using the opposite of the method in the stego key as show in Figure 2.

The embedding process is done using LSB technology on the sender side, where a sequential set of bits is embedding to a randomly arranged set of image pixels. Each pixel gets two secret bits, this process is done on the sender side in order to produce at the end of the process an embedded image of the secret data called a stego image, and after the embedding process we make an evaluation of the image to make sure that the information is hidden correctly and accurately before sending. The stereo image is sent through the communication channels to the other party, which is the recipient, so that the data included in the image is extracted by a process called extraction, which is the process of reversing the embedding process.

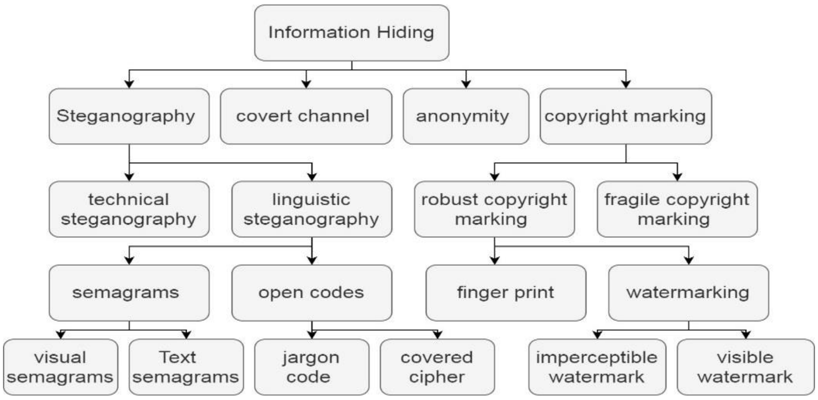


Figure 1. Classification of information hiding

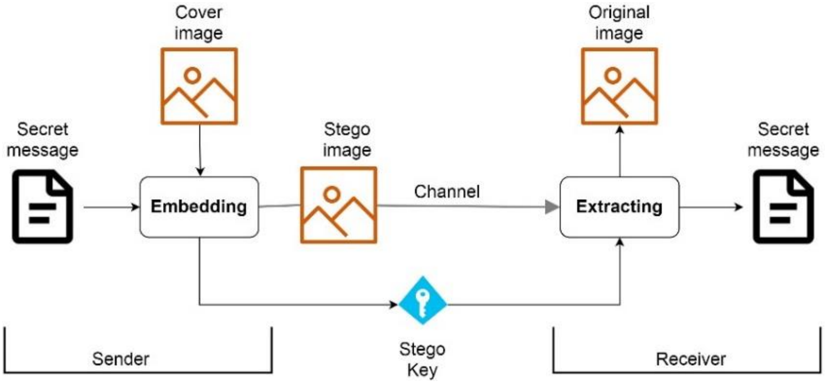


Figure 2. Steganography system as proposed method

3. AGENT SYSTEM

The software agent is a very important field of research, especially in artificial intelligence systems, and there is no general definition for it. However, the common definition for it is: a computer system that works independently and according to the requirements of the work environment environment to meet a specific goal, which is what it was designed for [13]. Through the word agent, which is within the general term, a software agent is a special, independent entity that has the authority to act within the environment to be operated on. Particularly here, the agent can be defined as a system that has the ability to predict and perceive as the human mind to act on new inputs and take appropriate action.

Among the main properties of agent software is independence, meaning that decisions that are made during the workflow are independent and are not affected by the initial main inputs, and decisions are taken with complete independence [14]. One of the properties is also the lack of presence that distinguishes the agent in that it operates as a hidden system as it is not completely linked to the secondary system within one working environment. Flexibility is also an important characteristic that indicates the extent of the immediate response to the work, in a timely and proactive manner, which gives the ability to give and predict decisions before the failure of the system [15]. Many applications of artificial intelligent (AI) used to solve problems that need prediction such as pattern recognition [16], [17] that behave like agent and looking for the best output by estimating technique.

The software agent can be classified after installing the important definition in several ways, including to sub-groups and according to the properties they have, and each sub-branch is defined according to its properties in the program. Also, they are classified on the basis of a hierarchical definition and according to the tasks performed and their importance. There is another classification according to the outputs of the process, which may be primary or secondary, or they can be classified according to the control environment and may be external by the user or internal according to the possible result.

4. RELATED WORK

Many of the previous studies used steganography, and the most important of these studies was through [18] where a hybrid algorithm was used to increase the compression (Huffman) of the confidential data to be included in the image, in addition to encryption to increase data security. Edges were used in the embedding of the image via most-significant bit (MSB) and two axes of LSB and using edges to increase the payload capacity of the image for the data, where the number of bits used in one pixel reached three bits, where the result of the embedding was efficient in terms of capacity, but few in terms of security [19]. A new method was proposed by [20] of using training for the method of hiding deep information, which proved its worth in including a large amount of data in the cover of the image, and through the outputs represented by peak signal to noise ratio (PSNR) and structural similarity index measure (SSIM), the method proved the efficiency of the method, which reflects the strength and applicability of the method used.

An effective method presented by [21] to hide the data in the image using BIM, choosing pixels at random, the night tour method, and compressing the data before embedding to increase the storage capacity of the cover image, as the result was encouraging and scalable for future development. Lu *et al.* [22] use a large-capacity and inverted steganography network in the image hiding process, which carries high confidentiality when reversing the embedding (extraction) process, and using the forward and backward propagation method to get good results. A new strategy was proposed by [23] to dividing the payload capacity to be embedded to the image through three channels in the red, green, and blue (RGB) color image and calculating the three channels' priority probabilities in order to increase the security of embedding and prevent repetitions that reduce imperceptibility, the result in this method was satisfactory [24], [25] presented a method to hide data in a compressed image cover of the JPEG type, which uses the intermediate image format and processed through high-frequency pixels and the adoption of histogram, but this method has a weak result due to the nature of the compressed image, which does not allow embedding more than the permissible limit.

A new method was suggested by [26] to embed confidential data in the cover image based on pixel locations, wavelength, and pixel selection used for Harris Hawks optimization (HHO) algorithm in embedding and extraction operations, the results were better than the rest of the methods and are characterized by high security and robustness. Baothman and Edhah [27] proposed a software agent method included with steganography to hide the data by LSB method in an image dataset in the process of training and testing, and the statistical characteristics of the agent's process were adopted in the working environment of MATLAB, and six statistical measures were adopted to achieve good results. A multifactor agent approach proposed by [28] based on steganography when it is transferred in cover image between sender and recipient, where the agent encrypts the text before embedding in the system on vehicular ad-hoc network (VANET). Fedorov *et al.* [29] develop a reliable method for building a software agent system to hide data in steganography by using classification and confusion matrix to find measurement of risks in a specific work environment and predict the output time of the proposed system.

A new approach to steganography was introduced by [30] without losing data via a software agent, which managed to hide a large amount of data in the cover of the selected images without increasing the image size or any distortion in it. To solve the security problems in the steganography system and data integrity, Araoye *et al.* [31] developed a styakanocrifi system based on digital encryption called elliptic curve cryptography (ECC) and used a value of 5 for PHP. Comparing the result with the commercial systems in the local market and found that the response time is very good. He addressed the weaknesses in the agent system, which is the basis of security, and adopted encryption in addition to steakanocrifi to evaluate the performance of the software agent system, and adopted the time of loading and execution of the agent in order to evaluate the performance of the system. He addressed the weaknesses in the agent system, which is the basis of security, and adopted encryption in addition to steakanocrifi to evaluate the performance of the software agent system, and adopted the time of loading and execution of the agent in order to evaluate the performance of the system. Araoye *et al.* [32] addressed the weaknesses in the agent system, which is the basis of security, and adopted encryption in addition to steganography to evaluate the performance of the software agent system, and adopted the time of loading and execution of the agent in order to evaluate the performance of the system.

5. METHOD

The algorithm was implemented in the form of stages and in a working environment that is MATLAB and using the LSB method. The method of work is divided into two main stages, the first is steganography, which is to hide the data in the part of the LSB of the selected pixels in certain image randomly, and the second stage includes choosing the appropriate cover image using the agent technique. LSB technology is considered one of the important methods in the image approach and without losing important data, two bits are used to embed in each color contrast. In color image there are three contrast of RGB in one pixel, so the addition is six bits per pixel because one pixel consists of 24 bits, 8 bits for each color, which allows embedding more storage capacity.

Choosing pixels at random increases the security of the data in the image, and the selection process is unique so that it is not repeated, and this work with the randomness in (1) is stored in the stego key. The process of random selection of pixels starts from a number called the seed number and can be found in (2) [30]. Initial value start from seed number and continue through random number till scan all the image then order of pixels will start in random vector.

$$X_{n+1} = (a \times X_n + b)\ (mod\ m), (X_n)\ n \geq 1 \qquad (1)$$

$$Seed = \sum_{i=1}^{10}(Max\ Pixels\ mod\ i) + i \qquad (2)$$

a, *b* and *m* are the parameters chosen expermentaly that stay as constant at all the process. Xn is the initial number. Long serial of random number will start from seed until the last pixel in the image.

First the image enters the system and performs a preprocessing in order to choose the appropriate pixels for the embedding process, or rather the new sequence of pixels in the image. Then we embed in a way that ensures non-perceptibility, which is as follows: if the bit taken from the secret message is 1 or 0, it will be embed only if the condition that takes into consideration the amount of color change in adjacent pixels is met. The peak limit of the color change (threshold) is 50 or more between the center and the edges neighbour of the pixel. In the case of less than threshold, we use the other pixel in the embeding sequence as in Figure 3

The purpose of choosing threshold 50 value due to embedding two pixels give impact into pixel value such as 20-22 give 7 decimal value then have to find wide range to hide it. LSB pixel not always suitable for embed so selecting strategy play main rule in this issue. Condition help system to hide data with high security and imperceptibility. Strategy of embedding in LSB occurs in the first and third bits in grey pixel image but in color image the pixel consist of 24 bits in three channels for each 8-bits that's mean there are three LSBs then the embedding will be 6 secret bits instead of 2 bits in grey image as is seen in Figure 4.

Decisions are made about the software agent and its responsibility, which is relatively independent, and a decision is made about the use of cover image in the event that the required conditions are met, and this image is a candidate for embedding among a group of images in the dataset. The software agent plays an important role in choosing the image in which the data will be hidden. There are images whose proparities are weak in hiding the data and others with high efficiency and according to the contents of the image. High color contrast, hold secret bits better and difficult to detect because a small change in the uniform color is easy to detect and is visible to the human eye however statistical equations are accurate than human eyes.

According to Li and Hilliges [33], the agent depends on a set of conditional procedures and important rules within a particular environment. The agent predetermines the appropriate environment (image) for hiding data and follows interactive factors in that. When the user needs some process or procedures such as choosing an image, hiding secret data, inserting a file or other things, the proposed system sends a request to the agent to carry out this task and responds immediately whether the task is executed or not as shown in Figure 5.

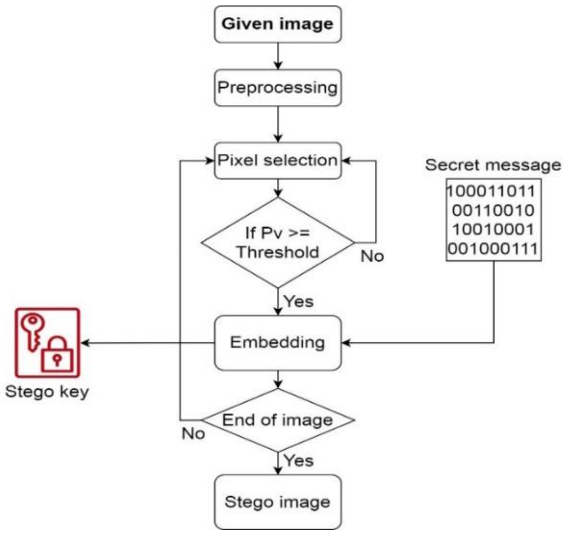


Figure 3. Embedding strategy in proposed steganography system

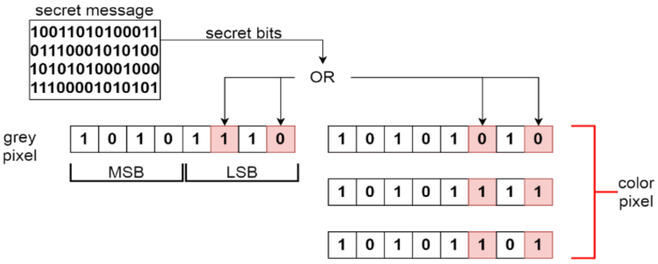


Figure 4. LSB embedding in both color and grayscale pixels

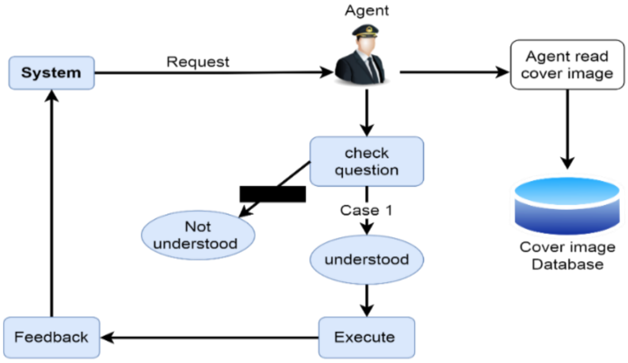


Figure 5. Agent steganography system

The proposed agent system embeds the secret message in the cover image and extracts it from the second party, and this image that carries the secret must be chosen or recommended by the system depends on the properties and statistical features that are extracted from the image. Then the agent analyzes many of the image's properties, and then, based on the results of the tests of those properties, the agent chooses the best cover image to hide the data from. As for the statistical properities [30] of the image, it can be summarized as follows:

– Histogram: it represents a diagram showing the numerical distribution of data in the form of numerical values to compare the grayscale against the number of pixels in the image.

$$P(g) = \frac{N(g)}{M} \tag{3}$$

N(g) represent pixels of grey level *g* and *M* is image pixels. The histogram reflects the ability of the image to absorb secret data in a way that is not noticed by the hacker, so that the image with a close color gradation is not good for holding secret data and vice versa.

– Average: measures the average value of the brightness of the pixels in the image, this related with neighbour pixels when scan the image. Can illustrate in (4).

$$g = \sum_r \sum_c \frac{I(r,c)}{M} \tag{4}$$

I(r,c) represent pixels’ value of certain image, brightness of the image reflects the varying of pixel value at the subgroup to determine which area in the image be suitable for embedding that agent to decide. Accumulation averaging of the subgroups give good illustration of image compatibility.

– Standard deviation: that consider the square root of vaiance. It defines the image contrast as (5):

$$\delta_g = \sqrt{\sum_{L-1}^{g=0} (g_L - g_{L+1}) P(g)} \tag{5}$$

Standard deviation represent variance of contrast as which pixel appropriate for embedding in such area.

– Entropy: used to measure the bits required to embed the amount of data in certain image. It is increases if the distribution of pixel values is large within the gray area and can fellow in (6):

$$Entropy = \sum_{L-1}^{g=0} P(g) log_2[P(g)] \tag{6}$$

High entropy allows the agent to select the best area for embedding and return specific area to the system accordingly.

- Variance: it measures the level of roughness in specific areas of the image, being the areas that are more likely to be embed than others as in (7):

$$\delta_g^2 = \sum_{g=0}^{L-1}(g - g)^2 P(g) \tag{7}$$

- Energy: it measures the intencity variation in a specific area of the image and also the rate of the sum of the intencity in the areas of the image represented in (8):

$$E = \sum_{g=0}^{L-1}(p(g))^2 \tag{8}$$

The agent in the proposed steganography seeks to find the cover image among a set of images in the database with different features of contrast, entropy, and histogram. The agent chooses the cover of the image in order to embed secret data to the image by the proposed steganography method, and as in the Figure 6. The agent will do some stages before embedding the text randomly and choose the image based on the six mentioned statistical features such as histogram, mean, entropy, variance, energy, and standard deviation. Which are determined by the ability of the image to be embedded and compared with what is in the database in order to obtain stego image suitable for sending.

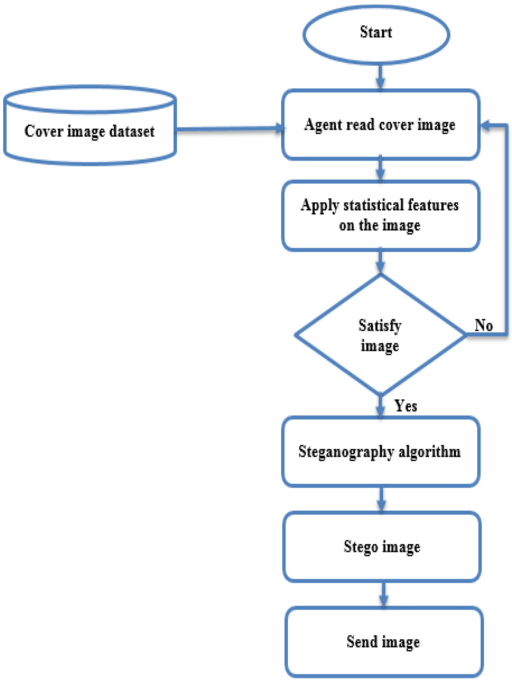


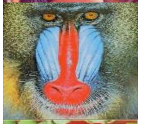
Figure 6. Agent system follow by proposed method

6. RESULTS AND DISCUSSION

The image that has been recommended has a size of 512×512 pixels and an 8-bit resolution for each individual pixel. The random function needs to be used to choose which pixels will be utilized for embedding. Utilizing the random function, which contributes to an increased level of safety, is the first step in the security process. In spite of the fact that the pixels are selected at random, the outcome of the random function is not used in its raw form. Instead, it is subjected to the method that has been proposed in order to be able to identify which pixel would be the location of the hidden data. Each cluster of pixels inside a picture has the same color value, and this is the point at which the methodology comes into play, as seen in Figure 4.

The agent helps maintain the secret of information through steganography by choosing the best cover image from a database in which the data can be hidden in a good way. Choosing cover image is the first stage of data hidding and is considered the cornerstone that affects the safety of the stego image. The agent analyzes the features of the images and chooses the best image according to what he concludes from the statistical analysis of the images. The LSB method is used with the same two bits to embed data in the selected image, and according to the image features of entropy, contrast, mean and other features, the agent will search for the highest and most appropriate criteria to be done and choose the appropriate cover for the embedding. In Table 1 shows the features extracted from images taken from a database downloaded from the SIPI website [34] and for images with a resolution of 256×256 pixels.

Table 1. Features of some cover images from standard dataset

Image	Mean	Entropy	Variance	Contrast	Energy	Standard deviation
	172.234	9.6540	3013.53	6.735	1.35462×10-1	4.25142×10-1
	174.832	10.6660	3863.51	7.532	1.94462×10-1	4.82241×10-1
	190.382	12.2378	3999.83	5.478	1.89933×10-1	4.76290×10-1

The MATLAB environment was used to extract features from the image, as well as in the steganography algorithm, and the software agent architecture was developed to match what was required in the program to get the best result. On this basis, the vector image was chosen to be the best suitable cover for hiding the data, and the agent was trained using a support vector machine (SVM) classifier in order to obtain an accurate output, as some of the statistical properties are on it and some are few for a particular image, so the classifier's task is to choose the best among them. Cost function [35] used to find the good prediction in term of multidimensional features as in (9).

$$j(\theta) = \frac{1}{m}\sum_{i=1}^{m} Cost\left(h_\theta\left(x^{(i)}\right), y^{(i)}\right) \tag{9}$$

$$Cost(h_\theta(x), y) = \begin{cases} \{-\,log\ log\ \left(h_\theta(x)\right)\ when\ y\ is\ 1 \\ -\,log\ log\ \left(1 - h_\theta(x)\right)\ when\ y\ is\ 0 \end{cases} \tag{10}$$

Where j is logistic regression, m is number of features, h is feature function, x and y are features domain in x and y vectors. Dataset labeled to train the classifier in 80% while 20% used for testing mode. To fit parameter of (θ) use (11):

$$\min_{\theta} J(\theta) \tag{11}$$

And to make prediction we give the new features factor as (12):

$$Output\ h_\theta(x) = \frac{1}{1+e^{-\theta^T x}} \tag{12}$$

T represent the threshold that determines feature item belong to which class.

6.1. Steganography implementation

During the processing stage, the quality of the image is degraded, and there is a possibility that part of the image's information would be lost. The standard techniques of assessment in such a scenario are either subjective or objective. Finding the difference between the starting image and the final image and then using statistical analysis to that difference is essential to maintaining objectivity. The subjective technique is characterized by the absence of any reference to standards or criteria in the observation and evaluation processes. In this section, we will explore the importance of a wide variety of criteria that are used to assess such systems.

6.2. Imperceptibility

It is an essential prerequisite for the method of concealing data in photographs and it is essential that the image maintains its quality once the embedding procedure has been completed. It is not feasible to view the concealed data in most cases, and the image gives the impression of being innocent since it does not include any confidential data. PSNR [36] is a measurement used to determine how perceptible a system is. This ratio may be calculated as in (13):

$$PSNR = 10.\,log_{10}\left(\frac{MAX_1^2}{MSE}\right) \tag{13}$$

In addition, to determine the mean square error (MSE), the results in (14) [37]:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \tag{14}$$

where MAX is the highest possible pixel value in the image, n and m denote the dimensions of the image, and I and K denote the original and stego pixel values respectively (from cover and stego image). In the LSB component of the image, embedding 16,384 bytes equals about 6.25% of the total amount of information that the image is capable of holding. Because the tables that follow will show that the increase in capacity is inversely proportional to the value of the PSNR, and because a fair evaluation requires taking into consideration the percentage of the embedding in order to standardize the embedding, it is imperative that this percentage be taken into consideration. After embedding 6.25% of the payload capacity, Figure 6 displays the cover image as well as the stego image.

For more accurate benchmarking, we employed the recommended technique, which consisted of using three different quantities of payload capacity and three different types of pictures taken from the standard dataset of the USC SIPI. The suggested system makes use of three different text sizes totaling 16,384; 32,768; and 49,152 bytes. These text sizes correspond, respectively, to 6.25%, 12.5%, and 18.75% of the system. This procedure used an image size of 512 by 512 pixels for a color image. The procedure used for embedding a grayscale image was the same as that used for a color image; the only difference was that grayscale images only used one channel for embedding, while color images used all three channels of the RGB color space. The percentage of payload capacity is calculated as follows: if one pixel is equal to eight bits, then one eighth is equal to twelve and a half percent during the embed one bit process. If two pixels are equal to sixteen bits, then one sixteenth is equal to six and a half percent during the embed one bit process, as shown in Table 2

Because each contrast value in this form of image originates from three independent pixels, one for red, one for green, and one for blue, the embedding process will take place in three channels. This is owing to the fact that the combination of these pixels produces a single-color pixel. Steganography presents a different set of challenges when dealing with color images, such as the fact that each color image contains three different types of mixed color-RGB, which stands for red, green, and blue. However, as seen in Table 3 embedding in color images has a little greater imperceptibility than doing so in grayscale images.

Since we can see from Table 2 that the PSNR will fall when the payload capacity is increased because more data will be more detectable, it is important to strike a balance between increasing the payload capacity and maintaining a high level of image quality. It is also possible to see that images with a more variable contrast will produce more sub-images, which, in accordance with the suggested strategy, means that images with a more variable contrast in their pixels may store more hidden bits. Because pepper's image has a greater variety of colors and a greater number of distinct areas, it is more suited to conceal data inside these regions.

Table 2. PSNR of common photos with various pixel counts (grey-images)

Image/pixels	Percentage (%)	Peppers image (dB)	Baboon image (dB)	Lena image(dB)	Payload capacity
512×512	6.25	89.5	86.8	82.7	16,384
512×512	12.5	85.0	82.9	80.2	32,768
512×512	18.75	80.3	79,8	76.9	65,536
1,024×1024	6.25	55.7	54.4	51.3	16,384
1,024×1024	12.5	47.9	47.4	46.2	32,768
1,024×1024	18.75	45.6	44.5	43.1	65,536

Table 3. Results of PSNR of color images

Payload capacity	Percentage (%)	Lena image	Baboon image	Peppers image
16,384	6.25	84.7	88.5	90.9
32,768	12.5	82.4	83.0	86.1
65,536	18.75	77.2	80.3	82.2

6.3. Chi-square attack (X2)

The chi-square statistic, which stands for the second kind of assessment, is one of the aims that have been taken into account in the approach that has been suggested. One of the goals that has been taken into account is to identify the existence of the attack known as chi-square, which is considered to be one of the most harmful assaults that can be used to determine whether or not data is present in an image. When opposed to a cover photograph, an image that really contains data is far more prone to being targeted by an attacker. The attacker in this case only has access to the stego image; as a result, he utilizes the arithmetic mean to compute the frequencies, and for each LSB bit, he calculates their frequency in order to determine whether or not the image contains any hidden data. Since the attacker only has access to the stego image, he cannot compare it to the original image.

The chi-square test is seen here in Figure 7. The x-axis indicates the proportion of each image, while the y-axis displays the chance that the hidden message is included in the image. This sort of attack is known as the statistical attack, and it exhibits the potential of incorporating data in the picture by repeatedly including the LSB bits in the stego image. This type of attack may be distinguished from other types of attacks by its name. There is a possibility of 0.065 percent that the first 3 percent will be right if the units meet the objective of each pixel. This test is used to calculate the frequency ratios in order to establish the embedding rate. This is possible since the vast majority of the letters in the English language begin with the same frequencies and the value of the repeated bits. If the frequency of the LSB stego pixels in the image 512×512 pixels is increased in order to embed 18.75% of them, then the embedding ratio will clearly increase logarithmically. This is because the frequency of the rest of the pixels, which are not embedded, has an effect on the statistical issue of the chi-square equation, as shown in Figure 8.

In a nutshell, the chi-square test may determine the security of pictures by analyzing the manner of embedding them. It also depends on the statistical distribution of the values of the pixels in the image, and it is possible to obtain better results for the first part of the image despite the fact that the beginning of the iterative calculation is inaccurate in this part of the image. This is because the iterative calculation starts off inaccurately in this part of the image. As a consequence of this, the location where the frequencies are computed in line with the equation is an acceptable location for adding the information, which can be found as (15) [38]:

$$X^2 = \sum \frac{(Observed - Expected)^2}{Expected} \tag{15}$$

Proposed method consists of two main stages select the best cover image by the agent software then hiding secret message into this image by steganography algorithm. In this regard two stages of evaluation performed for eagent and steganography each with its details.

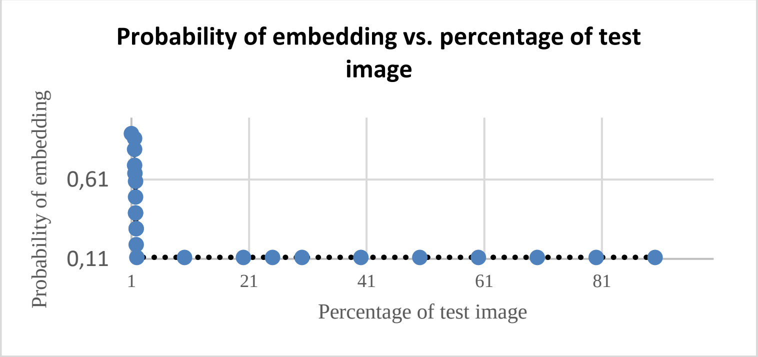


Figure 7. Chi-square analysis performed on the pepper image as part of the suggested technique

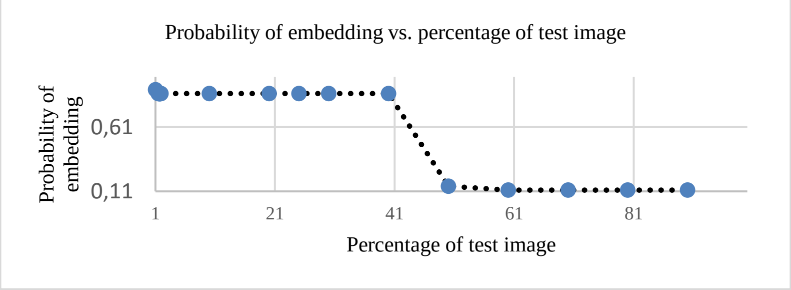


Figure 8. The behavior of the chi-square test after embedding 18.75% of the payload capacity

7. CONCLUSION AND FUTURE NETWORK

In the past few years, interest has increased to hide data in various media, including images, and there are three important factors in steganography: payload, security, and imperceptibility. Payload capacity defines it as the amount of information that is embedded to the cover image and the extent of the capacity of that image, and security is that the data is not disclosed when analyzing it except with the presence of the stego key only, and the imperceptability is that the image when sent is innocent and does not contain signs indicating the presence of any information. A software agent was used with six features including histogram, entropy, mean, standard deviation, variance, and power. The features were classified using the well-known SVM classifier to select the best cover images to be filtered for steatography, to maintain the security of the information sent to the second part.

Many studies have focused on the agent of software and linking it to steganography, as well as the use of statistical properties. It is possible in the future studies to shorten these statistics and add statistics that will reveal more about the possibility of embedding data to the image as the relationship of pixels to the far neighborhood that is more than 8 neighbor pixels. We recommend adopting the remaining classifiers or using the deep learning method as a first step for effective prediction.

REFERENCES

[1] M. H. Muhammad, H. S. Hussain, R. Din, H. Samad, and S. Utama, "Review on feature-based method performance in text steganography," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 1, pp. 427–433, Feb. 2021, doi: 10.11591/eei.v10i1.2508.

[2] K. Manjunath, G. N. K. Ramaiah, and M. N. G. Prasad, "Backward movement oriented shark smell optimization-based audio steganography using encryption and compression strategies," *Digital Signal Processing*, vol. 122, p. 103335, Apr. 2022, doi: 10.1016/j.dsp.2021.103335.

[3] E. H. J. Halboos and A. M. Albakry, "Hiding text using the least significant bit technique to improve cover image in the steganography system," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 6, pp. 3258–3271, Dec. 2022, doi: 10.11591/eei.v11i6.4337.

[4] M. O. Igbinovia, "Internet of things in libraries and focus on its adoption in developing countries," *Library Hi Tech News*, vol. 38, no. 4, pp. 13–17, Nov. 2021, doi: 10.1108/LHTN-05-2021-0020.

[5] J. D. Priest, A. Kishore, L. Machi, C. J. Kuhlman, D. Machi, and S. S. Ravi, "CSonNet: an agent-based modeling software system for discrete time simulation," in *2021 Winter Simulation Conference (WSC)*, Dec. 2021, pp. 1–12, doi: 10.1109/WSC52266.2021.9715287.

[6] W. Fan, P. Chen, D. Shi, X. Guo, and L. Kou, "Multi-agent modeling and simulation in the AI age," *Tsinghua Science and Technology*, vol. 26, no. 5, pp. 608–624, Oct. 2021, doi: 10.26599/TST.2021.9010005.

[7] N. J. Alhyani, O. K. Hamid, S. Y. Ali, and A. M. Ibrahim, "Efficient terrestrial digital video broadcasting receivers based OFDM techniques," *Przegląd Elektrotechniczny*, vol. 1, no. 11, pp. 74–77, Nov. 2021, doi: 10.15199/48.2021.11.13.

[8] H. Zhou, W. Zhang, K. Chen, W. Li, and N. Yu, "Three-dimensional mesh steganography and steganalysis: a review," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 12, pp. 5006–5025, Dec. 2022, doi: 10.1109/TVCG.2021.3075136.

[9] V. R. F. Signing, T. F. Fonzin, M. Kountchou, J. Kengne, and Z. T. Njitacke, "Chaotic jerk system with hump structure for text and image encryption using DNA coding," *Circuits, Systems, and Signal Processing*, vol. 40, no. 9, pp. 4370–4406, Sep. 2021, doi: 10.1007/s00034-021-01665-1.

[10] X. Wang, C.-C. Chang, and C.-C. Lin, "High capacity reversible data hiding in encrypted images based on prediction error and block classification," *Multimedia Tools and Applications*, vol. 80, no. 19, pp. 29915–29937, Aug. 2021, doi: 10.1007/s11042-021-11143-0.

[11] W. Alexan, A. Elkhateeb, E. Mamdouh, F. A.-Seba'Ey, Z. Amr, and H. Khalil, "Utilization of corner filters, AES and LSB steganography for secure message transmission," in *2021 International Conference on Microelectronics (ICM)*, Dec. 2021, pp. 29–33, doi: 10.1109/ICM52667.2021.9664947.

[12] M. Ulker and B. Arslan, "A novel secure model: image steganography with logistic map and secret key," in *2018 6th International Symposium on Digital Forensic and Security (ISDFS)*, Mar. 2018, pp. 1–5, doi: 10.1109/ISDFS.2018.8355320.

[13] C. Burr, N. Cristianini, and J. Ladyman, "An analysis of the interaction between intelligent software agents and human users," *Minds and Machines*, vol. 28, no. 4, pp. 735–774, Dec. 2018, doi: 10.1007/s11023-018-9479-0.

[14] S. Sahu, R. Agarwal, and R. K. Tyagi, "A novel αβevolving agent architecture for designing and development of agent-based software," in *Transforming Management with AI, Big-Data, and IoT*, Cham: Springer, 2022, pp. 169–184, doi: 10.1007/978-3-030-86749-2_10.

[15] K. Pal, "Software agent-based simulation for pan-european transport corridor management in supply chain," in *Handbook of Research on Decision Sciences and Applications in the Transportation Sector*, Pennsylvania, USA: IGI Global, 2021, pp. 325–339, doi: 10.4018/978-1-7998-8040-0.ch015.

[16] G. Sulong and A. Mohammedali, "Recognition of human activities from still image using novel classifier," *Journal of Theoretical and Applied Information Technology*, vol. 71, no. 1, pp. 115–121, 2015.

[17] G. Sulong and A. Mohammedali, "Human activities recognition via features extraction from skeleton," *Journal of Theoretical and Applied Information Technology*, vol. 68, no. 3, pp. 645–650, 2014.

[18] O. F. A. Wahab, A. A. M. Khalaf, A. I. Hussein, and H. F. A. Hamed, "Hiding data using efficient combination of RSA cryptography, and compression steganography techniques," *IEEE Access*, vol. 9, pp. 31805–31815, 2021, doi: 10.1109/ACCESS.2021.3060317.

[19] D. R. I. M. Setiadi, "Improved payload capacity in LSB image steganography uses dilated hybrid edge detection," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 2, pp. 104–114, Feb. 2022, doi: 10.1016/j.jksuci.2019.12.007.

[20] L. Liu, L. Meng, X. Wang, and Y. Peng, "An image steganography scheme based on ResNet," *Multimedia Tools and Applications*, vol. 81, no. 27, pp. 39803–39820, Nov. 2022, doi: 10.1007/s11042-022-13206-2.

[21] A. M. Fadhil, "Bit inverting map method for improved steganography scheme," M.S. thesis, Dept. Comput. Sci, Universiti

Teknologi Malaysia, Johor Bahru, Malaysia, 2016.

[22] S. P. Lu, R. Wang, T. Zhong, and P. L. Rosin, "Large-capacity image steganography based on invertible neural networks," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 10811–10820, 2021, doi: 10.1109/CVPR46437.2021.01067.

[23] Y.-H. Chuang, B.-S. Lin, Y.-X. Chen, and H.-J. Shiu, "Steganography in RGB images using adjacent mean," *IEEE Access*, vol. 9, pp. 164256–164274, 2021, doi: 10.1109/ACCESS.2021.3132424.

[24] X. Hu, J. Ni, W. Zhang, and J. Huang, "Efficient JPEG batch steganography using intrinsic energy of image contents," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 4544–4558, 2021, doi: 10.1109/TIFS.2021.3109464.

[25] L. Zhu, X. Luo, C. Yang, Y. Zhang, and F. Liu, "Invariances of JPEG-quantized DCT coefficients and their application in robust image steganography," *Signal Processing*, vol. 183, pp. 1–15, Jun. 2021, doi: 10.1016/j.sigpro.2021.108015.

[26] R. Sonar and G. Swain, "Steganography based on quotient value differencing and pixel value correlation," *CAAI Transactions on Intelligence Technology*, vol. 6, no. 4, pp. 504–519, Dec. 2021, doi: 10.1049/cit2.12050.

[27] F. A. Baothman and B. S. Edhah, "Toward agent-based LSB image steganography system," *Journal of Intelligent Systems*, vol. 30, no. 1, pp. 903–919, Jul. 2021, doi: 10.1515/jisys-2021-0044.

[28] V. Gautam, "MASSS—multi-agent-based steganography security system for VANET," in *Proceedings of 3rd International Conference on Computing Informatics and Networks*, Singapore: Springer, 2021, pp. 159–172, doi: 10.1007/978-981-15-9712-1_14.

[29] V. K. Fedorov, E. G. Balenko, S. I. Shterenberg, and A. V Krasov, "Development of a method for building a trusted environment by using hidden software agent steganography," *Journal of Physics: Conference Series*, vol. 2096, no. 1, pp. 1–6, Nov. 2021, doi: 10.1088/1742-6596/2096/1/012047.

[30] D. D. L. Nascimento, F. R. P. Couto, L. Z. Wolski, and I. A. Kuhnen, "Image steganography using LSB and software agents," *International Journal of Engineering Research and*, vol. 6, no. 3, pp. 191–195, 2017, doi: 10.17577/ijertv6is030243.

[31] O. I. Araoye, O. S. Adewale, B. K. Alese, and R. O. Akinyede, "Developing a secured mobile-agent-based electronic commerce using crypto-steganography," *International Journal of Sciences*, vol. 4, no. 2, pp. 82–88, 2018, doi: 10.18483/ijsci.1550.

[32] O. I. Araoye, K. A. Akintoye, and M. K. Adu, "Performance evaluation models for a secure agent-based electronic commerce," *International Journal of New Technology and Research (IJNTR)*, vol. 4, no. 4, pp. 116–120, 2018.

[33] Y. Li and O. Hilliges, *Artificial intelligence for human computer interaction: a modern approach*. Cham: Springer, 2021, doi: 10.1007/978-3-030-82681-9.

[34] D.-H. Kim and H.-Y. Lee, "Deep learning-based steganalysis against spatial domain steganography," in *2017 European Conference on Electrical Engineering and Computer Science (EECS)*, Nov. 2017, pp. 1–4, doi: 10.1109/EECS.2017.9.

[35] S. Koda, A. Zeggada, F. Melgani, and R. Nishii, "Spatial and structured SVM for multilabel image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 10, pp. 1–13, 2018, doi: 10.1109/TGRS.2018.2828862.

[36] N. Singh, "High PSNR based image steganography," *International Journal of Advanced Engineering Research and Science*, vol. 6, no. 1, pp. 109–115, 2019, doi: 10.22161/ijaers.6.1.15.

[37] S. Farrag and W. Alexan, "Secure 2D image steganography using recamán's sequence," in *2019 International Conference on Advanced Communication Technologies and Networking (CommNet)*, Apr. 2019, pp. 1–6, doi: 10.1109/COMMNET.2019.8742368.

[38] W.-B. Lin, T.-H. Lai, and C.-L. Chou, "Chi-square-based steganalysis method against modified pixel-value differencing steganography," *Arabian Journal for Science and Engineering*, vol. 46, no. 9, pp. 8525–8533, Sep. 2021, doi: 10.1007/s13369-021-05554-2.

BIOGRAPHIES OF AUTHORS

Estabraq Hussein Jasim Halboos currently is a Master's researcher at Informatics Institute for Postgraduate Studies (IIPS), Iraqi Commission for Computers and Informatics (ICCI), Baghdad, Iraq. She received her Diploma in computer science from Informatics Institute for Postgraduate Studies at the Iraqi Commission for Computers and Informatics, Iraq, in 2019. She received a B.Sc. in computer science from the University of Technology, Baghdad, Iraq, in 2010. She researches interests include artificial intelligence, machine learning, bioinformatics, and cyber security. She can be contacted at email: ms202030596@iips.icci.edu.iq.

Prof. Abbas M. Al-Bakry he is currently a senior lecturer at University of Information Technology and Communications. Research interest: software agents, internet, web sites, artificial intelligence, bar codes, biomedical MRI, brain, data handling, data mining, face recognition, feature extraction, image classification, image denoising, image segmentation, information filtering, intelligent transportation systems, learning (artificial intelligence), matrix algebra, medical image processing, multi-agent systems, object recognition, query processing, radiofrequency identification, real-time systems, and road vehicles. He can be contacted at email: abbasm.albakry@uoitc.edu.iq.