

Exploring the Use of Steganography and Steganalysis in Forensic Investigations for Analysing Digital Evidence

KRISTIYAN MICHAYLOV, University of Twente, The Netherlands

Image steganography and steganalysis have gained significant popularity in recent years. The military, medical, e-governmental and social media fields are some places where image steganography and steganalysis are utilized. To answer the growing demands, digital forensic investigators (DFI) are interested in analysing the implications of image steganography and steganalysis domains. Nevertheless, these professionals have limited qualitative resources dedicated to a systematic analysis of techniques, tools and metrics used in these domains. This research concentrates on 3 parts. First, an extensive literature review of the existing papers for Artificial Intelligence (AI), statistical and signature steganalysis techniques is performed. The study suggests that AI-driven steganalysis techniques are not strictly better at detecting image steganography, compared to the rest. Second, some popular, non-paid steganography (F5, Steghide and Outguess) and steganalysis (Aletheia, StegExpose) tools are utilized on a JPEG dataset. The detection accuracies are compared to answer why despite having a lower accuracy, Aletheia is more appropriate for DFI than StegExpose. Finally, features such as size, colour, mean squared error (MSE) and peak signal-to-noise ratio (PSNR) between stego (with different sizes of embedded secret text) and cover JPEG images from the dataset are examined. It is found that none of the chosen features produces a direct indication regarding the possible existence of hidden messages inside JPEG images. Overall, the current research performs a novel qualitative approach, performing a literature review as well as experimentation. Based on that, the results and conclusions could help professionals to tackle and analyse image steganography and steganalysis more systematically, to obtain more insightful results.

Additional Key Words and Phrases: Image Steganography Tools, Image Steganalysis Tools, Forensic Investigators, Joint Photographic Experts Group (JPEG)

1 INTRODUCTION

Steganography has been utilized by people from ancient times [22]. The term has Greek origin and means covered writing, indicating that the main idea is to hide a message inside another media [22]. Steganography has evolved, in many ways, since the beginning of its practice. From the creation of an astragal during the time of Aeneas the Tactician [22], to the current state of steganography [8, 35] which has not changed the foundational principles. In fact, nowadays, steganography is more concentrated on concealing messages in digital resources. Currently, depending on the type of data that should be concealed, there exist different types of steganography, such as text, image, audio, network, and video [33]. In order to find hidden messages inside the above-mentioned steganography resources, it is important to follow a systematic approach. The process of discovering the messages hidden inside data payload is called steganalysis. There are several attacking techniques which

are used for steganalysis, such as a stego-only attack, known cover attack, known message attack and more [24] that can be used from digital forensics, in order to find hidden messages.

For the last 20 years of technological advancements, steganography has posed numerous challenges to cybersecurity specialists. Part of the problems include detecting malicious data, embedded inside digital images, web advertisement banners and videos [7, 11, 34]. Most people would likely not suppose that an image or other digital resource received online contains hidden information or malware script. They will not be aware that this could corrupt their computer and even format their hard disk [11]. For this reason, it is crucial that attention is focused on steganography and detecting the presence of it. DFI are some of the professionals who are responsible for finding solutions to tackle the aforementioned challenges introduced by steganography. Furthermore, their job includes analysing in detail computer programs and software, in order to obtain the data from a damaged device, trace sources of a breach or analyse electronic data [1].

Image steganography and steganalysis have been intriguing topics for researchers in the past 20 years. Multiple steganalysis techniques are used for identifying the presence of a message inside a digital file, such as signature-based, statistical, feature-based, deep learning (DL) and more [14, 32]. For most of the mentioned techniques, comparisons have been performed by various researchers [14, 24, 25, 32], with different datasets, such as IStego100K [42], BOSS [4] utilizing different categories of methods and techniques (AI, signature or statistical) [9, 14, 24, 31]. The results provide a solid foundation upon which future work in the image steganography and steganalysis domain could be built.

1.1 Problem statement

There have been papers focusing on the topic of image steganography and steganalysis [24, 31, 39, 41]. All the above-mentioned references have provided detailed information regarding the currently existing methods and algorithms. However, not many studies have concentrated on explaining the advantages and disadvantages of each existing steganalysis algorithm and technique. Therefore, part of the current research focuses on exploring and discussing the strengths and weaknesses of existing image steganalysis methods. The aim was to facilitate the work of DFI and serve as a road map indicating which approach or technique should be taken. Another limitation in the field of steganography and steganalysis constitutes the fact that not many papers compare the available forensic tools and their detection capabilities. This is another niche area that was explored in the current study. It could provide valuable information to DFI, especially regarding the JPEG format, which is one of the most used for image steganography [5, 6]. Another interesting question in the domain of steganography that could produce valuable

TScIT 39, July 7, 2023, Enschede, The Netherlands

© 2023 University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

results is, whether image features such as colour, size, peak signal-to-noise ratio (PSNR) and mean square error (MSE), indicate the presence of steganography.

1.1.1 Main goal. Explore and analyse the use of image steganography and steganalysis, to aid DFI and other interested parties.

1.1.2 Research questions. Given the problem statement and main goal, the research aimed to answer the following 3 research questions:

RQ1: To what extent are AI-based techniques more accurate in detecting steganography in digital image files, compared to the traditional signature and statistical techniques, and should DFI prefer them?

RQ2: To what extent, one of Aletheia and StegExpose steganalysis tools is more accurate (measured by dividing the sum of correctly identified images with or without hidden messages by the total pool of images) in detecting steganography in JPEG images?

RQ3: To what extent do any of the feature(s), such as size, colour, PSNR and MSE produce an indication of possible hidden information inside a JPEG image, independent of the utilized steganography tool?

2 RELATED WORK

Multiple researches [14, 16, 24, 25, 32] have been performed in the field of image steganography and steganalysis. The next lines summarize part of the most interesting ones.

Laskar and Hemachandran in 2014 [25] analysed the main existing techniques and methods used by forensic investigators in image steganography and steganalysis. Their work offers a comprehensive comparison between the image steganography techniques such as transform and spatial domain based and on the other hand visual, statistical and structural techniques for the steganalysis. This study is concentrated on the important features of each of the investigated techniques. Similarly, a paper by Nissar and Mir [32] focuses on the classification of the existing steganalysis techniques. They provide an indication with respect to the best circumstances in which certain categories (signature, statistical) of techniques should be utilized. The papers showed that the existing image steganalysis techniques still cannot universally detect hidden messages inside digital images.

A study by Karampidis et al., from 2018 [24] also focuses on the existing image steganalysis methods and techniques. The paper compares visual, statistical, spread spectrum, and universal methods used in image steganalysis. It provides insight suggesting the appropriate usage of some of the aforementioned methods, depending on the information available to the digital forensic investigator. For instance, in case only the stego object is known, then a statistical image steganalysis technique might be the most beneficial. Nevertheless, the main conclusion from the paper indicates that currently, it is impossible to utilize an image steganalysis algorithm with low computational needs and high accuracy. This is expected to be achieved once DL concepts are embedded in the image steganalysis tools.

A study by Farooq and Selwal from 2020 [14] depicts the current trends in AI and offers a systematic review of the usage of DL

principles for image steganalysis by presenting a comparative analysis of the existing DL-based methods and algorithms. The paper concentrates on the promising future of DL techniques in image steganalysis. It also identifies the importance of creating and building more robust and efficient image steganalysis models that could be utilized for the learning process by the DL image steganalysis methods.

A study by Giarimpampa from 2018 [16] focuses on examining a set of steganographic tools from both the transform and spatial domains and developing a blind image steganalysis method. The goal is to detect the presence of a hidden message inside a digital source with the lowest possible error rate. The paper emphasizes that due to the lack of standard scientific datasets, it is hard to prove that universal image steganalysis performs well. Therefore, some tools/algorithms have high detection rates for the spatial domain but do not work as effectively with the transform domain. This possesses difficulties in creating a completely decisive blind image steganalysis method for detecting steganography in images.

After examining the related literature on the topic of image steganography and steganalysis, it was evident that there were limitations in the existing techniques and tools. Currently, a universal/blind tool for detecting steganography is not yet implemented. Thus, DFI need mostly in-depth knowledge regarding the steganography tool used for embedding the secret message to obtain the message. All these problems motivated the need for the current research and especially the formulation of the first research question.

3 METHODOLOGY

With the goal to help DFI in the field of digital image steganography and steganalysis, the following procedures were taken for answering the research questions.

For the first research question, a literature review was performed on 2-3 steganalysis sub-techniques (more information in the Results RQ1 subsection) from the AI, signature or statistical domain. The papers for the literature review were found via a Google Scholar search, using "image steganalysis", "steganalysis techniques", "image steganography", "forensic" key terms. The main goal was to identify, whether AI-related techniques produce more accurate results for detecting steganography inside digital images. The main metric of interest for the comparison process was the detection accuracy (in case it was present), obtained from the consulted papers during the literature reviews [24, 28]. Furthermore, the general advantages and disadvantages of each category of techniques were also discussed.

Moving to the dataset for the second and third research questions, several steps were followed. The dataset, consisting of 30 random JPEG image files, obtained from Unsplash [40], was composed of 15 colourful (RGB) and 15 grayscale (GS) that were identical to the colourful [30], however, converted to grayscale with a free online tool called ImageOnline.co [2]. Finally, all the images were resized to 512×512 dimensions with the ILoveIMG website[21]. There were several reasons for choosing specifically JPEG images and composing a dataset of 30 images, rather than using an existing one like BOSS [4]. For instance, most of the non-paid and widely used steganography and steganalysis tools work predominantly or exclusively with JPEG images. Furthermore, it is one of the most widely

used image formats in the world [3]. Continuing with the dataset size, 30 was chosen, because it is neither big nor small to investigate and would produce significant and measurable results. In terms of the image dimension decision, 512×512 was chosen, since it is one of the smallest possible values, for which the selected steganography tools would work. Regarding the concealment of messages, 5 text sizes (6 Bytes, 53 Bytes, 125 Bytes, 529 Bytes, and 1040 Bytes) were used in order to hide text inside 16 of the 30 images. The main idea was to analyse whether the detection accuracy is influenced by the embedded message size. Especially for 529 and 1040 Bytes, the focus was to identify whether surpassing the dimension threshold would influence significantly the exposure of the hidden message.

For the second research question, using the above-mentioned dataset, 3 steganography tools for hiding information inside 16 images were utilized (most of the chosen cover images were more colourful and with bigger sizes in Bytes), with the remaining 14 being untouched. For detecting steganography inside the images, 2 steganalysis tools were used. For the purpose of this research, **StegHide**, **F5** and **Outguess** were chosen as the steganography options for concealing messages. The 3 tools were one of the most popular non-paid and compatible with JPEG images [10, 28]. Once the dataset with the stego and non-stego images was constructed for the experimentation process, then, 2 popular steganalysis tools which support steganography detection for JPEG files were used, namely **Aletheia** [26] and **StegExpose** [38]. The former offers a novel approach by implementing state-of-the-art machine learning techniques [26], whereas the second is an older tool, yet, still currently used tool by DFI [10]. Initially, it was planned to utilize **StegDetect** [27] alongside Aletheia and StegExpose, however, due to the old version of the tool and incompatibility problems, it was decided to be skipped. In terms of the actual experiment, after obtaining the stego files for each image, both steganalysis tools were run on the images, in order to check the accuracy, which was calculated by summing the correctly identified as stego images and correctly identified as non-stego images divided by the total image pool size (30). The approach was done against each steganography tool and embedded hidden message size, in order to answer the second research question.

For the third research question, given the stego images from the composed dataset, image features such as peak signal-to-noise ratio (PSNR), mean square error (MSE), colour, and image size were analysed. The first one measures the quality of the image between the original and changed image in decibels (cover and stego images), with the following formula, $PSNR = 10 \log_{10} \cdot \frac{[MAX(cr)]^2}{MSE}$, where $MAX(cr)$ is the maximum pixel value of the image (8 for grayscale, 24 for RGB)[10], whereas MSE is an estimator, measuring the differences between 2 images (cover and stego), using this formula, $MSE = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (cr_{mn} - st_{mn})^2$, where M and N are the dimensions of the image, cr_{mn} st_{mn} , are the individual pixel values from the cover and stego image, respectively [37]. To check the importance of the PSNR, MSE and image size, several steps were taken. For each of the features, the mean and standard deviation (SD) values were computed, based on the stego images with different embedded messages. Both **SD** and **mean** were chosen, since they demonstrate how the stego image deviates from the cover one

and if some of the utilized steganography tools, introduce common changes in the stego images, or possibly independent of the tool, that could be detected by DFI.

4 RESULTS & DISCUSSION

This section answers the proposed 3 research questions from the Problem statement. The answer to each question is separated into an individual subsection, where in detail the findings are discussed and analysed.

4.1 Results RQ1

The **first research question** was dedicated to the comparison of AI, signature and statistical techniques used in steganalysis and identifying whether a certain category is more accurate in detecting steganography in digital images. To answer that, the preparation phase involved an extensive reading process on the available literature, as described in the Methodology section. The following studies [13, 14, 23–25, 28, 32, 36] were examined and analysed, by observing the advantages, disadvantages, and use cases of techniques. From the above-referenced papers, a minimum of 2 and a maximum of 3 techniques from each domain (signature, statistical and AI-based) were chosen.

4.1.1 Signature based. Signature steganalysis is a method, used by DFI and other cybersecurity specialists, to search for steganography in digital images by observing repetitive patterns, introduced by common steganography tools [24], with the help of computer programs such as **HxD** that show the hexadecimal representation of the image. It has been found in papers [24, 32] that, the presence of specific characters, also referred to as signatures, could indicate that the image is stego (an image that has an embedded secret message inside it). For example, 5B 3B 31 53 00 is a common byte sequence embedded at the end of JPEG files (when the files are represented as sequences of bytes) that indicates the image might contain steganography [24, 32], or "CDN" appearing somewhere in the byte sequence representation of the image [23, 24]. Following this general introduction, in the next lines, more information and details are presented regarding the results. The selected 2 main sub-techniques from the signature domain, that are not AI-based, were specific signature steganalysis (**SSS**) [32] & universal signature steganalysis (**USS**) [32].

Specific signature steganalysis is a technique, developed to tackle a concrete steganography algorithm, such as LSB and F5 [17]. Some steganography algorithms leave traces as the ones identified in the previous paragraph (5B 3B 31 53 00, CDN, etc.). Therefore, SSS algorithms provide the option to detect such traces, in case they are embedded in the stego file. Following that, the biggest limitation of the SSS methods is the required exact knowledge regarding which tool has been used.

Universal signature steganalysis focuses on identifying steganography in images, independent of the utilized algorithm for concealing the message. An example of such an algorithm has been proposed by Fridrich [15], which identifies steganography in JPEG images, by dividing the image into 8×8 blocks and extracting the quantization matrix from the discrete cosine transform (DCT) coefficients, which is compared with the standard JPEG quantization

table. This technique has been tested and referenced as reliable, capable of detecting hidden information even in case of flipping the LSB of one-pixel [32]. Regarding the disadvantages, in case of rescaling of the image, the JPEG signature may be lost [32] and the algorithm only works on JPEG images [23, 24].

During the review part of the existing studies, not enough information was found on the topic of detection accuracy in signature-based image steganalysis. However, from the consulted literature [23, 24, 32] several conclusions can be made. In case the steganography tool leaves signature patterns, similar to **Masker** or **JPEGx**, then signature-based steganalysis techniques, especially the specific ones, can produce decent results that signify the presence of steganography. For the universal ones, only the algorithm proposed by Friedrich [15] was found to be useful, and very reliable, but it works solely in the JPEG domain.

4.1.2 Statistical based. Statistical techniques are a category in image steganalysis that analyses the properties of the image by utilizing statistical methods like Chi-square analysis, Raw Quick Pair (RQP), RS steganalysis [13, 14, 23–25, 32]. Most of the statistical techniques, including the aforementioned methods, have been applied in various currently used steganalysis tools, such as **StegDetect** and **VSL – Virtual Steganography Laboratory**. An important point that researchers have discovered in 2 papers [24, 32], is that in case only the stego object is known, then statistical steganalysis outperforms the signature one in terms of effectiveness and robustness [24, 32]. Following the general introduction and aim of the statistical techniques, in the next paragraphs, more information will follow on the obtained results from the research done on 3 popular and utilized statistical steganalysis sub-techniques [18]. Based on the popularity and information available in the literature, the chosen sub-techniques were Chi-square analysis, RS steganalysis and Raw Quick Pair (RQP) [20, 24, 25].

Starting with the Chi-square analysis, it concentrates on Pairs of Values (POVs) exchanged during the secret data embedding inside the digital image [20]. The POVs could be pixel values, quantized DCT coefficients or pallet indices, depending on the steganography algorithm used [24]. It has been found that, if there is hidden information inside an image, then the embedded information alters the histogram of colour frequencies inside the image in a particular way. This constitutes a change in the least significant bits of the image, and the frequencies of POVs are prone to be situated further from the mean POV [24, 25]. Overall, the chi-square analysis detects reliably messages that are sequentially embedded, however, doesn't provide high accuracy for randomized ones [24, 32].

RS steganalysis is another important technique, used to detect steganography inside images. It has been utilized for detecting the least significant bit (LSB) steganography in both colourful and grayscale images [20, 24, 25, 32]. The RS analysis divides the image into groups and measures the noise within each group [24]. The groups could be classified as either "regular" or "singular", depending on the results obtained after the LSBs of fixed pixels in the groups have been flipped and analysed whether the noise within the corresponding groups has increased or decreased [23, 32]. Regarding the effectiveness of the RS steganalysis, most of the examined papers during the literature review have concluded that it is more reliable

in detecting steganography compared to the Chi-square method [24, 32], however in case the embedded message inside the image is less than 0.005 bits per pixel, it is undetectable by RS steganalysis [32].

Raw Quick Pair (RQP) is a statistical steganalysis technique, that detects steganography in 24-bit colour images by analysing close colour pairs that only differ in their LSB [23]. The method has been shown to work well, especially in case the number of unique colours in the cover image is less than 30 % of the number of pixels [32]. The most notable disadvantage of RQP is the compatibility to work with only 24-bit colourful images and not with grayscale ones [24, 32].

After examining the above-mentioned statistical sub-techniques and the statistical techniques in general, several important points could be emphasized. First, it has been suggested by certain researchers [24, 32] that statistical methods are more accurate in detecting steganography, compared to the signature ones, in case only the stego object is accessible [32]. For the detection accuracy, similarly to the signature-based techniques, in the examined literature no concrete information related to the accuracy of the sub-techniques was found.

4.1.3 AI-based. AI-based techniques are a set of methods used in steganalysis that aim to detect steganography in images, by using different types of AI algorithms and approaches, including machine learning (ME) and deep learning (DL) [29]. Generally, the AI approaches in image steganalysis involve 2 steps, initially to extract interesting and valuable features (could be size, colour, entropy, histogram) from the stego object, which are then compared with the cover image. Afterwards, the detection mechanism is built upon the examined data from the first step [23, 29]. As discussed in the previous sections, steganalysis has different domains of techniques for detecting steganography. Signature and statistical are one of the most popular, however, in most cases these techniques require information regarding the steganography algorithm used for embedding the secret message inside the image [14, 24]. An alternative category of techniques that do not utilize any additional information for the embedding algorithm is known as universal or blind steganalysis [14, 24]. These techniques are in most cases implemented together with AI-based approaches. In the next paragraphs, the information obtained regarding 3 popular and widely referenced blind/universal steganalysis techniques will be discussed and analysed.

A paper by Zhang et al. [43], proposed in 2013 a universal steganalysis algorithm, based on sparse representation, connected to finding steganography in JPEG images. The algorithm concentrates on transporting the main body of information with the minimum possible amount of information, in order to solve the information processing [24, 28, 43]. This method has proven in experimental results to overcome several shortages which are common for SVM-based classifiers, like achieving high detection accuracy (around 90 % [28, 43] compared to SVM) and solving the over-fitting problem of traditional classifiers. Furthermore, the algorithm used by Zhang et al. was demonstrated to be more effective when the image that needs to be detected has a Gaussian or Salt Pepper noise. [24, 28, 43]. Nevertheless, the most noticeable disadvantage of the algorithm is the limitation to perform the steganalysis on only JPEG images.

A paper by Zong et al., proposed in 2012 for blind steganalysis in JPEG images, concentrating on the correlation of inter- and intra-wavelet sub-bands in the wavelet domain and feature extraction from co-occurrence matrix [28, 44]. The algorithm depicts high-detection capabilities for several popular steganography tools (F5, Jsteg, Outguess) with around 95 % detection accuracy and good detection capability for double compressed images [24, 28, 44]. Similarly to the method utilized by Zhang et al., the solution is limited to the JPEG domain only.

A paper by Desai et al., proposed in 2016 a universal image steganalysis using Fisher Criterion and ANOVA techniques, by extracting features from the wavelet sub-bands and binary similarity patterns, obtained from DCT domain [12, 23, 24]. The proposed method, demonstrates more than promising results with an overall accuracy of 97%, obtained, against steganography algorithms such as Outguess and F5 [12].

To conclude the AI techniques section, during the literature review, numerous papers were examined, mainly from the universal/blind steganalysis techniques domain. Overall, it was found that, currently, there does not exist a complete solution which has a high detecting rate independent of the steganography algorithm and image type used [24, 28]. The sub-techniques that were examined, namely by Zhang et al. and Zhang et al., which are working solely in the JPEG, whereas for Desai et al. no specific information was found regarding the working domain, demonstrated excellent results, against famous steganalysis algorithms such as F5 and Outguess. However, from the literature review, it was found that no universal/blind technique exists that could detect steganography in any image with high accuracy and low computational needs [23, 24].

4.1.4 Final results. In the previous subsections, the results obtained from the literature review regarding signature, statistical and AI-based techniques, their advantages, and disadvantages were discussed. The following flowchart depicts the most important observations and could serve as a road map **Fig. 1** for DFI. Overall, it was examined that, currently, no best technique exists in the steganalysis domain that could detect steganography independent of the algorithm and image type used. In terms of the research question, no details regarding the detection accuracy were demonstrated for the signature and statistical techniques, whereas AI-based information was present in the literature. Despite the lack of concrete numbers for the detection accuracy in statistical- and signature-based techniques, several recommendations could be followed. DFI should prefer specific algorithms from the statistical domain when information is present regarding the steganography tool, utilized for embedding the message. Furthermore, forensic investigators should generally refrain from using signature techniques since the statistical ones are more accurate. Regarding AI-based techniques, such should be used when specific information is not available, however, these methods do not guarantee high detection in all cases. Finally, from the examined sub-techniques, it cannot be stated that, overall, AI-based techniques are more accurate than a signature or statistics ones in detecting steganography.

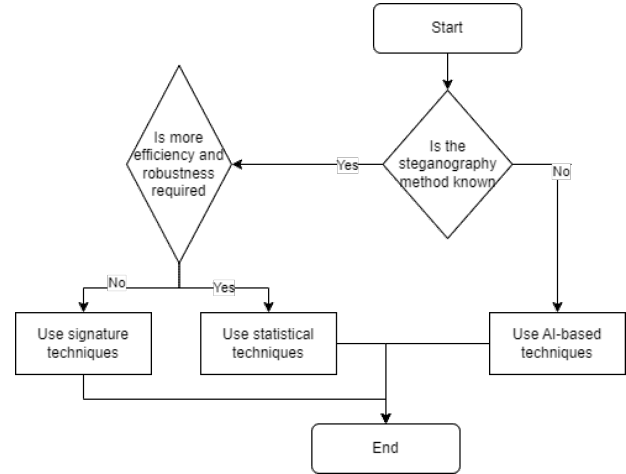


Fig. 1. Steganalysis technique selection



Fig. 2. Sample JPEG image dataset

























4.2 Results RQ2

To answer the **second research question**, the steps, identified in the methodology section, were followed. Starting with the images, **Fig. 2** shows a sample of the JPEG image dataset. Of all the 30 images, 16 images were used as cover ones. For each of the cover images, 5 different stego images were obtained, due to the 5 variants of embedded hidden messages. **Table 1** shows examples of the messages, where for m4 and m5, the "..." signifies that more text is present after the three dots. **Table 2** illustrates the cover images and the respective stego images based on the embedded message size (m1 to m5) and steganography tool (F5, Outguess, Steghide) used. Regarding the experiment, in the following subsections, information regarding the detection capabilities of both steganalysis tools will be discussed, so that it could be identified which of the tools is more accurate in detecting steganography. Furthermore, in order to replicate the job of the DFI and produce valuable results for

Table 1. Secret messages

Message Acronym	Message Text	Size
m1	Secret	6B
m2	This message is secret and not supposed to be found.	53B
m3	The idea of the current research is to find how certain steganalysis tools perform against common steganography algorithms.	125B
m4	Now is the winter of our discontent Made glorious summer by this sun of York ...	529B
m5	Now is the winter of our discontent Made glorious summer by this sun of York; And all the clouds that lour'd upon our house In the deep bosom of the ocean buried. Now are our brows bound with victorious wreaths...	1040B

Table 2. Cover & Stego Images

Tool	Cover Image	Stego Image m1	Stego Image m3	Stego Image m5
F5				
F5				
Outguess				
Outguess				
Steghide				
Steghide				

them, both Aletheia and StegExpose tools were used, without prior knowledge of the embedded algorithm or steganography tool used.

4.2.1 StegExpose. Starting with StegExpose [38], the tool achieved, overall, 50% accuracy, independent of the steganography tools used, or embedded message length, as evident from **Table 3**. From the conducted experiment, StegExpose was depicted to have correctly identified all images without embedded messages as non-steganographic.

Table 3. StegExpose detection rate

StegExpose detection accuracy						
Secret Message	Outguess		F5		Steghide	
	GS	RGB	GS	RGB	GS	RGB
m1	50%	50%	50%	50%	50%	50%
m2	50%	50%	50%	50%	50%	50%
m3	50%	50%	50%	50%	50%	50%
m4	50%	50%	50%	50%	50%	50%
m5	50%	50%	50%	50%	50%	50%

However, for the actual steganographic images, both grayscale and colourful, all the images were labelled as non-steganographic, which corresponds to 50 % false negative rate ($FNR = \frac{FN}{FN+TP}$) and unreliability of the tool to detect steganography. In contrast, StegExpose achieves a 100% accuracy on the test dataset provided by the developer, which consists of cover and stego images in the PNG format. Therefore, from the experimentation process, solely based on the utilized dataset of colourful and grayscale JPEG images and the provided PNG test dataset, it could be stated that StegExpose shouldn't be utilized by DFI for detecting steganography, at least in the JPEG domain, but could demonstrate reliable results in the PNG domain.

Table 4. Aletheia detection rate

Aletheia detection accuracy						
Secret Message	Outguess		F5		Steghide	
	GS	RGB	GS	RGB	GS	RGB
m1	26.6%	33.3%	46.6%	40.0%	33.3%	33.3%
m2	26.6%	33.3%	46.6%	40.0%	33.3%	33.3%
m3	26.6%	33.3%	46.6%	40.0%	33.3%	33.3%
m4	40.0%	33.3%	46.6%	46.6%	40.0%	33.3%
m5	40.0%	46.6%	60.0%	46.6%	33.3%	33.3%

4.2.2 Aletheia. Aletheia, is a novel steganalysis tool with multiple possible options, such as performing calibration, RS and other attacks [19]. However, for the scope of the current research, the default exploratory "auto" detection functionality was utilized. It has built-in detection against 4 common steganography tools, namely Outguess, nsF5, Steghide and J-UNIWARD [26]. As it is evident, 3 of the common steganography tools corresponded to what was used for concealing the messages inside the JPEG images. Nevertheless, all the probability results were utilized for computing the detection accuracy, by taking the average of all probabilities. Such an approach was chosen to be utilized since a digital forensic investigator is in most cases not aware of the steganography algorithm used to embed a secret message inside an image. The following formula depicts the calculation of the accuracy:

$$P = \frac{(P_{Outguess} + P_{nsF5} + P_{Steghide} + P_{J-UNIWARD})}{4} \quad (1)$$

If $P \geq 50\%$, the image is identified as stego $P < 50\%$, the image is labelled as non-stego.







The experiment with Aletheia, demonstrated that the tool obtained different results, depending on the steganography tool used for embedding the secret message, **Table 4**. For the first 3 sizes of hidden messages with lengths 6B, 53B and 125B, the results per tool are identical, independent of the colour of the image (RGB or GS), achieving the highest accuracy against F5 and lowest against Outguess. Regarding the larger secret message length size of 529B and 1040B, Aletheia has still the highest detection rate against F5, whereas against Steghide, it has the lowest. Moreover, given the results, several aspects could be observed. Aletheia performed worse in detecting secret messages in grayscale images, embedded with Outguess, compared to the RGB ones. On the other hand, F5 and Steghide, in both cases the detection accuracy in grayscale images is bigger or equal to the colourful ones. Furthermore, given the secret message lengths and detection accuracy, it is evident that after surpassing the 512 threshold, the detection accuracy in most cases is improved, except for Steghide, which would indicate that the bigger the hidden message is, the higher the detection rate Aletheia could achieve.

4.2.3 Main observations. In summary, from the utilized popular steganography and steganalysis tools on the JPEG dataset, it has been observed that both steganalysis, Aletheia and StegExpose tools, are still unreliable. With StegExpose, providing 50% false negative rates, which raised the question, of whether such a tool could be useful in any circumstances. On the other hand, Aletheia, depicted different results and findings, however with an average accuracy of 37.96% against Outguess, Steghide and F5, which could puzzle researchers whether they should trust it. Therefore, to answer the research question, based solely on the accuracy metric, StegExpose has a higher score compared to Aletheia, however with low reliability, given the false negative rate. Following that, several recommendations could be shared with DFI. First and foremost, StegExpose should be avoided as the choice of steganalysis tool, based on the observed unreliability and more attention should be dedicated to Aletheia. Given the default functionality of the latter one, which provides information regarding the possibility of a specific steganography tool that has been used, it could be used in an exploratory phase in detecting steganography in JPEG image files. Subsequently, in case the steganography algorithm is F5 or Steghide, a concrete set of steps could be utilized by DFI, such as DL models and calibration attacks that perform well in detecting steganography, against the aforementioned tools [19].

4.3 Results RQ3

To answer the **third research question**, the same dataset from RQ2 was utilized, however, consisting solely of stego images. The metrics evaluation was performed on the images, to detect whether values such as peak signal-to-noise ratio (PSNR), mean square error (MSE), colour and size indicate if an image is stego, or are totally independent in that respect. The results were analysed and divided into separate subsections based on the steganography tool that was utilized. Only a sample of 6 images, 3 colourful (*BC*, *CC*, *SC*) and their respective grayscale (*BG*, *CG*, *SG*) were analysed, **Table 5**. In the table, the “Average Stego Image Size in Bytes” column, indicates the average value that was obtained, given the sizes of stego images with

Table 5. Sample Stego Images Values

Image Name	Average Stego Image Size in Bytes	Type	Colour	Stego Image
BC	19859.0	jpg	RGB	
BG	19341.6	jpg	GS	
CC	36188.0	jpg	RGB	
CG	32496.4	jpg	GS	
SC	45863.0	jpg	RGB	
SG	42713.4	jpg	GS	

6B, 53B, 125B, 529B and 1040B embedded secret messages. Regarding 2 of the chosen features, PSNR and MSE, several assumptions were made. Based on the definition of MSE (given in the methodology section), the closer a value is to 0, the closer the stego image is to the cover one. Regarding PSNR, following the formula, since MSE appears in the denominator of the logarithm base 10 part, it can be concluded, that the bigger the PSNR is, the better quality the stego image has. Therefore, to analyse the results more systematically, statistical techniques such as mean values and standard deviation were computed. They should serve to check whether the embedded message size significantly alters the stego image size. In the next subsections, the results are analysed, based on the steganography tool.

Table 6. F5 stego images feature analysis results

Image Name	Colour	Mean Size in Bytes	SD Sizes in Bytes	Mean MSE	SD MSE	Mean PSNR in dB	SD PSNR in dB
BG	GS	19341.60	377.54	0.69	0.40	50.34	2.24
BC	RGB	19859.00	412.04	1.35	0.43	47.01	1.25
CG	GS	32496.40	203.42	0.80	0.34	49.40	1.59
CC	RGB	36188.00	182.61	1.93	0.44	45.37	0.91
SG	GS	42713.40	150.46	0.98	0.35	48.45	1.38
SC	RGB	45863.00	164.11	1.86	0.39	45.52	0.83

4.3.1 F5. From the chosen sample of 6 images, several interesting observations were found **Table 6**. The grayscale stego (GS) images have a smaller size in terms of Bytes compared to the respective colourful ones. However, such a detail may be viewed as expected, since the former one has 1 channel with 8 bits, whereas the latter one

has 3 channels with 8 bits each. Another observation related to the size was the increase in the hidden messages' lengths, correlating to the respective decrease in the stego image sizes. Regarding the MSE, from the chosen sample of 6 images, it was demonstrated that the MSE value increased with the increase of the embedded message size. Furthermore, especially for the **RGB** images, the mean MSE value is higher than the respective **GS** images, indicating that the **RGB** stego image quality is compromised and worse, compared to **GS**. Regarding the PSNR, the same trends can be observed, with **GS** stego images being with better quality.

Table 7. Outguess feature analysis results

Image Name	Colour	Mean Size in Bytes	SD Sizes in Bytes	Mean MSE	SD MSE	Mean PSNR in dB	SD PSNR in dB
BG	GS	18751.40	62.28	1.00	0.51	48.64	2.04
BC	RGB	19365.60	71.46	1.51	0.72	46.77	1.88
CG	GS	30787.20	33.84	1.29	0.54	47.36	1.67
CC	RGB	34424.00	40.98	2.62	0.79	44.13	1.21
SG	GS	40361.00	19.67	1.52	0.55	46.56	1.45
SC	RGB	43570.40	36.09	2.60	0.76	44.14	1.16

4.3.2 Outguess. Compared to F5, Outguess, delivered different values for some examined features **Table 7**. The sizes of the stego images were still smaller than the respective cover images, however, compared to F5, with the increase of the embedded message size, the stego size also increased. Moreover, compared to F5, the stego sizes with different secret messages were closer in terms of size than the respective ones in F5. As a result, the standard deviation of the aforementioned feature was approximately 2 times smaller, compared to F5, illustrating that Outguess hides the messages more efficiently. In terms of MSE and PSNR, the dataset from Outguess the values are higher and lower, compared to F5, indicating that F5 delivers better quality for the stego images.

Table 8. Steghide feature analysis results

Image Name	Colour	Mean Size in Bytes	SD Sizes in Bytes	Mean MSE	SD MSE	Mean PSNR in dB	SD PSNR in dB
BG	GS	27099.40	25.78	0.05	0.03	61.62	2.74
BC	RGB	19915.80	32.99	0.30	0.17	54.18	2.69
CG	GS	44737.20	21.19	0.06	0.04	61.19	2.81
CC	RGB	35839.80	12.67	0.36	0.21	53.49	2.88
SG	GS	58438.00	17.74	0.06	0.04	60.97	2.76
SC	RGB	45328.40	15.12	0.32	0.19	54.02	3.03

4.3.3 Steghide. Steghide, showed the most interesting results, compared to F5 and Outguess **Table 8**. In contrast to F5 and Outguess's stego images, Steghide had a bigger image size in terms of Bytes for the grayscale stego images than the colourful ones. Furthermore, the average Bytes size of all stego images was the highest, compared to their respective counterparts in F5 and Outguess, indicating that

Steghide doesn't perform efficiently with grayscale images. However, the standard deviation of the stego image sizes with different embedded messages was observed to be the lowest, depicting that the size of the stego image wasn't influenced significantly by the increase of the concealed message size. The MSE and PSNR values, were the lowest and highest, respectively, from all examined tools, indicating that the image quality has been the best.

4.3.4 Main observations. From the experimentation process, regarding the third research question, several conclusions could be drawn. Each of the utilized steganography tools affected differently the dataset of images. For instance, Steghide produced larger sizes for grayscale stego images, compared to F5 and Outguess, whereas F5 had the highest value for the standard deviation of the stego image sizes, indicating the inefficient embedding of secret messages inside the cover images. In terms of the MSE and PSNR values, no general information was found, that would indicate an image being stego. From the conducted experiment, Outguess demonstrated the highest and lowest average values for MSE and PSNR, in contrast, Steghide depicted the lowest and highest, from all the 3 steganography tools, respectively. Finally, to answer the research question, based on the conducted experiment, it can be stated that none of the tested features (PSNR, MSE, colour and size), directly indicate the existence of a hidden message inside an image. Following that, DFI cannot rely on any of the tested features. This is, because none of the 3 examined steganography tools, provoked similar or predictable changes in the tested features. Therefore, PSNR, MSE, colour and image size should not be used as markers to discover steganography in JPEG images.

5 CONCLUSION & FUTURE WORK

The current study focused on different aspects of the domain of steganography and steganalysis. The general aim of the research was to aid DFI in tackling challenges in the field of image steganography and steganalysis. The results showed that AI-based blind/universal steganalysis techniques are not better than statistical or signature-based. Nevertheless, blind/universal techniques should be used when the utilized steganography algorithm is unknown. On the other hand, statistical and signature-based should be utilized when the concrete steganographic algorithm is known. Regarding the tool comparison, Aletheia was identified as more reliable, despite achieving lower accuracy than StegExpose. Finally, for the study of the metrics, no concrete information was found that identified the size, colour, MSE or PSNR as an indication of the existence of a hidden message inside a JPEG image.

For the future work, several new directions can be followed. For example, more emphasis could be focused on experimenting with different image formats (PNG, GIF). Furthermore, some paid, contemporary steganography and steganalysis tools, could be studied and compared, on a dataset consisting of images with different dimensions. From such newly followed approaches, more insightful conclusions could be made, regarding the image steganography and steganalysis domain.

REFERENCES

- [1] 2023. How to Become a Digital Forensic Investigator. <https://www.wgu.edu/career-guide/information-technology/digital-forensic-investigator-career.html>
- [2] 2023. Make grayscale image online - Free tool. <https://grayscale.imageonline.co/>
- [3] 2023. Usage Statistics of Image File Formats for Websites, June 2023. https://w3techs.com/technologies/overview/image_format
- [4] Patrick Bas, Tomáš Filler, and Tomáš Pevný. 2011. "Break Our Steganographic System": The Ins and Outs of Organizing BOSS. In *Information Hiding*, Tomáš Filler, Tomáš Pevný, Scott Craver, and Andrew Ker (Eds.). Vol. 6958. Springer Berlin Heidelberg, Berlin, Heidelberg, 59–70. https://doi.org/10.1007/978-3-642-24178-9_5 Series Title: Lecture Notes in Computer Science.
- [5] Black Hat. 2014. Advanced JPEG Steganography and Detection by John Ortiz. <https://www.youtube.com/watch?v=BQPKRlbfVEs>
- [6] Jan Butora, Pauline Puteaux, and Patrick Bas. 2022. Errorless Robust JPEG Steganography using Outputs of JPEG Coders. <https://arxiv.org/abs/2211.04750v1>
- [7] Lori Cameron. [n. d.]. With Cryptography Easier to Detect, Cybercriminals Now Hide Malware in Plain Sight. Call It Steganography. Here's How It Works. <https://www.computer.org/publications/tech-news/research/how-steganography-works>
- [8] Abbas Cheddad, Joan Condell, Kevin Curran, and Paul Mc Kevitt. 2010. Digital image steganography: Survey and analysis of current methods. *Signal Processing* 90, 3 (March 2010), 727–752. <https://doi.org/10.1016/j.sigpro.2009.08.010>
- [9] Shaveta Chutani and Anjali Goyal. 2019. A review of forensic approaches to digital image Steganalysis. *Multimedia Tools and Applications* 78, 13 (July 2019), 18169–18204. <https://doi.org/10.1007/s11042-019-7217-0>
- [10] Mukesh Dalal and Mamta Juneja. 2021. Steganography and Steganalysis (in digital forensics): a Cybersecurity guide. *Multimedia Tools and Applications* 80, 4 (Feb. 2021), 5723–5771. <https://doi.org/10.1007/s11042-020-09929-9>
- [11] Keith Debattista. 2010. The Threats of Steganography. <https://techtalk.gfi.com/threats-steganography/> Section: Network.
- [12] Madhavi B Desai, S V Patel, and Bhumi Prajapati. 2016. ANOVA and Fisher Criterion based Feature Selection for Lower Dimensional Universal Image Steganalysis. (2016).
- [13] Wafa M. Eid, Sarah S. Alotaibi, Hasna M. Alqahtani, and Sahar Q. Saleh. 2022. Digital Image Steganalysis: Current Methodologies and Future Challenges. *IEEE Access* 10 (2022), 92321–92336. <https://doi.org/10.1109/ACCESS.2022.3202905> Conference Name: IEEE Access.
- [14] Numrena Farooq and Arvind Selwal. 2023. Image steganalysis using deep learning: a systematic review and open research challenges. *Journal of Ambient Intelligence and Humanized Computing* (March 2023). <https://doi.org/10.1007/s12652-023-04591-z>
- [15] Jessica Fridrich, Miroslav Goljan, and Rui Du. 2001. Steganalysis based on JPEG compatibility. In *Multimedia Systems and Applications IV*, Vol. 4518. SPIE, 275–280. <https://doi.org/10.1117/12.448213>
- [16] Despoina Giarpampapa. 2018. *Blind Image Steganalytic Optimization by using Machine Learning*. <https://urn.kb.se/resolve?urn=urn:nbn:se:hh:diva-38150>
- [17] Baraa Tareq Hammad, Ismail Taha Ahmed, and Norziana Jamil. 2022. A Steganalysis Classification Algorithm Based on Distinctive Texture Features. *Symmetry* 14, 2 (Feb. 2022), 236. <https://doi.org/10.3390/sym14020236> Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.
- [18] Muhammad Hassan, Murad Amin, and Suzan Mahdi. 2020. STEGANALYSIS TECHNIQUES AND COMPARISON OF AVAILABLE SOFTWARES. In *Proceedings of the Proceedings of the 1st International Multi-Disciplinary Conference Theme: Sustainable Development and Smart Planning, IMDC-SDSP 2020, Cyperspace, 28-30 June 2020*. EAI, Cyberspace. <https://doi.org/10.4108/eai.28-6-2020.2297970>
- [19] Daniel Lerch Hostalot. [n. d.]. Introduction to steganalysis with Aletheia. <https://daniellerch.me/stego/aletheia/intro-en/>
- [20] Mehdi Hussain, Aimuddin Wahid Abdul Wahab, Yamani Idna Bin Idris, Anthony T. S. Ho, and Ki-Hyun Jung. 2018. Image steganography in spatial domain: A survey. *Signal Processing: Image Communication* 65 (July 2018), 46–66. <https://doi.org/10.1016/j.image.2018.03.012>
- [21] iLovePDF. 2023. Resize multiple images at once! <https://www.iloveimg.com/resize-image>
- [22] David Kahn. 1996. The history of steganography. In *Information Hiding (Lecture Notes in Computer Science)*, Ross Anderson (Ed.). Springer, Berlin, Heidelberg, 1–5. https://doi.org/10.1007/3-540-61996-8_27
- [23] Konstantinos Karampidis. 2020. *Image Steganalysis for Digital Forensics*. Ph. D. Dissertation. UNIVERSITY OF THE AEGEAN.
- [24] Konstantinos Karampidis, Ergina Kavallieratou, and Giorgos Papadourakis. 2018. A review of image steganalysis techniques for digital forensics. *Journal of Information Security and Applications* 40 (June 2018), 217–235. <https://doi.org/10.1016/j.jisa.2018.04.005>
- [25] Shamim Ahmed Laskar and Kattamanchi Hemachandran. 2014. A Review on Image Steganalysis techniques for Attacking Steganography. *International Journal of Engineering Research* 3, 1 (2014).
- [26] Daniel Lerch-Hostalot. 2021. Aletheia. <https://doi.org/10.5281/zenodo.4655945>
- [27] Abel Luck. 2023. stegdetect. <https://github.com/abeluck/stegdetect> original-date: 2013-04-09T09:32:34Z.
- [28] Sherif MBadr, Goada Ismaail, and Ashgan H. Khalil. 2014. A Review on Steganalysis Techniques: From Image Format Point of View. *International Journal of Computer Applications* 102, 4 (Sept. 2014), 11–19. <https://doi.org/10.5120/17802-8617>
- [29] Prateek Mehta, Akshay Nair, Soumya Edappilly, Khushboo Manik, and Sunita Sahu. 2022. A Comprehensive Study of AI-Based Steganalysis Techniques on Image and Text Documents. In *Advances in Data and Information Sciences (Lecture Notes in Networks and Systems)*, Shailesh Tiwari, Munesh C. Trivedi, Mohan Lal Kolhe, K.K. Mishra, and Brajesh Kumar Singh (Eds.). Springer, Singapore, 53–63. https://doi.org/10.1007/978-981-16-5689-7_5
- [30] Kristiyan Michaylov. 2023. Image dataset - Google Drive. <https://drive.google.com/drive/folders/1qQhRQeBUV6s8on27BnhOjyhXfM2osza>
- [31] Ramadhan Mstafa and Christian Bach. 2013. Information Hiding in Images Using Steganography Techniques. <https://doi.org/10.13140/RG.2.1.1350.9360>
- [32] Arooj Nissar and A.H. Mir. 2010. Classification of steganalysis techniques: A study. *Digital Signal Processing* 20, 6 (Dec. 2010), 1758–1770. <https://doi.org/10.1016/j.dsp.2010.02.003>
- [33] Masoud Nosrati, Ronak Karimi, and Mehdi Hariri. 2011. An introduction to steganography methods. *World Applied Programming* 1 (Aug. 2011), 191–195.
- [34] Tom Olzak. 2019. The Security Challenges and Defense of Hidden Data. <https://www.spiceworks.com/it-security/data-security/articles/the-security-challenges-and-defense-of-hidden-data/>
- [35] Rainer Poisel and Simon Tjoa. 2011. Forensics Investigations of Multimedia Data: A Review of the State-of-the-Art. In *2011 Sixth International Conference on IT Security Incident Management and IT Forensics*. 48–61. <https://doi.org/10.1109/IMF.2011.14>
- [36] Feng Ruan, Xing Zhang, Dawei Zhu, Zhanyang Xu, Shaohua Wan, and Lianyong Qi. 2020. Deep learning for real-time image steganalysis: a survey. *Journal of Real-Time Image Processing* 17, 1 (Feb. 2020), 149–160. <https://doi.org/10.1007/s11554-019-00915-5>
- [37] Umme Sara, Morium Akter, and Mohammad Shorif Uddin. 2019. Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study. *Journal of Computer and Communications* 07, 03 (March 2019), 8. <https://doi.org/10.4236/jcc.2019.73002> Number: 03 Publisher: Scientific Research Publishing.
- [38] snoop. 2023. StegExpose. <https://github.com/b3dk7/StegExpose> original-date: 2014-08-03T08:38:52Z.
- [39] Nandhini Subramanian, Omar Elharrouss, Somaya Al-Maadeed, and Ahmed Bouridane. 2021. Image Steganography: A Review of the Recent Advances. *IEEE Access* 9 (2021), 23409–23423. <https://doi.org/10.1109/ACCESS.2021.3053998> Conference Name: IEEE Access.
- [40] Unsplash. 2023. Beautiful Free Images & Pictures | Unsplash. <https://unsplash.com/>
- [41] Muhammad Arslan Usman and Muhammad Rehan Usman. 2018. Using image steganography for providing enhanced medical data security. In *2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC)*. 1–4. <https://doi.org/10.1109/CCNC.2018.8319263> ISSN: 2331-9860.
- [42] YangzI THU. 2023. IStego100K. <https://github.com/YangzITHU/IStego100K> original-date: 2019-07-10T03:03:05Z.
- [43] Zhuang Zhang, Donghui Hu, Yang Yang, and Bin Su. 2013. A Universal Digital Image Steganalysis Method Based on Sparse Representation. In *2013 Ninth International Conference on Computational Intelligence and Security*. 437–441. <https://doi.org/10.1109/CIS.2013.99>
- [44] Han Zong, Fen-lin Liu, and Xiang-yang Luo. 2012. Blind image steganalysis based on wavelet coefficient correlation. *Digital Investigation* 9, 1 (June 2012), 58–68. <https://doi.org/10.1016/j.diin.2012.02.003>