Forensic Science International: Digital Investigation 45 (2023) 301515

Contents lists available at ScienceDirect

Forensic Science International: Digital Investigation

journal homepage: www.elsevier.com/locate/fsidi

Benford's law applied to digital forensic analysis

Pedro Fernandes [a, c, *], Mário Antunes [a, b]

[a] School of Technology and Management, Polytechnic of Leiria, Morro do Lena, Alto do Vieiro, Leiria, 2411-911, Portugal
[b] INESC TEC, CRACS, R. Campo Alegre 1021/1055, Porto, 4169-007, Portugal
[c] Technological University of the Shannon, Moylish, Limerick, V94 EC5T, Ireland

ARTICLE INFO

Article history:
Received 3 August 2022
Received in revised form
7 January 2023
Accepted 12 January 2023
Available online 30 January 2023

Keywords:
Benford's law
Digital forensics
Digital images manipulation
First digits law
Statistical coefficient correlation
Pearson's correlation

ABSTRACT

Tampered digital multimedia content has been increasingly used in a wide set of cyberattacks, challenging criminal investigations and law enforcement authorities. The motivations are immense and range from the attempt to manipulate public opinion by disseminating fake news to digital kidnapping and ransomware, to mention a few cybercrimes that use this medium as a means of propagation.

Digital forensics has recently incorporated a set of computational learning-based tools to automatically detect manipulations in digital multimedia content. Despite the promising results attained by machine learning and deep learning methods, these techniques require demanding computational resources and make digital forensic analysis and investigation expensive. Applied statistics techniques have also been applied to automatically detect anomalies and manipulations in digital multimedia content by statistically analysing the patterns and features. These techniques are computationally faster and have been applied isolated or as a member of a classifier committee to boost the overall artefact classification.

This paper describes a statistical model based on Benford's Law and the results obtained with a dataset of 18000 photos, being 9000 authentic and the remaining manipulated.

Benford's Law dates from the 18th century and has been successfully adopted in digital forensics, namely in fraud detection. In the present investigation, Benford's law was applied to a set of features (colours, textures) extracted from digital images. After extracting the first digits, the frequency with which they occurred in the set of values obtained from that extraction was calculated. This process allowed focusing the investigation on the behaviour with which the frequency of each digit occurred in comparison with the frequency expected by Benford's law.

The method proposed in this paper for applying Benford's Law uses *Pearson's* and *Spearman's* correlations and *Cramer-Von Mises* (CVM) fitting model, applied to the first digit of a number consisting of several digits, obtained by extracting digital photos features through *Fast Fourier Transform* (*FFT*) method.

The overall results obtained, although not exceeding those attained by machine learning approaches, namely Support *Vector Machines (SVM)* and *Convolutional Neural Networks (CNN)*, are promising, reaching an average F1-score of 90.47% when using Pearson correlation. With non-parametric approaches, namely Spearman correlation and CVM fitting model, an F1-Score of 56.55% and 76.61% were obtained respectively. Furthermore, the *Pearson's* model showed the highest homogeneity compared to the *Spearman's* and *CVM* models in detecting manipulated images, 8526, and authentic ones, 7662, due to the strong correlation between the frequencies of each digit and the frequency expected by Benford's law.

The results were obtained with different feature sets length, ranging from 3000 features to the totality of the features available in the digital image. However, the investigation focused on extracting 1000 features since it was concluded that increasing the features did not imply an improvement in the results.

The results obtained with the model based on Benford's Law compete with those obtained from the models based on CNN and SVM, generating confidence regarding its application as decision support in a criminal investigation for the identification of manipulated images.

1. Introduction

Global digitalization has favoured the exponential growth of

* Corresponding author. School of Technology and Management, Polytechnic of Leiria, Morro do Lena, Alto do Vieiro, Leiria, 2411-911, Portugal.
E-mail addresses: pedro.a.fernandes@ipleiria.pt (P. Fernandes), mario.antunes@ipleiria.pt (M. Antunes).

https://doi.org/10.1016/j.fsidi.2023.301515
2666-2817/

illicit activities in cyberspace, such as phishing, spear-phishing and ransomware, constituting some of the main vectors of attack on countries' national security Enisa (2021).

Factors such as the pandemic and war in Europe have brought about a greater awareness of the risks resulting from the use of the Internet and the total dependence on digital services such as shopping, reading, and even chatting online. Moreover, the ease and speed of access to the Internet have provided unique opportunities for cybercriminals to commit illicit acts, Rajan et al. (2017). In this context, cyberattacks have been gaining strength, characterised mainly by sophistication and harmful impact on a society thirsty for information, creating psychological, economic and social problems, Ferreira et al. (2021).

According to *Europol*, cybercrime is a dynamic problem among the EU Member States, with cybercriminals taking advantage of the sufficiently robust Internet infrastructure, the negligence of most people in making online payments and the complete exposure of what they do in their daily lives, Europol.

It is precisely from this exposure that cybercriminals have found their way to perpetuate illicit activities, using powerful tools such as Photoshop and Gimp, allowing them to manipulate any multimedia digital content, namely digital photos, by using splice and copy-move techniques.

In the presence of manipulated content, the investigation carried out by digital forensic teams must include observations that integrate various aspects, such as the image's physical, digital and semantic integrity. Over time, various methods have been applied to expose certain inconsistencies in the image, such as shadows due to low light and low contrast. Moreover, manipulating an image leaves certain traces, visible or not, mainly detectable through computational tools using machine learning, Lin et al. (2018); Thakur and Rohilla (2020).

The impact of such manipulations is high and can have disastrous consequences. Following the rapid growth of computing power, the tools used in image processing have allowed the introduction of a set of new techniques based on artificial intelligence in the form of surface forgery, cheap forgery and deformation, which allow the manipulation of digital content in a fast, cheap and realistic way, LI, jian Zhou, guo Yuan, cui Guo and xin Niu (2014); Thakur and Rohilla (2020).

The motivation for crimes involving image manipulation is diverse, whether personal or political. Generally, revenge pornography or paedophilia involving people in a more vulnerable context, Harris (2019), and blackmail for ransom are the most prominent, leading to severe multi-level implications in people's lives.

The emergence of a set of legal procedures and standards, as well as the application of computer techniques and tools, has enabled digital forensic analysis, carried out by the criminal investigation police, to collect, preserve and analyse digital evidence consisting of a lengthy and complex process, Unodc (2019). On the other hand, the investigation must be quick and unambiguous as to which facts may or may not constitute a computer crime. The investigation carried out by an expert without using a set of tools to help him in such an arduous task when faced with millions of data is a time burden hardly feasible, Ferreira et al. (2021).

In recent years, a set of forensic tools has emerged, such as *Forensic Toolkit (FTK)*, *Autopsy* or *ImageNet*, with the ability to extract, analyse and reconstruct digital evidence, constituting a real help for researchers.

On the other hand, the development of several classifiers based on conventional machine learning models made it possible to automate the identification of a set of forensic artefacts based on statistical models, such as Bayesian algorithms (Naive Bayes, K-Nearest Neighbor), as well as the introduction of a set of performance metrics such as precision, recall and F-measure to benchmark the classification models. Recently, there has been substantial growth in techniques based on deep learning, with excellent results in many computer vision applications, Saini and Kapoor (2016); Ferreira et al. (2021); Mar-Raave et al. (2021).

The use of deep learning-based methods requires enormous computational power (where the need for data training is added when neural networks are used), with expensive GPU cards, making applications heavy on data processing; the current analysis of portability, which translates into the inability to migrate software components between hosts, and problems in data validation, driven by rapid technological evolution, makes existing models obsolete, in a process identical to the proverbial "there is no beauty without a catch". As a result of all this, these types of methodologies have not been implemented in digital forensic tools, Ferreira et al. (2021), increasing the enormous challenges faced by digital forensic investigators.

The problems arising from the application of unusual methods and the potential that statistical methods have presented in recent years, Kumar et al. (2021), are the motivational basis of this paper.

This paper describes the application of Benford's Law to detect digital images manipulated by splicing and copy-move. The operation of the proposed model is based on the extraction of a differentiated set of features from the digital images, calculated by the Fast Fourier Transform (FFT) method.

Apart from the results obtained rivalling those resulting from the application of machine learning-based techniques, the proposed model requires less CPU and memory processing as it does not require the use of data for training and does not require specific hardware to produce results. The possibility of creating lightweight modules that can be included in the most diverse digital forensic tools is an advantage worth exploring.

To evaluate the reliability of the results, statistical correlations have been used, such as Pearson's chi−square correlation coefficient, Spearman's correlation coefficient and Cramer-Von Misses (CVM) goodness of fit test, Singh and Bansal (2015).

The dataset consists of 9000 authentic images and 9000 images manipulated by splicing and copy-move, for a total of 18000 images. The overall results obtained with the statistical methods employed are an average precision of 86.14% and an F1 score of 90.40% in Pearson's distribution; an average precision of 50.79% and an F1 score of 56.55% in Spearman's distribution, and an average precision of 65.27% and an F1 score of 76.61% in the Cramer-Von Mises distribution methodology. The proposed model is not implemented as part of any digital forensic tools, unlike other machine learning-based solutions, Ferreira et al. (2021) increasing the enormous challenges that digital forensic investigators face.

The contributions of this paper can be described as follows.

- An architecture to preprocess and process the digital images to extract their features in the form of an array of numbers.
- An original model based on Benford's Law for processing digital images, centered on extracting the first digit.
- A set of MatLab scripts which implements Benford's law-based method and the datasets preprocessing and processing tasks. The scripts can be found as a GitHub project available in https://github.com/Pacfes/Benford-Law.
- A comparison of the results obtained to SVM and CNN machine learning and deep learning methods.

The present paper is organised as follows. Section 2 describes specific digital image manipulation techniques, some of the main algorithms that allow extracting features from images, and finally a set of statistical models that classify whether an image is authentic or manipulated. Section 3, mathematically describes Benford's law

and its application in different research areas related to the topic under review. Section 4 discusses the general architecture of the proposed model and the dataset where the experiments were processed. The section describes the pre-processing and processing steps in feature extraction from images. Next, the statistical correlation coefficient, namely Pearson and Spearman, and Cramer-Von Mises goodness of fit test, are discussed, as well as the techniques that allowed the analysis and evaluation of the results. This section further describes the metrics used, and the hypothesis tests carried out, which served as a basis for evaluating the images. Section 5 describes the results obtained from the performed experiments and their analysis. Finally, Section 6 presents the main conclusions obtained from the research and delineates possible future work activities.

2. Digital photos manipulation fundamentals

This section describes specific techniques for manipulating digital images, such as splicing and copy-move, some of the main algorithms that allow extracting characteristics from images and, finally, a set of statistical models that classify whether an image is authentic or manipulated.

2.1. Digital image manipulation techniques

The proliferation of low-cost electronic devices, such as smartphones, digital cameras, and tablets, has enabled the acquisition, distribution and sharing of digital images through several online platforms, among which TikTok and Instagram stand out.

The ease and speed with which this phenomenon occurs enable a range of cybercriminals to manipulate this digital content for various criminal purposes. However, any manipulation that may have happened to a digital image leaves a body of evidence that can be exploited by digital forensic tools Amerini et al. (2020).

Coupled with all these problems, the human inability to keep up with the increasing computational power associated with the physical limitations of their visual ability has allowed the creation of a new area of research named computer vision. The major goal of this new area of investigation is to try to overcome certain obstacles, specifically in the detection of manipulations that a digital image may have been subject to. Unfortunately, the human eye does not give absolute certainty that the image may or may not have suffered some manipulation.

Currently, there has been a vertiginous growth in the number of techniques and tools that allow the manipulation of an image, where copy-move, splicing and deepfake stand out as the most popular, but also in the automatic generation of manipulated images, as the case of "*this person does not exist*", using Generative Adversarial Networks (GAN). In the opposite direction, several technologies based on artificial intelligence have been applied, namely Convolutional Neural Networks (CNN), Support Vector Machines (SVN), watermarks, digital signatures and the use of statistical resources from the extraction of the characteristics of images in a dataset, allow their classification between authentic and manipulated, Muzaffer and Ulutas (2019).

Fig. 1. Left: Original image, **Right:** Manipulated image, Sreenivasu and Vani (2017).

However, the similarity caused between an authentic and a manipulated digital image can be a considerable challenge. Fig. 1 is an excellent example of this situation, where the manipulation resulted from applying the copy-move technique. The procedure is simple and consists of copying a specific component existing in the original image to another region of the resulting image, giving the sensation that there are more elements than those in the original image.

The copied regions are usually processed before being pasted to hide or remove any incriminating details. Certain areas of the image with texture such as grass, tree leaves or cloudy sky can be an asset to hide details of the pasted object and thus be invisible to the human eye.

Another technique widely used in manipulating a digital image is splicing, which employs AI techniques to copy and paste a specific region of input digital photos to a new resulting one (Fig. 2). When the manipulation results from splicing, several inconsistencies can be exploited to detect the manipulation, namely the presence of different illumination points in the various pasted regions, the existence of varying noise levels of the manipulated image, and other statistical properties present in the image, different from the original characteristics of the image when it was obtained from the digital camera.

2.2. Image feature extraction algorithms

Following the frenetic pace of crimes through the manipulation of digital content, researchers felt the need to develop a set of algorithms capable of extracting features from images to detect various manipulations, such as splicing, and copy-move, among others. These include algorithms based on the Discrete Cosine Transform (DCT), whose range of action focuses on compression, filtering and feature extraction from images.

The method proposed in this paper is described in Section 4. The feature extraction process divides the digital image into 8×8 non-overlapping blocks where DCT is applied to each of the three data channels (RGB). Such a procedure allows generating 64 DCT coefficients representing the various horizontal, vertical and composite frequencies, enabling the extraction of a set of features that are subsequently stored in a matrix of values. Recent investigations have shown that applying the DCT coefficients in Convolutional Neural Networks (CNN) is possible, obtaining good results with the neural networks being trained with the DCT coefficients compressed into JPEG, Rajesh et al. (2019). The DCT coefficients can be calculated from equation (1):

$$X[k] = \alpha[k] \sum_{n=0}^{N-1} x[n] \cos\left(\frac{\pi(2n+1)k}{2N}\right) \tag{1}$$

where $\alpha[k] = \sqrt{\frac{1}{N}}$ if $k = 0$ and $\sqrt{\frac{2}{N}}$ if $k = 1, 2, \ldots, N-1$.

Related to the DCT algorithm is the Discrete Fourier Transform (DFT). The DFT can facilitate harmonic analysis and signal processing in the frequency domain. Algorithms were developed to calculate its coefficients fast on the Fast Fourier Transform (FFT) approach of Cooley and Tukey by the composition of simple elementary transforms, usually known as *butterfly transforms*. The DFT maps the sequence in the time domain to a sequence in the frequency domain of the same length, allowing information about the amplitude and phase of the signal at each frequency and is calculated by Equation (2), Ferreira et al. (2021); Rao et al. (2010); Parfieniuk (2021); Cooley and Tukey (1965):

Fig. 2. Left: Original image, **Middle** Original image **Right:** Manipulated image, Wu et al. (2022).

$$X(k,l) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} x(n,m) e^{\left(-\frac{2\pi i}{N} k_n\right)} e^{\left(\frac{2\pi i}{M} l_m\right)} \tag{2}$$

Fast Fourier Transform is a DFT algorithm that allows reduce the number of complex operations from $O\left(N^2\right)$ to $O(Nlog_2 N)$.

The Fast Fourier Transform (FFT) allows finding the relationship between the indices of a vector $N = LC$ with the elements of a matrix of dimension $L \times C$, given by Equations (3) and (4):

$$n = n_1 + Ln_2 \tag{3}$$

$$u = u_1 + Cn_2 \tag{4}$$

where n_1, n_2, u_1, u_2 are the new indices of the transform.

The two-dimensional transform results from replacing the frequency and time indices given by 3 and 4, in the expression 5:

$$y_u = \frac{1}{N} \sum_{n=0}^{N-1} x_n W^{nu}, \text{ where } W = e^{\frac{-2\pi i}{N}} \tag{5}$$

where y_u represents the Fourier coefficient, resulting in equation (6), Pedrini and Schwartz (2008):

$$\frac{1}{N} \sum_{n_1=0}^{L-1} \sum_{n_2=0}^{C-1} X_{n_1} + L_{n_2} W^{n_1 u_1} W^{Cn_1 u_2} W^{Ln_2 u_1} \tag{6}$$

Fig. 3 represents the extraction of 40 features taken from the image. The vector produced is represented by the values resulting from the application of the FFT, which contains information about certain particularities existing in the input image, such as variations in the grey tones of the pixel's existence in areas with higher brightness intensity.

2.3. Statistical correlation coefficients

The introduction of hypothesis testing in the detection of manipulated images required using a set of statistical models, namely Pearson, Spearman, and Cramer-Von Mises, to calculate correlation coefficients and p-values based on the extraction of the first digits from the image's characteristics.

Such procedure allowed the construction of a decision rule to reject or not reject the statistical hypothesis that allowed classifying an image as authentic or manipulated, as described in section 5.

Pearson's correlation coefficients make it possible to describe whether or not there is a linear relationship between two quantitative variables, and if so, their correlation is calculated using the Equation described in Equation (7). It is commonly used in

(a) Authentic image

```
[1.                10201928.51602545   290088.9734832387   108522.14486396471   65993.07433980935
40637.57075600533  26876.532773295134  21745.230858866176  16936.23276289725    13666.158640986421
12102.807954481035 10063.993746706052  8854.909602119318   7988.83573556636     7096.777876799761
6192.121132291657  5736.0287205105815  5105.338371644688   4545.805398767385    4189.338649068494
3766.6801048116126 3464.6827682974076  3119.6321363929787  2722.4600320080226   2424.4727358661694
2222.8368172389846 1992.338626429163   1945.6995507819893  1826.9644500398388   1690.2877759648873
1521.1837025166378 1432.0265027198277  1149.766662624205   1083.178176462072    978.8052664684922
865.20032967722    772.1488081359927   763.3861198843238   728.1553505148144    730.5946598270873
1012.5727300295159]
```

(b) Extracted features

Fig. 3. Process of extracting features from an image.

inferential statistics to test statistical hypotheses.

$$r = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}(X - \bar{x})^2 \sum_{j=1}^{n}\left(Y - \bar{Y}\right)^2}} \tag{7}$$

where r represents Pearson's correlation coefficient.

To verify if the correlation between two quantitative variables is significant, it is necessary to calculate the p-value, given by Equation (8), known as the level of significance associated with a particular observed value of the test statistic. Briefly, the value of the p-test allows checking whether a given value is more unfavourable for the null hypothesis, assuming that the hypothesis is true.

$$t = \frac{r \times \sqrt{n-2}}{\sqrt{1-r^2}} \tag{8}$$

where r is the correlation coefficient, t is the p-value coefficient and n the number of observations.

When performing specific tests, certain problems may arise in the data set, such as small outliers and data that do not follow normality. Therefore, it is necessary to perform specific procedures to certify the validation of Pearson's model, such as studying the homoscedasticity and normality of the data set, Levine et al. (2021). Spearman's correlation coefficient can be used in cases where there are continuous variables, and the relationships are not linear but monotonic, Thirumalai et al. (2017).

Equation (9) calculates the correlation between variables:

$$r(X,Y) = 1 - \frac{6\sum_{i=1}^{n}(x_i - y_i)^2}{n^3 - n} \tag{9}$$

where n is the length of each column, Best and Roberts (1975). Equation 8 allows calculating the p-value of the Spearman correlation.

Calculating the Cramer-Von Mises correlation will verify the goodness of fit between two functions, where one of them is represented by the cumulative distribution of the values that constitute the images and another function that represents the empirical distribution of the values based on Benford's law.

The Cramer-Von Mises criterion is defined by Equation (10):

$$W^2 = \int (F^*(t) - F_0(t))^2 dF_0(t) \tag{10}$$

where $F^*(t) = \frac{k}{N}$ with k observations being less than or equal to $t = 0, 1, \ldots, N$ is the empirical cumulative distributions function, and $F_0(t)$ is theoretical cumulative distributions function, Anderson (1962), and the Equation that allows checking the p-test between the distributions is given by Equation (11):

$$P_{value} = \frac{[T - \mu T]}{\sqrt{45 \times Var(T)}} + \frac{1}{6} \tag{11}$$

where T is the expected value.

3. Benford's law fundamentals

This section describes Benford's law mathematical fundamentals, as well as its main contributions in the area of digital, image and video manipulation.

3.1. The math behind Benford's law

Suppose we are in the presence of an independent and identically distributed random variable, $X = (X_1, X_2, \ldots, X_i)$, $i = 1, 2, \ldots, n$, $\forall n \in \mathbf{N}$, and $D_i(X)$ represents the ith significant decimal digit of X.

The probability mass function that best describes Benford's law is given by Equation (12), Arno Berger (2015):

$$P(D_i(X)) = \log\left(1 + \frac{1}{d}\right), \text{ if } d = \{1, 2, 3, \ldots, 9\} \tag{12}$$

From Equation (12), we can calculate the empirical frequency of each digit. For example, the probability of the number 1 is given by $\log(1 + \frac{1}{1}) = \log(2) \cong 0.301$. If $d = 2$, then $\log(1 + \frac{1}{2}) = \log(\frac{3}{2}) \cong 0.176$ and so forth, until you get the probability of $d = 9$ and, producing the graph defined in Fig. 4.

However, Benford's law is not restricted to the first digit, having immediate implications for the construction of the second digit, third digit and so on. The probability for the second digit is given by Equation (13).

$$\sum_{k=1}^{9} \log(1 + (10k + d)^{-1}), d = \{1, 2, 3, \ldots, 9\} \tag{13}$$

The math expressions defined in 12 and 13, allow us to introduce the general Theorem 3.1, which allows us to calculate the empirical frequency of each digit.

Theorem 3.1. (*General law*). *Be* $k \in \mathbf{Z}$, $d_1 \in \{1, 2, 3, \ldots, 9\}$ *and* $d_j \in \{0, 1, 2, \ldots, 9\}$, $j = 2, \ldots, k$.

$$P(D_k = d_k) = \log\left(1 + \frac{1}{\sum_{i=1}^{k} d_i \times 10^{k-i}}\right) \tag{14}$$

Benford's general law has the ability to be generalized to all significant digits. Usually, the first significant decimal digit of a real number is defined as x as the first non-zero value in the expanding development of x, the second significant decimal digit as the second digit after the first significant digit, and so on, Arno Berger (2015).

The Definition 3.1 allows to explain the real meaning about what the significant digits mean, Arno Berger (2015).

Definition 3.1. For each real value x, different from zero, it is considered as the first significant digit of x, expressed by $D_1(x)$, the

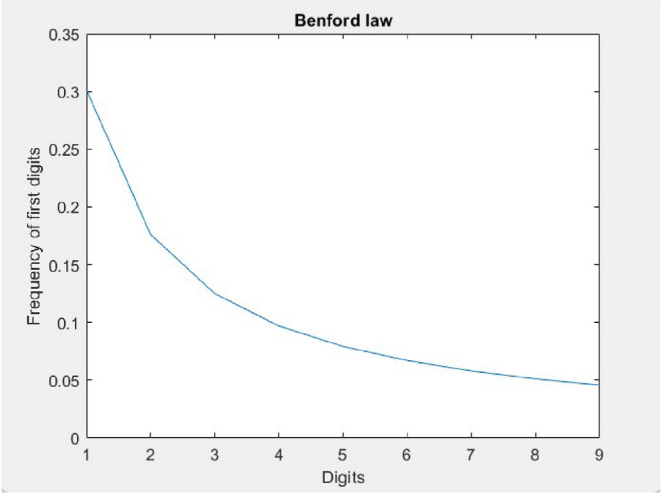

Fig. 4. Benford's law graph for the first digit.

only integer $j \in \{1, 2, 3, \ldots, 9\}$ satisfying: $10^k j \leq |x| < 10^k(j+1)$, $\forall k \in Z$.

Likewise, for values greater than 1, e.g., $n \geq 2$, $n \in N$, the nth significant digit of x, expressed by $D_n(x)$, can be defined inductively as the single integer $j \in \{1, 2, 3, \ldots, 9\}$, such that:

$$10^k\left(\sum_{i=1}^{n-1} D_i(x)10^{n-i} + j\right) \leq |x| < 10^k\left(\sum_{i=1}^{n-1} D_i(x)10^{n-i} + j + 1\right),$$

$k \in Z$.

By convention, $D_n(0) := 0$, $\forall n \in N$.

In addition to the definition of what significant digits are, discussed in the definition 3.1, there is a need to define what significant a real number is. The signifier of a real number refers to the coefficient of that number in the floating-point form. The definition 3.2, explain this type of numbers, Arno Berger (2015); Taimori et al. (2012).

Definition 3.2. Given a function $F : IR \rightarrow [1, 10)$.

If $x \neq 0$ then:

$S(x) := 10^{\log |x| - \lfloor \log |x| \rfloor}$, $\forall x \neq 0$.

By convention $S(0) := 0$.

According to Arno Berger book, the relationship between the significand of a number and the significant digits is visible. This relationship is translated into the property 3.1, where the significant digits can be expressed as the coefficients of a significand function.

Property 3.1. *x represents a real number.*

$D_n(x) := \lfloor 10^{n-1} S(x) \rfloor - 10\lfloor 10^{n-2} S(x) \rfloor$, $n \in N$.

Another important concept about digits and Benford's law is the definition of mantissa. The Definition 3.3, allows to establish a relationship between the significand numbers, present in definition 3.2, and the traditional definition of mantissa, Arno Berger (2015); Parnak et al. (2020).

Definition 3.3. (Mantissa)**.** Represents the decimal part in calculating the log of a number. The relationship between the *significands* translates into $\log S(x)$. The only number r in $[\frac{1}{10}, 1)$ with $x = r \times 10^n$ for some integer.

The property 3.2, reinforces the idea of a relationship between the significant of a real number.

Property 3.2. *The mantissa of a number does not change when we multiply the logarithm by a power of 10.*

Another important aspect, highlighted in the investigation carried out by Berger and Hill (2011); Arno Berger (2015), is related to the existence of a sequence of integer and positive digits. The definition 3.4, shows whether a given sequence conforms to Benford's law.

Definition 3.4. Given a sequence of positive integers $(x_1, x_2, \ldots)$ denoted by (x_n), if $t \in [1, 2, \ldots, 10)$ the limiting proportion of meaningful (x_n) and less than or equal to t is exactly $\log t$.

$\lim_{N \to \infty} \frac{\{1 \leq n \leq N : S(x_n) \leq t\}}{N} = \log t, \forall t \in [1, 10)$

Given the definition 3.4, checking whether or not a given sequence is Benford's is not a simple process. It is essential to introduce a set of properties, that allow a more straightforward solution.

Benford's law can be defined by four important properties defined in Property 3.3, Taimori et al. (2012); Volčič (2020); Hill (1995).

Property 3.3. *The main defining properties of Benford's law are:*

- *Uniform distributions, where $D_i(X) \sim B_i \Leftrightarrow Y \sim U(0, 1)$;*
- *Scale-invariance;*
- *Base-invariance;*
- *sum-invariance characterizations.*

Based on the Properties 3.3, we can introduce lemma 3.1, based on Weyl's theorem, Arno Berger (2015). The proof can be seen at Berger and Hill (2011).

Lemma 3.1. *The sequence $(n \times a) = (a, 2a, 3a, \ldots)$ is uniformly distributed mod 1 if and only if a is irrational.*

From the lemma indicated in 3.1, results Theorem 3.2, which allows defining a sequence, Arno Berger (2015); Berger and Hill (2011).

Theorem 3.2. *A given sequence is a Benford sequence if* $(\log |x_n|) = (\log |x_1|, \log |x_2|, \log |x_3|, \ldots)$ *is uniformly distributed mod*1.

From the mathematical concepts highlighted in this section, we can extract some important ideas.For a better understanding of these concepts, in section 3.3 an example was made based on two images, one manipulated and the other authentic, where the mathematical procedures relating to this section are described.

- In the first stage, extracting the first significant digit of the set of features extracted from the digital images was possible.
- To state that a given set of features obtained from the digital images follows Benford's law, then the behaviour of the first digits of these features should be identical to the behaviour shown in the graph of Fig. 4.
- After analysing the behaviour of the digits resulting from the extraction of the characteristics of the images, if they do not behave according to the graph of Fig. 4, it means that the image may have been manipulated, resulting in a true positive.

3.2. Benford's law related works applied to digital forensics

Benford's law has been widely used in the most diverse areas of investigation, namely the detection of financial fraud, anomalies in electoral data and in scientific fraud. All research work carried out in these areas has shown promising results, Said and Mohammed (2020); Nunes et al. (2019); Taimori et al. (2012).

The application of Benford's law in digital image processing is only a few years old. Jolion Wolf et al. (2000), Acebo del Acebo and Sbert (2005), and Sbert Bardera et al. (2006), have demonstrated its use in certain fields whose domain ranges from the magnitude of the gradient of a given image in the entropy field, as well as the intensity of the light in natural images. Based on that, the focus of a small part of researchers has been concentrated on the manipulation and adulteration of digital images. Recent studies point to investigations into the first digits of a Discrete Cosine Transform (DCT) block, characterised by a lossless compression process that allows reversion based on simple matrix multiplications, whose capacity allows the conversion of pixel values into coefficients Satapathy et al. (2020).

The transformation is performed in 8X8 bit blocks and considers the type of image, whether it is coloured or not. In addition to the type of image, there are other characteristics to consider, and namely, if the image is coloured, the blocks are applied to the chrominance (colour value); if the image is grey scaled, the blocks are applied to the luminance and the number of JPEG coefficients.

There is a record of several trials in applying Benford's law in digital forensic analysis, including topics such as compression of JPEG images to bitmap format, estimation of compression of JPEG images, and double compression. However, the growing interest in

the subject has led the investigation to new studies related to detecting tampered images, whether by clone, retouching, or splicing, among other techniques, Mire (2022); Parnak et al. (2022). It has also been applied to recover hidden data in digital images or even in the registration of the image itself, Singh and Bansal (2015); Wu et al. (2022).

The research has also focused its efforts on statistical models based on filters, restoration, and image analysis. Studies have shown that Benford's law-based method detects manipulations associated with double compression of JPEG images, Lesperance et al. (2016); Pasquini et al. (2017); Yao et al. (2020).

Other investigations point to JPEG2000 images (with higher quality than the JPEG format) and Discrete Wavelet Transform (DWT) coefficients, that allows decomposing an image in a single resolution level structured into four sub-bands, low−low, low-hight, hight-low, hight-high), Wang et al. (2015); Qadir et al. (2011).

In the JPEG2000 format, images have a very high quality compared to other formats. Studies based on this format have revealed that increasing the rate of compression of the images ends up increasing in equal proportion the deviation of the coefficients, inducing possible manipulations in the images, Yang et al. (2015); Singh and Bansal (2015).

Concerning the application of DWT, it was verified that the images that suffered double compression did not follow Benford's law since accentuated changes in the logarithmic curve were detected Singh and Bansal (2015). Other studies analysed the introduction of brightness in the images, obtaining a clear distortion in the curve. Still, in the field of brightness, other studies try to detect the presence of unbalanced lighting in images with the help of the Discrete Wavelet Transform (DWT), identifying certain irregularities in the intensity of brightness, with the accentuated presence of brightness in certain areas to the detriment of others and possibly losing some visual quality,Wei et al. (2021); Singh and Bansal (2015).

One of the limitations of Benford'law is characterised by the existence of malicious attacks based on the knowledge that attackers may have in the forensic context. Image manipulation and subsequent compensation are serious obstacles to the detection of manipulated images, all because, after compensation by compression, the Benford curve resembles the original curve Wang et al. (2009).

Recent studies report the application of Benford's law in the separation of images generated by computer graphics from photographic images, Meena and Tyagi (2019), and in the detection of unknown JPEG compression in semi-fragile watermarked images, obtaining good results, Zhao et al. (2009). The first digits extraction process is based on the DCT process.

3.3. Proof of concept using Benford's law

The dataset consists of two images, one manipulated and one authentic, and served only to test the model, available at *"This person does not exist"* website, Karras and Nvidia (2019), Flickr-Faces-HQ, Karras et al. (2019);

Table 1 details the datasets collected and used in the experiments (see Table 2).

Table 1
Dataset to prove the concept.

Name	Fake	Real
"This person does not exist"	1	–
Flickr-Faces-HQ	–	1
Total	1	1

Table 2
Model Comparison with 50, 150, 300 features.

Model	Correlation		P-Value	
50 features	Image 1	Image 2	Image 1	Image 2
Pearson	0.8635	0.7905	0.0027	0.0112
Spearman	0.9279	0.5108	0.0008	0.1620
Cramer-Von Mises	0.0200	0.0457	0.0057	0.0322
150 features				
Pearson	0.8513	0.8573	0.0036	0.0031
Spearman	0.9667	0.4667	0.0002	0.2125
Cramer-Von Mises	0.0262	0.0428	0.0121	0.0293
300 features				
Pearson	0.8242	0.8929	0.0063	0.0012
Spearman	0.9833	0.5833	0.0000	0.1080
Cramer-Von Mises	0.0224	0.0305	0.0081	0.0165

Initially, 150 and 800 features were extracted from the two images. Then, the total number of the first digits was counted, regardless of whether the image was genuine or manipulated. As a result, Fig. 5 depicts a graphic that shows a considerable distortion of each image compared with Benford's law, indicating the possibility of manipulations.

Analysing the images individually, as depicted in Fig. 6, we can observe the distortion caused by the two images by comparison with the Benford line. The objective is to verify whether the statistical models can detect which image is manipulated and which is authentic. Thus, several parameters were tested to find the optimal point, such as the number of features obtained from each image, and different significance degrees (0.01, 0.05 and 0.001), and the metric used for this scenario. In the other investigations presented in this article, the statistical procedure adopted followed the same line of thought.

The results of the correlations and the P-value obtained from the relative frequencies resulting from the application of the Pearson, Spearman and Cramer-Von Mises models with the empirical frequency of Benford's law, taking into account 50, 150 and 300 features.

Hypothesis testing.

- H(0): The correlation coefficient equals zero; there is no linear relationship between the pair of variables under analysis; possible manipulated image.
- H(a): The correlation coefficient is not equal to zero; there is a linear relationship between the pair of variables under analysis; a possible authentic image.

Rule that allows the decision.

- Do not reject H(0) if the degree of significance $> \alpha$.
- Reject H(0) if the degree of significance $\leq \alpha$.

Tables 3 and 4, shows the results of the tests considering the different degrees of significance 0.01, 0.05 (agreed upon) and 0.001.

4. Proposed architecture

This section describes the architecture implemented to process the digital images based on Benford's law, allowing images to be classified as authentic or manipulated.

4.1. Benford's law-based method

The proposed model is based on two phases. In a first phase, the analysis of the first digit extracted from the characteristics of the

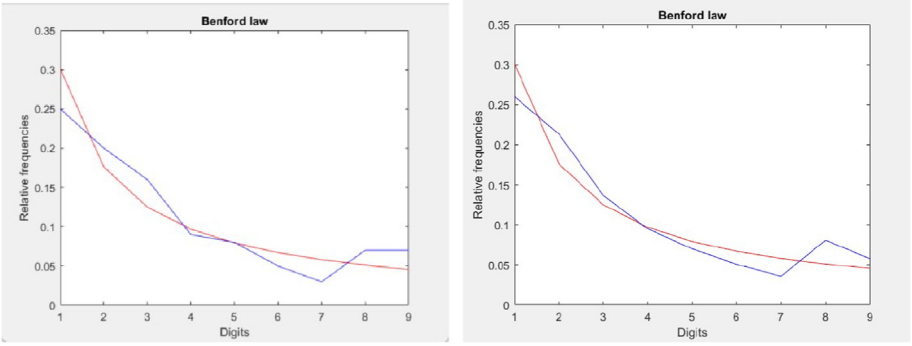

Fig. 5. Comparison of the first digits extracted from 150 (left) and 800 (right) of two-image features and Benford's law. It turns out that there is no difference between the graphics, concluding that increasing the extraction of the number of features does not affect the graph produced.

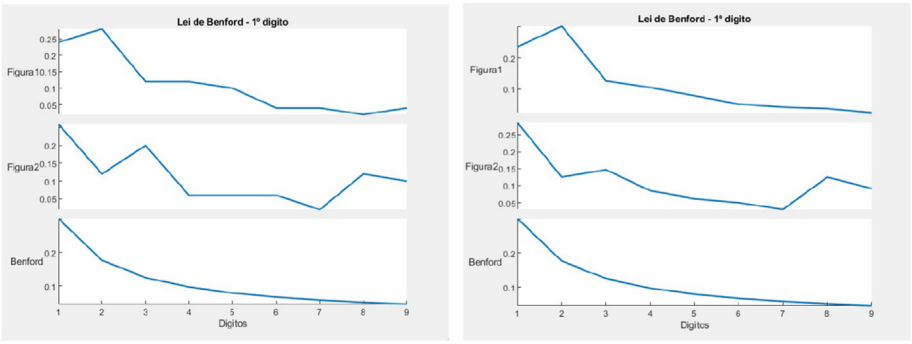

Fig. 6. Comparison between the first digits extracted from the features of each image and Benford's law, using 150 (left) and 800 (right) features.

Table 3
Hypothesis testing, where (R) rejects H(0) and (DR) does not rejects H(0).

Model	Image 1 (Manipulated)		
50 features	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.001$
Pearson	R	R	DR
Spearman	R	R	R
Cramer-Von Mises	R	R	DR
150 features			
Pearson	R	R	DR
Spearman	R	R	R
Cramer-Von Mises	DR	R	DR
300 features			
Pearson	R	R	DR
Spearman	R	R	R
Cramer-Von Mises	R	R	DR

Table 4
Hypothesis testing, where (R) rejects H(0) and (DR) does not rejects H(0).

Model	Image 2 (Authentic)		
50 features	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.001$
Pearson	DR	R	DR
Spearman	DR	DR	DR
Cramer-Von Mises	DR	R	DR
150 features			
Pearson	R	R	DR
Spearman	DR	DR	DR
Cramer-Von Mises	DR	R	DR
300 features			
Pearson	R	R	DR
Spearman	DR	DR	DR
Cramer-Von Mises	DR	R	DR

digital images is carried out, and in a second phase, the analysis is based on the second digit in the expectation of verifying if the results improve, worsen or remain the same, following a line of research aimed at answering the question: *"If we are faced with a database containing digital images, is it possible to detect whether there are authentic or manipulated images and which ones?"*. According to Benford's law, if there is a manipulation in the first digit, the graph will produce a different curve from the curve produced by Benford's law, Fig. 4 presented in section 3.1.

Fig. 7 illustrates the overall architecture designed to apply Benford's law under the context of manipulated digital image detection. It is based on the following three main building blocks: preprocessing, processing and analysis of the results.

In order to obtain a functional model for the detection of manipulated images, it is necessary to implement a set of procedures to obtain the input data and extract the first digit from it. Such a procedure will be vital to classify the images as manipulated or authentic.

The pre-processing depicted in Fig. 8 consists in extracting a set of n features from the images by applying the DFT (Discrete Fourier Transform) method. For this, a Python script was built, where besides the standard libraries (NumPy, pickle), the libraries OpenCV were used to process the image. A script was built for the radial profiling function, whose main function is to create a circular boundary in the image, extracting only the features within the circular zone. The extracted data is stored in a dataset, where through the development of a script built in MatLab, the first digit of all the obtained values was extracted and subsequently stored in a digit matrix.

The data relating to the extraction of the first digit from each image is appropriately stored into a feature vector, and each is labelled, that is, if the image is original, it is assigned the label 1;

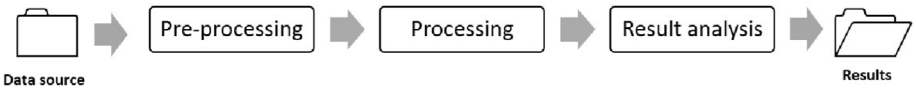

Fig. 7. General architecture of the method based on Benford's law.

Fig. 8. Preprocessing phase.

otherwise, the image is manipulated and is assigned the labelled with 0.

At the end of pre-processing, an adequately labelled dataset is available to apply a set of hypothesis tests based on Pearson, Spearman and Cramer-Von Mises statistical models.

The processing phase depicted in Fig. 9 consists of two steps. The first step consists in counting the first digits from the values obtained in the pre-processing phase for each image. In contrast, the second step calculates the absolute frequency of each digit, having the whole database as a reference. Then, the relative frequency of the values obtained in the two previous steps is calculated, consisting of the quotient between the absolute frequency of each digit and the sum of the total number of digits of each image under study, allowing the subsequent comparison with Benford's law. Finally, the values obtained by the relative frequency calculation are duly stored in a data set for further investigation, determined by two significant moments: hypothesis tests and graphically.

Fig. 10 schematically shows the processing performed by the hypothesis tests, and it is vital for the ongoing investigation. Three hypothesis tests were introduced from the relative frequencies based on three different models: Pearson, Spearman, and Cramer-Von Misses. Each model allowed the generation of labels related to the evaluation, indicating 1 if the image is genuine or 0 if the image was manipulated. These labels are stored according to the statistical model used, whether Pearson, Spearman or Cramer-Von Misses, and compared with the labels obtained in the pre-processing of the images, generating a set of results where the quantity of manipulated and authentic images is analysed.

Table 5 depicts the final dataset that allows comparing the labels to check which ones are authentic or manipulated.

4.2. Datasets

Two different scenarios were created with the dataset presented in Table 6 as the basis for carrying out the experimental tests.

The first scenario was derived from the main dataset, composed of two public datasets containing authentic and manipulated images, including various types of manipulation, from splicing to copy movement. The second scenario used the dataset in its entirety, which, like the first, contains authentic and manipulated images, including various types of manipulation, such as splicing and copy movement.

The dataset referring to the first scenario contains 280 manipulated and 280 authentic images for a total of 560 images. It consists of a compilation of images available on the COVERAGE dataset website, Wen et al. (2016) and Columbia Image Splicing Dataset,Ng, Hsu and Chang (2004). The dataset relating to the second scenario contains 9000 manipulated images and 9000 authentic images for a total of 18000 images and consists of a compilation of images available in various datasets, available on GitHub, Ferreira (2021).

Table 6 details the datasets collected and used in the experiments.

The databases are balanced, although it is not necessary to have data for training (as in machine learning models) where the amount of authentic and manipulated images may not necessarily be the same.

The experimental setup comprises various technological components and tools described in Table 7.

4.3. Results analysis

This section describes the metrics used, as well as the hypothesis tests carried out, which served as a basis for evaluating the images.

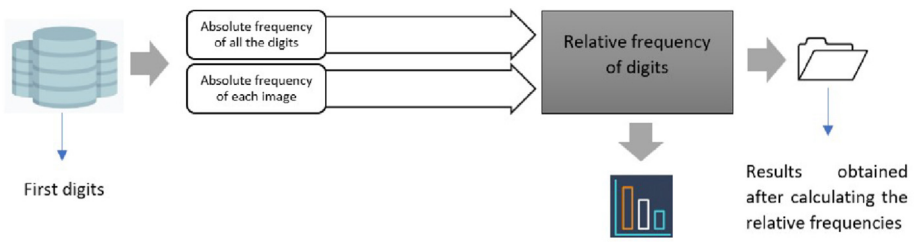

Fig. 9. Processing stage.

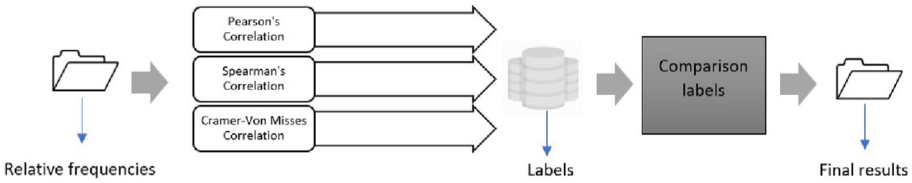

Fig. 10. Processing performed by the hypothesis tests.

Table 5
Dataset examples labelling.

ID Image	Original status	Pearson status
0	0	1
1	0	1
2	0	1
3	0	1
…	…	…

Table 6
Dataset to evaluate the proposed model, Ferreira et al. (2021).

Name	Fake	Real
Columbia Image Splicing Dataset	180	180
COVERAGE dataset	100	100
CelebA-HQ dataset	–	8600
This person does not exist	120	–
100K-Faces-HQ Datset	8600	–
Flickr-Faces-HQ Dataset	–	120
Total	9000	9000

Table 7
Technological components and tools.

Name	Characteristic
Operating system	Windows 11 Home 21H2
Processor	i7-10875H CPU @ 2.30 GHz 2.30 GHz
Graphics card	NVIDIA GeForce RTX 2070
Virtual machine	VMware Workstation 16 Player
Ram Installed	4 GB
Operating system	Ubuntu 64 bits
Processors	2
Python	Python 3.9.7
R Version	4.1.3
Plataform	x86-64-w64-mingw32/x64 (64-bit)
Distribution	GNU Public License
Matlab	Student - Individual

4.3.1. Hypothesis tests

The introduction of hypothesis testing in the context of manipulated image detection allows us to Equation te a decision rule to reject or not the statistical hypothesis based on our observations. The non-rejection of the statistical inference results from insufficient evidence to reject it, not implying that it can be true, hence the impossibility of claiming acceptance of the hypothesis. The complete set of features extracted from the images in the database was considered in the problem domain. Based on the information gathered, whether the hypothesis is true, i.e., the images are manipulated, or the hypothesis is false, i.e., the images are authentic.

The following notation is used.

- H_0: null hypothesis, or the statistical hypothesis to be tested;
- H_1: alternative hypothesis generally represents the conjecture to be proved.

The p-value is calculated, reflecting the probability of observing a more unfavourable sample for the null hypothesis. In this sense, when the calculated P-value is too small for a given value, the probability of being a more unfavourable sample than the observed one is small, and the initial hypothesis is rejected.

For the rejection or non-rejection of the initial hypothesis, the significance level, called α, represents the probabilities that lie outside the confidence intervals of a given distribution. The values considered to be accepted and well documented by the literature are 1%, 5% or 10%, Johnson (2013); Krzywinski and Altman (2013).

4.3.2. Evaluation metrics

Several metrics have been implemented for a correct evaluation of the models implemented in this research, namely Precision (P), Recall (R), F1-score (F1) and Accuracy (A), which were computed based on the confusion matrix, consisting of 2 rows to accommodate the prediction classes, and two columns for the instances of the real classes, Caelen (2017); Tharwat (2020), depicted in Table 8.

The positive classes refer to the manipulated images, and the negative classes refer to the authentic images. The True Positives (TP) refer to the events where the model correctly predicted the existence of manipulated images. In contrast, the True Negatives (TN) represent the events where the model predicted that the images were authentic. Both False Positives (FP) and False Negatives (FN) refer to events incorrectly predicted by the model, classifying authentic images as fake or fake images as genuine, Ferreira et al. (2021).

1. **Precision**, given by equation (15), allows calculating the percentage of images classified as true, referring to authentic images that correspond to authentic images.

$$P = \frac{TP}{TP + FP} \tag{15}$$

2. **Recall**, given by equation (16), which allows calculating the percentage of images classified as manipulated in the total number of manipulated images present in the database.

$$R = \frac{TP}{TP + FN} \tag{16}$$

Table 8
Confusion matrix.

	Positive	Negative
Positive	TP	FP
Negative	FN	TN

3. **F1-score**, given by equation (17), allows to make measurements between the accuracy and robustness of the classifier. It is often nicknamed harmonic mean, because it works with inversely proportional magnitudes and lies in the [0.1] range.

$$F1 = 2*\frac{P*R}{P+R} \tag{17}$$

4. **Accuracy**, given by equation (18), allows obtaining a percentage determined by the quotient between the number of images classified as genuine by the total number of images calculated as genuine and manipulated.

$$A = \frac{TP+TN}{TP+TN+FP+FN} \tag{18}$$

5. Results of experience-based research

This section describes and discusses the experiments made with the Bendford's model. After applying the model to the set of images, the correlation and p-value values were generated, enabling the generation of the confusion matrix. Specific metrics such as accuracy, recall, F1-score, and processing time are also analysed.

5.1. Scenario 1 - dataset with 560 examples

The experiment analysed 280 authentic and 280 fake images for a total of 560 images. In this experiment it was extracted 200 features, then 500 and finally 1000 features, from each digital image. Similar to what was done in Section 3.3, the aim is to check that the dataset complies with Benford's law graphically.

Observing the graphics in Fig. 11, it is visible that there is no change in the plotting between the curve produced by Benford's law and the curve produced by the frequency of occurrence of the digits, inferring that they are not affected by the number of features extracted from the images. We can also see that the curves are approximate, concluding that the dataset follows Benford's law. This fact is all the more critical as other authors in line with research carry it out, Wang et al. (2009); Singh and Bansal (2015).

However, the investigation is centred on a dataset composed of authentic and manipulated images, which allows us to obtain prior knowledge about the dataset. Graphically, the dataset complies with Benford's law giving a false idea that the data present in it has not been manipulated in any way. Therefore, from this moment on, the investigation will be conducted by an image-by-image analysis (Fig. 12), applying the models previously described in Section 3.3.

Tables 9–11 contain the results obtained after extracting 200, 500 and 1000 features from the authentic and manipulated images of the dataset, as well as the time processing for each model.

Comparing the average values obtained in Tables 9, 10 and 11, it can be seen that the Pearson model produces the best accuracy, 61.67%, relative to the other models evaluated. The number of misclassified images covering false positives and false negatives is high. Compared to the Spearman and Cramer-Von Mises models, the number of misclassified samples in the Spearman and CVM models is low for false negatives but high for false positives; the number of samples classified as manipulated when they were authentic is relatively high. In this comparison, the model advocated by Pearson proves to be more homogeneous, which can be explained by the direct relationship between the variables under analysis. As for the average accuracy, Pearson's model has the highest value, as it can detect a more significant amount of manipulated images compared to other models. Regarding recall and F1-score metrics, all models are in the same line of action but with better results for CVM. Compared to conventional machine learning-based Ferreira et al. (2021) models, the current model falls far short of the expected results. For example, the F1 score obtained by Support Vector Machines reaches 99.8%, considerably higher than the best result obtained with the same dataset. The research was based on statistical models with no training data, limiting the analysis only to the data extracted from each image.

Benford's law and the correlation methods explored are not based on learning of previously trained data, as in machine learning-based methods. A point in favour, and duly depicted in Tables 9–11, concerns the execution time. Comparing the processing times based on conventional machine learning models, where the detection of manipulated images reached very high running times (in some cases more than 6 h, Ferreira et al. (2021)), the maximum obtained by the proposed models was 54 s when extracting 40000 features. The results detailed in Tables 9–11, were obtained at three key times based on the number of features extracted from the dataset, namely 200, 500 and 1000.

For verification purposes, 3000 and 40, 000 features were also extracted. The tests confirmed the trend in the results, where one can conclude that the higher the number of features does not imply in a better performance of the model.

A possible justification for these results, being worse than those obtained by learning processes, is related to the number of images being processed and how they were obtained. The extraction of the characteristics of an image is performed from its pixels, and a retouching process constitutes a manipulation. It may affect the quantity and quality of the pixels. Therefore, it is essential that the data can be obtained from the source devices (cameras, sensors) and, in the case of forensic expertise, the characteristics of the images can be safely extracted in laboratories built for this purpose,

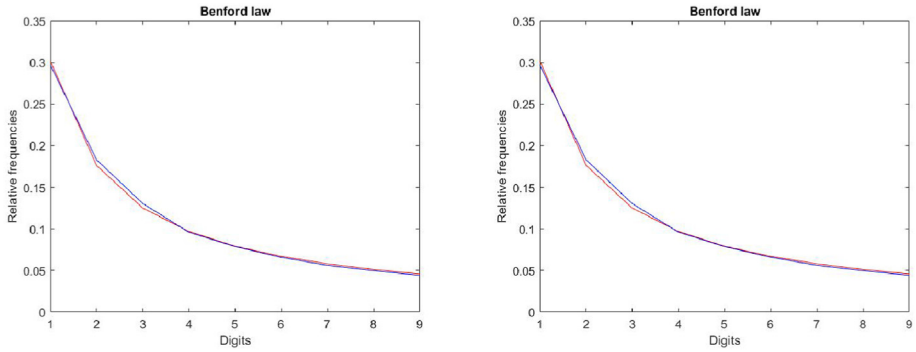

Fig. 11. Comparison between the first digits extracted from the features of two images and Benford's law using 200 features (left) and 500 features (right).

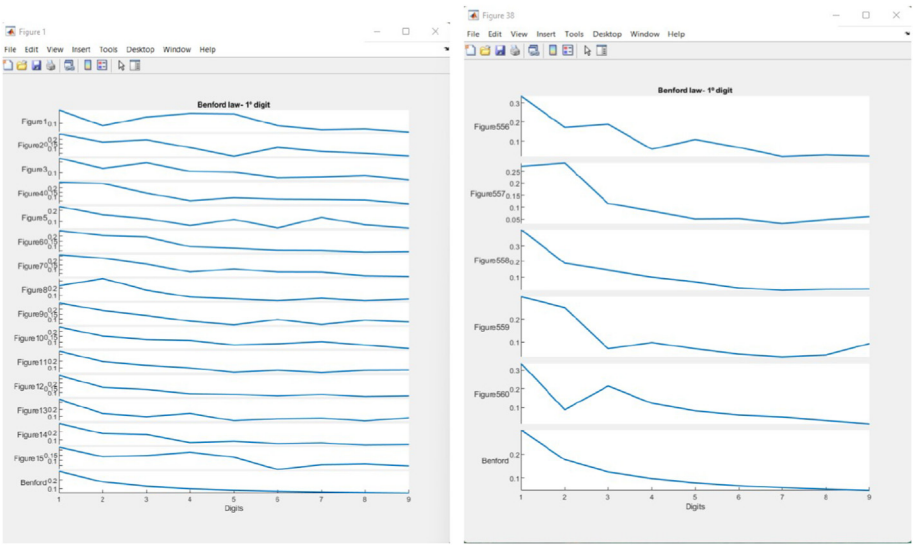

Fig. 12. Comparison of the curve constituted by the characteristics of each image with the curve of Benford's law.

Table 9
Results obtained after extracting 200, 500 and 1000 features from the images dataset, using Pearson $\alpha = 0.001$.

	TP	TN	FP	FN	PR	RE	F1	AC	Time
200	167	179	101	113	0.6231	0.5964	0.6095	0.6179	0.509 s
500	168	177	103	112	0.6199	0.6000	0.6098	0.6161	0.564 s
1000	166	179	101	114	0.6217	0.5929	0.6069	0.6161	0.546 s
Mean	167	178	101	113	0.6215	0.5964	0.6087	0.6167	0.539 s

Table 10
Results obtained after extracting 200, 500 and 1000 features from the image dataset, using Spearman $\alpha = 0.001$.

	TP	TN	FP	FN	PR	RE	F1	AC	Time
200	221	75	205	59	0.5188	0.7893	0.6261	0.5286	0.548 s
500	218	75	205	62	0.5154	0.7786	0.6202	0.5232	0.490 s
1000	218	75	205	62	0.5154	0.7786	0.6202	0.5232	0.477 s
Mean	219	75	205	61	0.5165	0.7821	0.6221	0.525	0.505 s

Table 11
Results obtained after extracting 200, 500 and 1000 features from the image dataset, using CVM with $\alpha = 0.001$.

	TP	TN	FP	FN	PR	RE	F1	AC	Time
200	253	25	255	27	0.4980	0.9036	0.6421	0.4964	0.505 s
500	259	27	253	21	0.5059	0.9250	0.6540	0.5107	0.473 s
1000	254	28	252	26	0.5020	0.9071	0.6463	0.5036	0.478 s
Mean	255	27	253	25	0.5019	0.9119	0.6474	0.5035	0.485 s

where the environment is appropriately controlled, preventing any manipulation of the original images.

5.2. Scenario 2 - dataset with 18000 examples

Similarly to what was carried out in Section 5.1, in the new experiment it was analysed 9000 authentic images and 9000 manipulated images, for a total of 18000 images. The experiment consisted in extracting 200, 500 and 1000 features. After a thorough study of the results produced with the different significance degrees, emphasis was placed only on the 0.001 significance degree in the first scenario and 0.01 in the second scenario, as they produced the highest hit rate in detecting manipulated and authentic images. The results described in Section 5.1 were encouraging but insufficient to give a concrete answer on the need and robustness of the model. Two important factors contributed to this: a too-small dataset and many false positives. Therefore, it was necessary to perform a more comprehensive investigation where the dataset is larger.

In this new test scenario, several procedures were performed (some already done in the previous experiments) to obtain new answers that could validate the robustness of the model. Fig. 13 depicts the comparison of the dataset with Benford's law.

Fig. 13, suggests the existence of manipulations in the images in the new dataset, so it becomes imperative to perform an analysis with greater detail and depth to verify if the statistical models (Pearson, Spearman and Cramer-Von Mises) previously used, allow us to obtain an affirmative answer about which images were manipulated.

Tables 12–14 contain the results obtained after extracting 200, 500 and 1000 features from the authentic and manipulated images

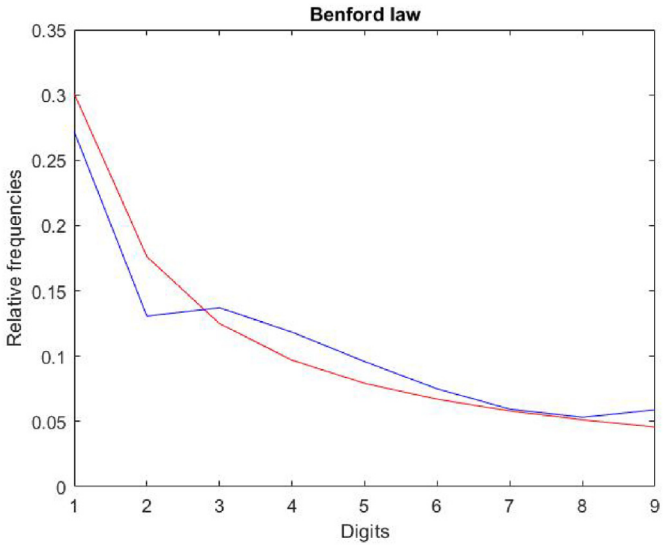

Fig. 13. Comparison between the curve resulting from Benford's law and the curve produced by the total characteristics of all the images in the dataset.

Table 12
Results obtained after extracting 200, 500 and 1000 features from the images dataset, using Pearson $\alpha = 0.01$.

	TP	TN	FP	FN	PR	RE	F1	AC
200	8529	7628	1372	471	0.8614	0.9477	0.9025	0.8976
500	8529	7686	1314	471	0.8665	0.9477	0.9053	0.9008
1000	8521	7674	1326	479	0.8653	0.9468	0.9042	0.8997
Mean	8526	7662	1337	474	0.8644	0.9474	0.9040	0.8993

Table 13
Results obtained after extracting 200, 500 and 1000 features from the image dataset, using Spearman $\alpha = 0.01$.

	TP	TN	FP	FN	PR	RE	F1	AC
200	5738	3411	5589	3262	0.5066	0.6376	0.5646	0.5083
500	5726	3453	5547	3274	0.5079	0.6362	0.5649	0.5099
1000	5751	3460	5540	3249	0.5093	0.6390	0.5669	0.5117
Mean	5738	3441	5559	3262	0.5079	0.6376	0.5655	0.5099

Table 14
Results obtained after extracting 200, 500 and 1000 features from the image dataset, using CVM with $\alpha = 0.01$.

	TP	TN	FP	FN	PR	RE	F1	AC
200	8333	4535	4465	667	0.6511	0.9259	0.7646	0.7149
500	8352	4572	4428	648	0.6535	0.9280	0.7669	0.7180
1000	8348	4573	4427	652	0.6535	0.9276	0.7668	0.7178
Mean	8344	4560	4440	656	0.6527	0.9272	0.7661	0.7169

from the new dataset. The processing time has been omitted, as it is identical to the time obtained in scenario 1.

Comparing the average values obtained in Tables 12 and 13 and 14, we can deduce that the Pearson model is the one that produces the best results in the two experiments performed, obtaining the best accuracy, 86.44%, relative to the other models (Pearson and Cramer-Von Mises). The number of misclassified images, covering false positives and false negatives, decreased, when compared to the previous experiment (Section 5.1), from 38.21% to 10.06%. Compared to Spearman's model, the number of misclassified samples is relatively high for false negatives and false positives, and the worst results were obtained compared to the other models.

About the Cramer-Von Mises model, the number of misclassified samples is high for false positives. About false negatives, it follows the line played by the Pearson model. As for the average accuracy, Pearson's model again presents the highest value, detecting a more significant amount of manipulated images than other models. Regarding the recall and F1-score metrics, Pearson's model offers the best result with values above 90% in detriment with the results obtained in the second experiment, where the CVM was better.

In this comparison, Pearson's model proves to be more homogeneous. Hence, the right choice for a correct classification of the manipulated images can be explained by several factors, among which we highlight the strong direct relationship between the variables under analysis, the fact that Pearson's model is a parametric model with the need for the variables to be normally distributed, as can be seen in Table 18, and the fact that the sample present in the last dataset was sufficiently large, which led to a considered increase in the accuracy of the results, Bonett and Wright (2000); Cohen (2013). Compared to conventional machine learning-based models, as can be seen in Table 15, Ferreira et al. (2021), the current model, underpinned by Pearson's classification model, rivals the results obtained by CNN and SVM-based methods. The F1 score obtained by Support Vector Machines

Table 15
Benchmark results.

	PR	RE	F1	AC	Time
DFT with SVM	0.9965	0.9941	0.9953	0.9951	00:00:51
CNN	0.9970	0.9966	0.9968	0.9967	06:36:00
BL-Pearson classification	0.8644	0.9474	0.9040	0.8993	00:25:23

reaches 99.8%, while Pearson's model reaches 90.4%, slightly lower than the best result obtained with the same dataset. The time factor is added as the main rival of such machine learning-based models.

In the presence of new image manipulation methodologies, machine learning-based models need the models to be re-trained, leading to a considerable increase in analysis and classification time. As the current model is based on statistical models, i.e. without training data, the data analysis and classification are limited only to the study of the data extracted from each image, with a shorter response time suitable for a forensic investigation.

Besides being necessary, the results obtained by Pearson's model are very encouraging but lack scientific confirmation. Thus, it is imperative to perform a set of tests that will serve as qualifiers of the association between variables under study, that is, between the frequency of digits empirically defined by Benford's Law and the frequencies of digits obtained by the extraction of the image characteristics. Among the tests proposed and performed is the homoscedasticity test, which checks whether the variances between the variables under study are equal, having been performed by Bartley's test and reinforced by the ANOVA test, Hair (2009); Levine et al. (2021).

5.2.1. Homoscedasticity between variables

The Bartlett test was used to study the homoscedasticity between the variables, which allows for checking whether the various data samples have equal variances against the alternative hypothesis that there are at least two variables with unequal variances.

The test statistic is given by Equation (19), Bartlett (1937).

$$T = \frac{(N-k)\,ln\,s_p{}^2 - \sum_{i=1}^{k}(N_i-1)\,ln\,s_i^2}{1 + \left(\left(\frac{1}{3\times(k-1)}\right)\right) \times \left(\left(\sum_{i=1}^{k}\frac{1}{N_i-1}\right) - \frac{1}{N-k}\right)} \tag{19}$$

where.

- S_i^2 is the ith group variance;
- N represents the total sample size;
- N_i is the size of each sample in the i-th group;
- k is the number of groups;
- s_p^2 represents the pooled variance.

Test statistics:

H_0: The frequencies with which digits occur in each image come from a normal distribution with the same variance; the variables are homogeneous.

H_1: The frequencies with which digits occur in each image do not come from a normal distribution with the same variance; the variables are not homogeneous.

The existence of a p-value less than the significance level of 0.05 (standard value) allows rejecting the initial hypothesis in favour of the alternative hypothesis. Finally, if the initial hypothesis is not rejected, Bartlett's test proves that the variables are homoscedastic, proving that the results are accurate.

Table 16, elucidates the homogeneity of the variables under study, showing that they are homoscedastic. This fact is proven by the p-value obtained 1, not rejecting the initial hypothesis.

Table 16
Bartlett's test results.

Image	Columns	Mean	Std
…	…	…	…
17995	9	0.1111	0.12059
17996	9	0.1111	0.07793
17997	9	0.1111	0.10284
17998	9	0.1111	0.11544
17999	9	0.1111	0.09089
18000	9	0.1111	0.10696
Pooled	162000	0.1111	0.09642
Bartlett's statistics	10185.2		
Degrees of freedom	17999		
P-value	1		

Furthermore, as can also be seen in the same figure, the standard deviation values are very close, reinforcing the equality of variances.

5.2.2. Test for equality of means

To reinforce the results obtained by Bartlett's test, an analysis of means was performed using the ANOVA analysis tool. This procedure aims to test whether the mean of the samples taken from a given population is equal, as opposed to the fact that the samples may have different means, Bartlett (1937).

The essential question we are confronted with at this point is related to the need to investigate whether the results obtained by including the Pearson model in the classification of images suffer from any effect that may be sensitive, rendering the results invalid. This way, we intend to determine if the population averages of all images are equal or if there are differences between them. The ANOVA model follows a structure in which the variability of the results presents a central tendency with essential attributes that must be verified. Among these attributes are the normality of the results; the results are independent and give a null mean and constant or similar variance (homoscedasticity verified in 5.2.1), Erjavec (2011).

Test statistics:

H_0: The frequencies with which the digits occur in each image come from samples with the same mean.

H_1: The frequencies with which the digits occur in each image do not come from samples with the same mean.

As seen in Table 17, the p-value obtained by performing the ANOVA test allows us to conclude that the mean of the samples is significantly equal, not rejecting the initial hypothesis.

To validate the hypotheses of the ANOVA model, it is imperative to demonstrate the normality of the data. To do this, we studied the frequency with which the digits occur in their entirety dataset.

Test statistics:

$H_0 : X \sim \mathbf{N}(\mu, \theta)$: Digit frequencies come from a normal distribution.

$H_1 : X \nsim \mathbf{N}(\mu, \theta)$: The digit frequencies do not come from a normal distribution.

Table 18 shows that the frequency with which the digits occur in the complete data set follows a normal distribution, duly confirmed by the p-value obtained ($0.1376 > 0.05$), not rejecting the initial hypothesis.

Table 17
ANOVA results.

Source	SS	df	MS	F	Prob > F
Columns	0	17999	0	1.28184e-32	1
Error	1338.84	144000	0.0093		
Total	1338.84	161999			

Table 18
Test of normality on the occurrence of digits.

Lillieffors - normality test	
Data	Relative frequency
D	0.24104
P-value	0.1376

We can conclude the integrity of the results obtained by Pearson's model through the normality test carried out, confirming the results obtained by the homoscedasticity and equality of the averages. Furthermore, these results showed that the frequency with which the digits occur admits similar means and variations, allowing us to conclude that the results obtained by the image classification are adjusted to reality.

6. Conclusions and future work

This paper described the application of Benford's law to a dataset containing authentic and manipulated digital images to detect manipulated digital content. Initially, a set of features was extracted from the images obtained by calculating the DFT for each multimedia file, allowing the extraction of the first digit. Then, from the vector of available digits, absolute and relative frequencies were calculated, allowing the application of a set of statistical models (Pearson, Spearman and Cramer-Von Mises) in the form of hypothesis tests, whose aim was to create a mechanism to verify whether a given digital image was authentic or manipulated.

To this end, correlations and P-values were calculated between the relative frequencies obtained by the digits of the images and the empirical frequency of Benford's law. This operation generated a set of labels that classified the images with 1 in case they were authentic and 0 in case they were manipulated. Finally, the original labels of the images and the new labels obtained from the statistical models were compared, generating a set of valid results for further analysis.

A careful review of the most recent and up-to-date literature, widely related to the problem domain, was carried out. Previously existing works related to the topic under analysis address the application of Benford's law to a dataset in its entirety (creating a false illusion that it contains no manipulated images), not focusing the investigation on the classification of each image and whether or not it was subject to manipulation.

In this paper, we investigated the individual classification of each image to detect any manipulated images. Two scenarios, created from various sources, were contributed for this research. Two test scenarios were carried out: 1) 280 manipulated and 280 authentic images, for a total of 560 images; 2) c9000 manipulated and 9000 authentic images, in a total of 18000 images.

The manipulated images underwent specific manipulations such as splicing and copy-move techniques. The results obtained by applying the probabilistic models took into account the existing correlation between the frequencies of each digit obtained from the characteristics of the images and the frequencies obtained by Benford's law.

The performance evaluation between the three statistical models employed was based on a set of performance metrics, where some comparisons with machine learning approaches were also calculated. The maximum F1 score attained was 90.40% for the detection of manipulated images, very close to the results obtained by other methods, such as CNN or SVM. However, the processing time required for detecting manipulated images using these methodologies far exceeds the time needed for the proposed

model, where the worst time obtained was 0.548 *s* versus the 51*s* using SVM, as documented in Ferreira et al. (2021).

Concerning the first experiment, increasing the number of image features did not change the results obtained. Therefore, it is concluded that the use of 1000 features, at most, is sufficient to decide whether an image has undergone any manipulation.

Regarding the second experiment, increasing the number of images in the dataset allowed for significantly improved results obtained by comparison with the results of the first experiment. For this improvement, the major contribution is given by the Pearson model, and being a parametric model, it allowed a strong impact on the results.

The calculation of homoscedasticity and the application of the ANOVA model allowed us to conclude that both the means and the variance deviation were in agreement with the frequencies obtained by each image, attesting to the homogeneity of the sample. It is inferred that the number of elements that constitute the database will make it possible to obtain greater reliability of the proposed model, to the detriment of the low number of elements, which will sacrifice the model's sensitivity.

The possible causes for a lower performance of the proposed model, in the first experiment, when compared with models based on machine learning, are related to how the images were obtained, in which the process of reducing an image affects the quantity and quality of the pixels, the possibility of being in front of images with low resolutions, the possibility that the FFT method used may present some limitations or in the low sensitivity of the correlation coefficients used.

In this study, it was demonstrated that the proposed model attained better results in the presence of larger datasets, duly confirmed by homogeneity tests between the variables under study. It was also shown that the model based on Benford's law should use parametric models for a correct classification of images, to the detriment of non-parametric models where the worst results were obtained.

The second experiment, Benford's law obtained better results with the application of Pearson's model, with a decrease in false positives from 38, 21% to 10, 06% by comparison with the first experiment. The results are consistent with those obtained in the first experiment for the Spearman model, where it can be concluded that this model is not adequate to be used as a classifier of image manipulation.

In the opposite direction, further research is necessary for applying the Cramer-Von Mises model as a classifier of image manipulation, as it allowed the detection of an extensive set of manipulated images but obtained also a high number of false positives. A possible solution is to obtain all the characteristics of the image provided by the original data source and then apply the mean absolute deviation from the averages that may arise.

Possible lines of research are: to employ Benford's Law based models on the detection of anomalies, in general; to obtain more digits from Bedford' law; to process the complete digital image; to investigate the benefits of using Fourier-Mellin transform to extract features; and to apply new correlations such as the Monte Carlo or the Kolmogorov methods.

CRediT authorship contribution statement

Pedro Fernandes: ¡Credit authorship details¿. **Mário Antunes:**.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

https://github.com/Pacfes/Benford-Law

Abbreviations

A	Accuracy
AI	Artificial Intelligence
CNN	Convolutional Neural Networks
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
FFT	Fast Fourier Transform
FP	False Positive
FN	False Negative
FTK	Forensic Toolkit
GPU	Graphics Processing Unit
JPEG	Joint Photographic Experts Group
P	Precision
R	Recall
SVM	Support Vector Machine
TP	True Positive
TN	True Positive

References

del Acebo, E., Sbert, M., 2005. Benford's Law for Natural and Synthetic Images, pp. 169–176. https://doi.org/10.2312/COMPAESTH/COMPAESTH05/169-176.

Amerini, I., Li, C.T., Memon, N., Huang, J., 2020. IEEE access special section: digital forensics through multimedia source inference. IEEE Access 8, 209657–209659. https://doi.org/10.1109/access.2020.3036772.

Anderson, T.W., 1962. On the distribution of the two-sample cramer-von Mises criterion. Ann. Math. Stat. 33, 1148–1159. https://doi.org/10.1214/aoms/1177704477. URL: https://doi.org/10.1214/aoms/1177704477.

Arno Berger, T.P.H., 2015. An Introduction to Benford's Law. Princeton University Press. URL: https://www.ebook.de/de/product/23323656/arno_berger_theodore_p_hill_an_introduction_to_benford_s_law.html.

Bardera, A., Feixas, M., Boada, I., Sbert, M., 2006. Compression-based image registration. In: 2006 IEEE International Symposium on Information Theory, pp. 436–440. https://doi.org/10.1109/ISIT.2006.261706.

Bartlett, M.S., 1937. Properties of Sufficiency and Statistical Tests, vol. 160. Proceedings of the Royal Society of London, pp. 268–282. URL: http://www.jstor.org/stable/96803.

Berger, A., Hill, T.P., 2011. A basic theory of Benford's Law. Probab. Surv. 8, 1–126. https://doi.org/10.1214/11-PS175. URL: https://doi.org/10.1214/11-PS175.

Best, D.J., Roberts, D.E., 1975. Algorithm AS 89: the upper tail probabilities of spearman's rho. Appl. Stat. 24, 377. https://doi.org/10.2307/2347111.

Bonett, D.G., Wright, T.A., 2000. Sample size requirements for estimating pearson, kendall and spearman correlations. Psychometrika 65, 23–28. https://doi.org/10.1007/bf02294183.

Caelen, O., 2017. A bayesian interpretation of the confusion matrix. Ann. Math. Artif. Intell. 81, 429–450. https://doi.org/10.1007/s10472-017-9564-8.

Cohen, J., 2013. Statistical Power Analysis for the Behavioral Sciences. Routledge. https://doi.org/10.4324/9780203771587.

Cooley, J.W., Tukey, J.W., 1965. An algorithm for the machine calculation of complex fourier series. Math. Comput. 19, 297–301. https://doi.org/10.1090/s0025-5718-1965-0178586-1.

Enisa, 2021. URL: https://www.enisa.europa.eu/topics/threat-risk-management/threats-and-trends/. (Accessed 30 June 2022).

Erjavec, N., 2011. Tests for homogeneity of variance. In: International Encyclopedia of Statistical Science. Springer Berlin Heidelberg, pp. 1595–1596. https://doi.org/10.1007/978-3-642-04898-2_590.

Europol. Cybercrime. URL: https://www.europol.europa.eu/crime-areas-and-statistics/crime-areas/cybercrime. (Accessed 30 June 2022).

Ferreira, S., 2021. Photos-Videos-manipulations-dataset. Github. URL: https://github.com/saraferreirascf/Photos-Videos-Manipulations-Dataset.

Ferreira, S., Antunes, M., Correia, M.E., 2021. Exposing manipulated photos and videos in digital forensics analysis. J. Image. 7, 102. https://doi.org/10.3390/jimaging7070102.

Gimp. The free open source image editor. https://www.gimp.org/. [Online. (Accessed 18 July 2022).

Hair, J.F., 2009. Multivariate Data Analysis.

Harris, D.A., 2019. Deepfakes: false pornography is here and the law cannot protect you. Duke Law Technol. Rev. 17, 99–127.

Hill, T.P., 1995. Base-invariance implies Benford's law. Proc. Am. Math. Soc. 123.

Johnson, V.E., 2013. Revised standards for statistical evidence. Proc. Natl. Acad. Sci. USA 110, 19313–19317. https://doi.org/10.1073/pnas.1313476110.

Karras, T., Laine, S., Aila, T., 2019. A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
Karras, T., Nvidia, 2019. This Person Does Not Exist. https://thispersondoesnotexist.com/.
Krzywinski, M., Altman, N., 2013. Significance, p values and t-tests. Nat. Methods 10, 1041–1042. https://doi.org/10.1038/nmeth.2698.
Kumar, A., Singh, G., Kansal, A., Singh, K., 2021. Digital image forensic approach to counter the JPEG anti-forensic attacks. IEEE Access 9, 4364–4375. https://doi.org/10.1109/access.2020.3048246.
Lesperance, M., Reed, W.J., Stephens, M.A., Tsao, C., Wilton, B., 2016. Assessing conformance with benford's law: goodness-of-fit tests and simultaneous confidence intervals. PLoS One 11, e0151235. https://doi.org/10.1371/journal.pone.0151235.
Levine, D., Stephan, D., Szabat, K., 2021. Statistics for Managers.
Li, Y., jian Zhou, Y., guo Yuan, K., cui Guo, Y., xin Niu, X., 2014. Exposing photo manipulation with inconsistent perspective geometry. J. China Univ. Posts Telecommun. 21, 83–104. https://doi.org/10.1016/S1005-8885(14)60320-4.
Lin, X., Li, J.H., Wang, S.L., Liew, A.W.C., Cheng, F., Huang, X.S., 2018. Recent advances in passive digital image security forensics: a brief review. Engineering 4, 29–39. https://doi.org/10.1016/j.eng.2018.02.008.
Mar-Raave, J.R.D., Bahşi, H., Mršić, L., Hausknecht, K., 2021. A machine learning-based forensic tool for image classification - a design science approach. Forensic Sci. Int.: Digit. Invest. 38, 301265. https://doi.org/10.1016/j.fsidi.2021.301265.
Meena, K.B., Tyagi, V., 2019. Image forgery detection. Survey Future Direct. 163–194. https://doi.org/10.1007/978-981-13-6351-1_14.
Mire, A., 2022. Tampering localization using divergence in generalized benford's model of first digit probability distribution in JPEG images. J. Informatic. Assurance Security. 17, 46–54.
Muzaffer, G., Ulutas, G., 2019. A new deep learning-based method to detection of copy-move forgery in digital images. In: 2019 Scientific Meeting on Electrical-Electronics &amp Biomedical Engineering and Computer Science (EBBT). IEEE. https://doi.org/10.1109/ebbt.2019.8741657.
Ng, T.T., Hsu, J., Chang, S.F., 2004. Columbia image splicing detection evaluation dataset. URL: https://www.ee.columbia.edu/ln/dvmm/downloads/AuthSplicedDataSet/AuthSplicedDataSet.htm.
Nunes, A., Inacio, H., Marques, R.P., 2019. Benford's law and fraud detection in Portuguese enterprises. In: 2019 14th Iberian Conference on Information Systems and Technologies (CISTI). IEEE. https://doi.org/10.23919/cisti.2019.8760922.
Parfieniuk, M., 2021. Lifting-based alternatives to the FFT for computing the 4-, 8-, and 16-point discrete fourier transforms. In: 2021 29th European Signal Processing Conference (Eusipco). IEEE. https://doi.org/10.23919/eusipco54536.2021.9616225.
Parnak, A., Baleghi, Y., Kazemitabar, J., 2020. A Novel Forgery Detection Algorithm Based on Mantissa Distribution in Digital Images. https://doi.org/10.1109/icspis51611.2020.9349611.
Parnak, A., Baleghi, Y., Kazemitabar, J., 2022. A novel image splicing detection algorithm based on generalized and traditional benford's law. Int. J. Eng. 35, 626–634. https://doi.org/10.5829/ije.2022.35.04a.02.
Pasquini, C., Boato, G., Perez-Gonzalez, F., 2017. Statistical detection of JPEG traces in digital images in uncompressed formats. IEEE Trans. Inf. Forensics Secur. 12, 2890–2905. https://doi.org/10.1109/tifs.2017.2725201.
Pedrini, H., Schwartz, W.R., 2008. Analise de imagens digitais principios, algoritmos e aplicacoes. Cengage Learning Edições, lda. URL: https://www.ic.unicamp.br/helio/book_aid/index.html.
Photoshop. Photoshop. https://www.adobe.com/products/photoshop.html. (Accessed 18 July 2022).
Qadir, G., Zhao, X., Ho, A.T., Casey, M., 2011. Image forensic of glare feature for improving image retrieval using benford's law. In: 2011 IEEE International Symposium of Circuits and Systems (ISCAS). IEEE. https://doi.org/10.1109/iscas.2011.5938152.
Rajan, A.V., Ravikumar, R., Shaer, M.A., 2017. UAE cybercrime law and cybercrimes — an analysis. In: 2017 International Conference on Cyber Security and Protection of Digital Services (Cyber Security). IEEE. https://doi.org/10.1109/cybersecpods.2017.8074858.
Rajesh, B., Javed, M., Ratnesh, Srivastava, S., 2019. DCT-CompCNN: a novel image classification network using JPEG compressed DCT coefficients. In: 2019 IEEE Conference on Information and Communication Technology. IEEE. https://doi.org/10.1109/cict48419.2019.9066242.
Rao, K.R., Kim, D.N., Hwang, J.J., 2010. Discrete fourier transform. In: Fast Fourier Transform - Algorithms and Applications. Springer Netherlands, pp. 5–40. https://doi.org/10.1007/978-1-4020-6629-0_2.
Said, T., Mohammed, K., 2020. Detection of anomaly in socio-economic databases, by benford probability law, in: 2020 IEEE 6th International Conference on Optimization and Applications (ICOA). IEEE ASME Trans. Mechatron. https://doi.org/10.1109/icoa49421.2020.9094466.
Saini, M., Kapoor, A.K., 2016. Biometrics in forensic identification: applications and challenges. J. Forensic Med. 1. https://doi.org/10.4172/2472-1026.1000108.
Satapathy, G., Bhattacharya, G., Puhan, N.B., Ho, A.T.S., 2020. Generalized benford's law for fake fingerprint detection. In: 2020 IEEE Applied Signal Processing Conference (ASPCON). IEEE. https://doi.org/10.1109/aspcon49795.2020.9276660.
Singh, N., Bansal, R., 2015. Analysis of Benford's Law in Digital Image Forensics. https://doi.org/10.1109/icspcom.2015.7150688.
Sreenivasu, Vani, S., 2017. Copy-Move Digital Image Forgery Detection techniques: a Review URL: https://www.recentscientific.com/sites/default/files/7700-A-2017.pdf.
Taimori, A., Razzazi, F., Behrad, A., Ahmadi, A., Babaie-Zadeh, M., 2012. A Proper Transform for Satisfying Benford's Law and its Application to Double JPEG Image Forensics. https://doi.org/10.1109/isspit.2012.6621294.
Thakur, R., Rohilla, R., 2020. Recent advances in digital image manipulation detection techniques: a brief review. Forensic Sci. Int. 312, 110311. https://doi.org/10.1016/j.forsciint.2020.110311.
Tharwat, A., 2020. Classification assessment methods. Apll. Compute. Informatic. 17, 168–192. https://doi.org/10.1016/j.aci.2018.08.003.
Thirumalai, C., Chandhini, S.A., Vaishnavi, M., 2017. Analysing the concrete compressive strength using pearson and spearman. In: 2017 International Conference of Electronics, Communication and Aerospace Technology (ICECA). IEEE. https://doi.org/10.1109/iceca.2017.8212799.
Unodc, 2019, 2019. URL: https://www.unodc.org/e4j/en/cybercrime/module-4/key-issues/standards-and-best-practices-for-digital-forensics.html. (Accessed 2 July 2022).
Volčič, A., 2020. Uniform distribution, benford's law and scale-invariance. Boll. Unione Mat. Ital. 13, 539–543. https://doi.org/10.1007/s40574-020-00245-6.
Wang, J., Cha, B.H., Cho, S.H., Kuo, C.C.J., 2009. Understanding Benford's Law and its Vulnerability in Image Forensics. https://doi.org/10.1109/icme.2009.5202811.
Wang, Q., Zhang, R., Qing, K., 2015. Passive Detection of Tampered JPEG Image Based on First Digit Statistics. https://doi.org/10.1109/iih-msp.2015.42.
Wei, X., Guo, Y., Li, B., 2021. Black-box adversarial attacks by manipulating image attributes. Inf. Sci. 550, 285–296. https://doi.org/10.1016/j.ins.2020.10.028.
Wen, B., Zhu, Y., Subramanian, R., Ng, T.T., Shen, X., Winkler, S., 2016. Coverage — a novel database for copy-move forgery detection. In: 2016 IEEE International Conference on Image Processing. ICIP), pp. 161–165. https://doi.org/10.1109/ICIP.2016.7532339.
Wolf, C., Jolion, J.M., Kropatsch, W., Bischof, H., 2000. Content based image retrieval using interest points and texture features. In: Proceedings 15th International Conference on Pattern Recognition, vol. 4. ICPR, pp. 234–237. https://doi.org/10.1109/ICPR.2000.902902, 2000.
Wu, Y., Wo, Y., Han, G., 2022. Joint manipulation trace attention network and adaptive fusion mechanism for image splicing forgery localization. Multimed. Tool. Appl. https://doi.org/10.1007/s11042-022-13151-0.
Yang, J., Zhu, G., Huang, J., Zhao, X., 2015. Estimating JPEG compression history of bitmaps based on factor histogram. Digit. Signal Process. 41, 90–97. https://doi.org/10.1016/j.dsp.2015.03.014.
Yao, H., Wei, H., Qiao, T., Qin, C., 2020. Jpeg quantization step estimation with coefficient histogram and spectrum analyses. J. Vis. Commun. Image Represent. 69, 102795. URL: https://www.sciencedirect.com/science/article/pii/S1047320320300456, 10.1016/j.jvcir.2020.102795.
Zhao, X., Ho, A.T., Shi, Y.Q., 2009. Image forensics using generalised benford's law for accurate detection of unknown jpeg compression in watermarked images. In: 2009 16th International Conference on Digital Signal Processing, pp. 1–8. https://doi.org/10.1109/ICDSP.2009.5201261.