Microscopic printing analysis and application for classification of source printer

Quoc-Thông Nguyen, An Mai, Lionel Chagas, Nadège Reverdy-Bruas

Microscopic Printing Analysis and Application for Classification of Source Printer

Quoc-Thông Nguyen[a], An Mai[b], Lionel Chagas[c], Nadège Reverdy-Bruas[c]

[a]*GEMTEX Laboratory, Ecole Nationale Supérieure des Arts et Industries Textiles, Roubaix, 59560, France*
[b]*School of Computer Science & Engineering, International University, and Vietnam National University, Ho Chi Minh City, Vietnam*
[c]*University Grenoble Alpes, CNRS, Grenoble INP (Institute of Engineering, Univ. Grenoble Alpes), LGP2, F-38000 Grenoble, France*

Abstract

Identifying a forged printed document with scanned evidence can be a challenge. Microscopic printing is showing random shape which depends on the printing source as well as printing material. This paper presents a statistical analysis of the printing patterns under a microscopic scale, analyses the effect of printing direction, printing substrate (uncoated and coated paper), and printing technology (conventional offset, waterless offset, and electrophotography). The analysis shows a negligible effect of printing direction, yet, using the shape descriptor indexes, the printing materials and technologies are distinguishable under a microscopic scale. As a result, the algorithms based on Support Vectors Machine and Random Forest are developed, with shape descriptor indexes as features, for printing source identification. Both proposed algorithms, equally, achieve a high classification accuracy rate, over 92% accuracy with complex geometric-shape patterns. Thanks to the lightweight and efficiency of the Support Vectors Machine, the study shows promising applications for real-world and potential implementation in the Internet of Things devices.

Keywords: Microscopic printing, Source Printer Identification, Printer Forensics, Document Authentication, Support Vector Machine.

*Corresponding author
Email address: nguyenquocthong1111@gmail.com (Quoc-Thông Nguyen)

1. Introduction

Intellectual property infringement is always an issue in production industry. One of them, printing forgery using the high quality printers has been applied to reproduce the printed content of the original. In the recent years, the access to the high quality printing technologies becomes easily, the importance of document security has increased to prevent the counterfeiting or forgery of documents [3, 4]. As a matter of fact, it is a great motivation to create an affordable, portable, and reliable system to perform authentication on printed documents for the forensic specialists. There are various ways to perform the printed document authentication, for example, one of the common approaches consists in embedding an extrinsic signature as an authentic mark such as secure tags in the prints [32]. In another approach, it is relied on the non-invertible noises from the re-printing process of the original printed documents, the authentication is then performed based on a reliable statistical testing framework to detect the counterfeit prints [13, 29]. However, the opponents from the illegitimate source can employ some statistical tricks to estimate the printing parameters of the legitimate source. To address this issue, Phan Ho et al. [27, 28] modelled the testing process between two sources as a min-max game, which is still able to carry out the statistical test successfully and detect the illegitimate cases at some sort of magnitudes of test statistic.

Several examples of intrinsic features, it can be figured out here the banding artefact coming from fluctuation of the optical photoconductor angular velocity in case of electrophotography printers [1], dimple effect which is specific to inkjet printers or also texture features [33], or the analysis of quality signature of the unique print to differentiate one printer technology/supplier from another [22]. It is easy to realise that authentication can be performed based on paper statistics. For example, the paper statistics is extracted by capturing microstructure images from different paperboard packages under different light conditions from different cameras and with different angles. These microstructure images are considered as digital fingerprinting on which the authentication frameworks are derived [42]. In [35], the study proposed to characterise the intrinsic feature of paper based on a texture speckle pattern, specifically, a random

bright/dark region formation at the microscopic level when light falls on the paper.

It is known that the extracted signature of electrophotographic printer is a unique pattern of misplaced toner powder on each paper [43], and this property is very important in the applications of source identification. For instance, in the case of electrophotographic printers, the intrinsic signals extracted from banding artefact are used to identify the device [15]. In this field of application, ingeniously characterise and embedded a good intrinsic feature is an important approach in passive printer detection, for example in [32], the authors consider an extrinsic feature modelling as colour tracking dots, embedded in most of the modern colour laser printers. A list of clever techniques using characters to perform authentication of source scanner and printer have been proposed [8]. Source identification relying on the extracted features from the noises of printed characters in the documents has also been addressed in [34]. The printed character in different languages is also used in source identification [37, 36, 39, 38]. A recent study [6] focused on the use of local texture patterns from the scanned images of all printed letters, to perform source printer classification. A data-driven approach based on deep learning is explored for laser printer attribution [5].

There are a number of other applications concerning the printing models. For instance, in the human vision scale, halftoning technique is a method that takes advantage of the optical illusion of human vision to display continuous tone grayscale image with only black or/and white dots [40]. The scale of the model is the resolution of the printer, and an image of print not only contains the properties of a printing process but also includes image degradation. Under the microscopic scale printing, a probabilistic model for the average coverage of the toner particles ink of electrophotographic printing process is introduced in [21]. However, the model has a limited validation, and the estimation procedure has not been considered. The shapes of microscopic printed dots can be considered as the intrinsic feature of the printing process. Indeed, at the microscopic scale, a dot is a random pattern whose shape depends on the technology, the setting of the printer, the ink quality and/or the paper properties. From a statistical point of view, the digital acquisition of these random dots can be modelled as a spatial interaction binary model based on an exponential power kernel that depends on location parameter and shape parameters. In [17, 16, 18], the authors proposed a parametric

model consisting in a spatial distribution model that simulates the randomness of the printed dots shape. In the further study, Nguyen et al. [19] proposed a probabilistic model consisting of vector parameters describing a spatial interaction binary model with inhomogeneous Markov chain. These parameters determine the location and describe the diverse random structures of microscopic printed patterns. Recently, with the rising of 3D printing technologies, the study in [26, 25] constructed an equipment distortion model and introduced a uniform mark, designed to represent the inherent equipment distortion, as authentic signature. The authors employed Support Vector Machine for source identification of 3D printed objects.

This paper explores statistically the shape descriptor indexes of the microscopic printed pattern based on the printing technology and printing material. We develop an approach to identify the authentic printer source using micro-tags consisting of patterns of microscopic printed dots in the documents. The features of these patterns of printed dots are extracted and fed into multi-class Support Vector Machine (SVM) and Random Forest (RF) classification for the printer source identifications. The real patterns of printed dots from the common printing technologies (conventional offset, waterless offset, electrophotography) are used to assess the effectiveness of the proposed algorithms.

Section 2 is devoted to present the influences of the physical properties of ink, substrates as well as technology on the quality of the document with different patterns. The statistical analysis of patterns of printed dots is discussed in Section 3. The proposed algorithm and experimental results for printer identification are presented in Section 4. Finally, the concluding remarks are given in Section 5.

2. Materials and experiments

2.1. Printing technologies

In this study, two commonly used printing processes are investigated: offset and electrophotography. Offset printing is a lithographic technology, in which the printing areas are ink-accepting whereas the non-printing parts are ink-repellent. The inked image is transferred from a printing plate, which contains the pattern image, to a rub-

ber blanket, then to the printing surface [9]. Figure 1 depicts the process of offset printing. There are two common systems to obtain an ink-repellent areas in offset printing technology. In *Conventional Offset Printing*, the image-carrier plate is dampened with water and additives. Only the non-printing areas are water-receptive (hydrophilic), whereas the image areas are almost totally unreceptive to the dampening solution. While in *Waterless Offset Printing*, the printing plate is overlaid with a highly ink-repellent silicone on the ink-free areas. The printing ink is only accepted on the areas in which the silicone coating has been removed. The technical information of two offset printers used in this experiment is given in Table 1.

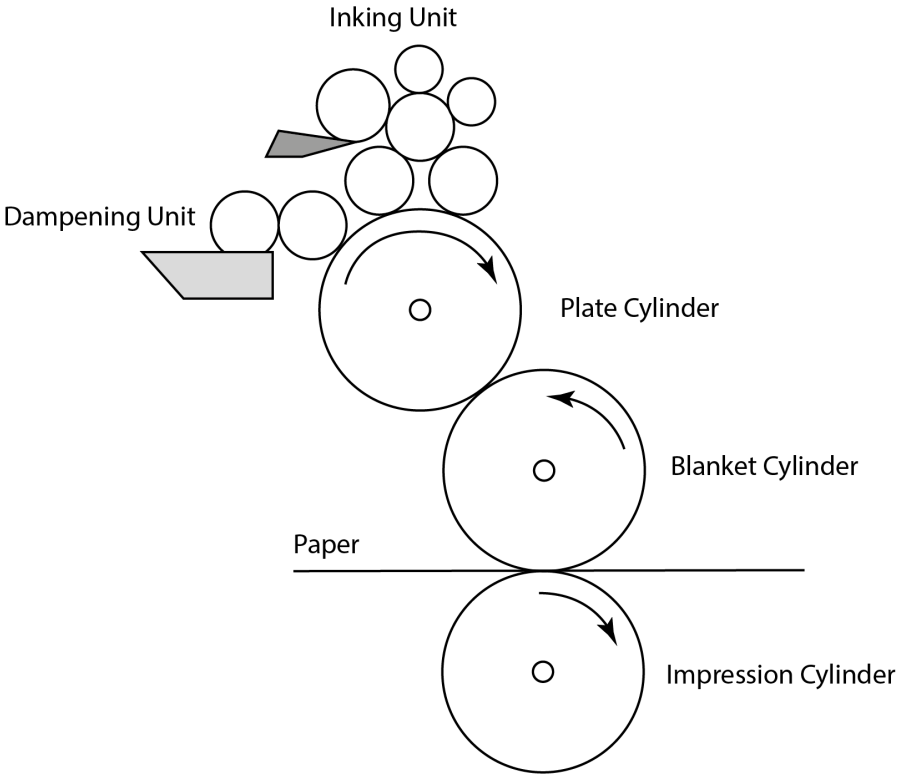

Figure 1: Schematic diagram of offset printing.

[1] dot per inch: the number of individual dots that can be placed in a line within the span of 1 inch (2.54 cm).

Name	CODIMAG	SEAILLE & TISON
Type	Waterless offset	Conventional offset
Web	Semi-rotating system	Rotating system
Width (mm)	420	480
Max Resolution (dpi [1])	2400	2400
V-max (m/mn)	50	200
Classical optical density	1.8	1.8

Table 1: Characteristics of offset printers used to print the test form.

In electrophotography, also called laser printer, photoconductor will be electrostatically charged with the laser beam, then it collects the particular inks, powder or liquid toners, having the opposite charge with the photoconductor surface [9]. The ink is then transferred to the substrate and fixed on it by an elevated temperature. The principle of electrophotography printing is illustrated in Figure 2. The printer used in this study is a HP-600 M620 with max resolution 1200 dpi.

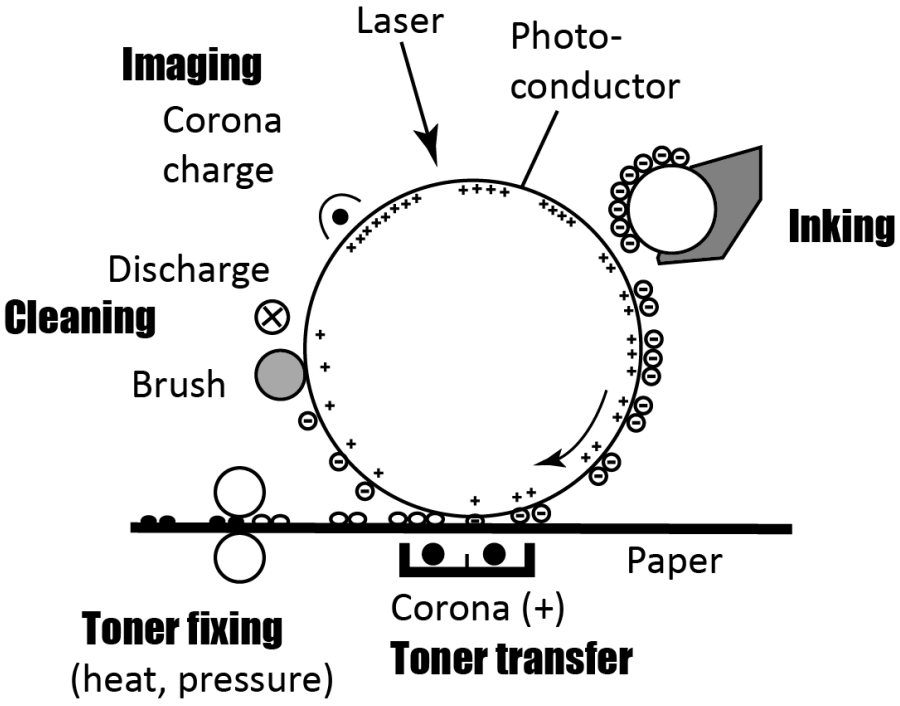

Figure 2: Electrophotography printing process.

2.2. Printing process and substrates

We present the influences of the physical properties of ink and substrates on the quality of the document with different patterns. As mentioned, the samples are printed by conventional offset, waterless offset, and electrophotographic printers, on two kinds of substrate: coated and uncoated paper. Coated paper is a paper coated with one or several layers to improve its surface roughness. During the coating process, the paper web is mechanically routed between rollers under pressure, friction and warmth which makes the paper thinner than the uncoated natural paper, also being used in the experiment. Due to the characteristics of each technology, the offset process will use offset paper, and paper for electrophotography is used for laser printer. In Table 2, the physical properties of each type of paper are presented, 10 samples of each type of paper are measured.

	Offset		**Electrophotography**	
Coated paper	Thickness (μm)	Roughness (μm)	Thickness (μm)	Roughness (μm)
Mean	85	0.9	98	2.1
Std.	1	0.1	1	0.15
Uncoated paper				
Mean	149	5.9	103	5.1
Std.	1	0.1	1	0.1

Table 2: Physical characteristics of coated and uncoated paper used in offset printing process and electrophotography printing process.

In addition, the physical properties of printing ink also plays a particular role in each technology. For offset printing, the printing ink has s high viscosity, dynamic viscosity = 40-100 Pa.s (see [9] page 137). For the conventional offset printing, the ink is able to achieve the additive of dampening solution, and typically contains water [11]. Waterless offset ink has a comparatively higher viscosity and is generally stiffer than conventional offset ink. Due to the fundamental difference between the two processes, conventional offset and waterless offset prints are different. In the "wet" process in conventional offset printing, the solution has lower cohesion than the ink, the water may remain on the ink film then make lower contrast on the image, Figure 3. On the other hand, for the electrophotography printing, special ink is applied, called toners. Toners can be powder or liquid, which explains the reason that the image has the shape of a

cluster of particles, see Figure 4. The printer HP-600 M620 has maximum resolution 1200 dpi. However, under the resolution 1200 dpi of the electrophotography printer, the shape of different printed patterns is visually indistinguishable. Therefore, the samples of electrophotography printer are printed under the resolution 600 dpi. For the same reason, the tested samples from the offset printers are printed under the resolution 1200 dpi, instead of the maximum solution 2400 dpi.

In the sample collection procedure, the samples are collected using an optical Zeiss Microscope with an AxioCam camera. The camera has a maximum solution of 2464 $\times$ 2056 (5Mp) with a sensor size of 8.5 mm $\times$ 7.1 mm and a pixel size of 3.45 μm $\times$ 3.45 μm.

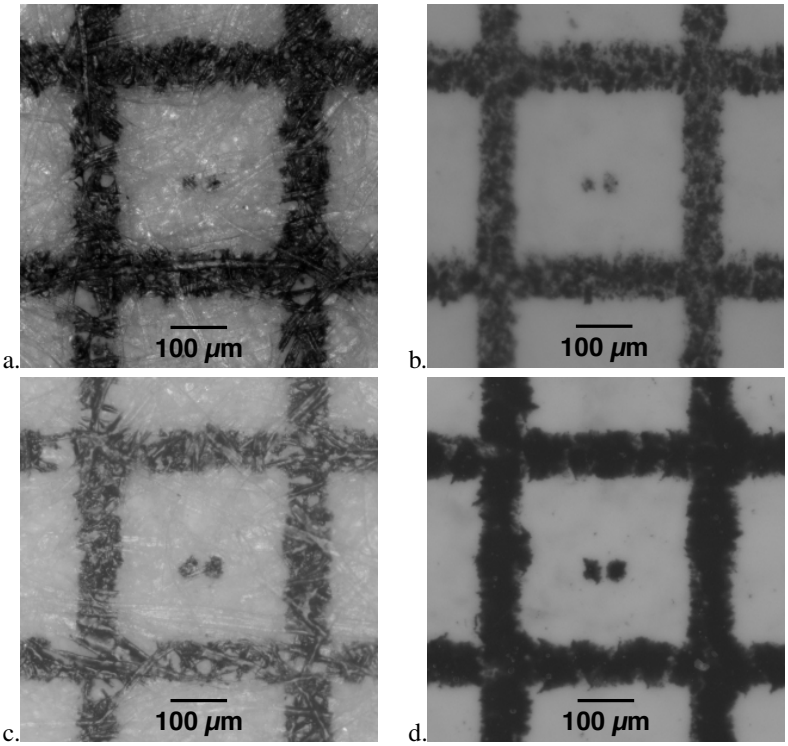

Figure 3: The conventional (a, b) and waterless (c, d) offset print on uncoated (a, c) and coated (b, d) papers, resolution 1200 dpi.

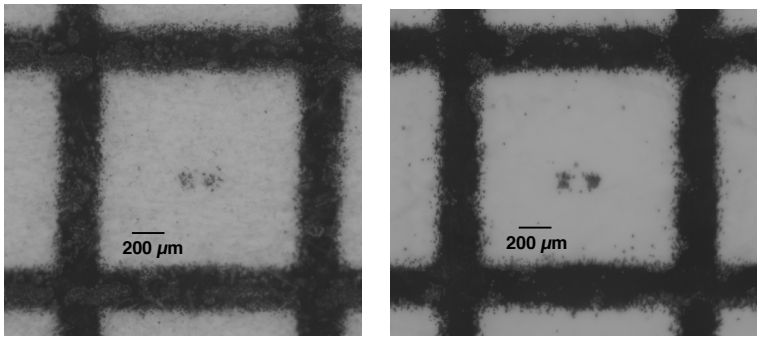

Figure 4: The print on uncoated (left) and coated (right) paper, HP-600 M620 printer with resolution 600 dpi.

2.3. Pattern analysis

Figure 5 presents eight different patterns selected for the analysis in our study. For each pattern, each kind of printer and paper, 100 copies were printed and captured by Zeiss Microscope. In total, the database has $100 \times 8 \times 2$ (paper)$\times 3$ (printers) = 4800 samples. From these different patterns, the objective is to understand if the direction of the printing process affects the degradation of printed patterns. The study on the single dot was carried out under the micro-scale [19]. The printed parts are the main concern in the experiment, thus the binary images are required. There are a number of threshold methods to convert grey-scale images to binary images [24, 10, 12]. Each method, for a specific purpose, has some advantages and also few drawbacks, based on the requirement of the application. In this study, we aim to discriminate the printers and the materials, so that a single binarization algorithm was chosen. The maximum entropy method mentioned in [7] performs properly. The images binarised by the maximum entropy method are demonstrated in Table 6.

For each printed pattern sample, the shape descriptor indexes [23, 41] are computed. Shape descriptor indexes are obtained from the printed segments by computing

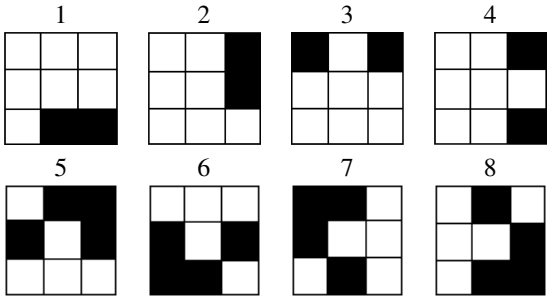

Figure 5: The 8 patterns used to print on 2 types of paper with 3 printing technologies.

their area, perimeter, and convex area,

$$\text{Solidity} = \frac{A}{A_c}, \tag{1}$$

$$\text{Convexity} = \frac{P_c}{P}, \tag{2}$$

$$\text{Circularity} = \sqrt{\frac{4\pi A}{P^2}}, \tag{3}$$

where A (px^2) is the area, A_c (px^2) is the convex area, P (px) is the perimeter, and P_c (px) is the convex perimeter. With the unit "px" stands for "pixel", since the unit of the indexes does not play a significant role in this study as long as it is the same for all samples. For each printed pattern sample, a vector of five features (*area, perimeter, solidity, convexity, circularity*) is considered.

3. Statistical analysis

In this section, we analyse statistically the features of the printed patterns with vertical and horizontal printing directions to investigate the effect of printing direction, see Figure 5. The mean values and standard deviations of the indexes of eight patterns with different combinations of printing technologies and substrates are presented in Table 3.

In order to compare the means of two populations, we use hypothesis testing method. Since the samples are not assumed to be normally distributed, the Mann-Whitney-

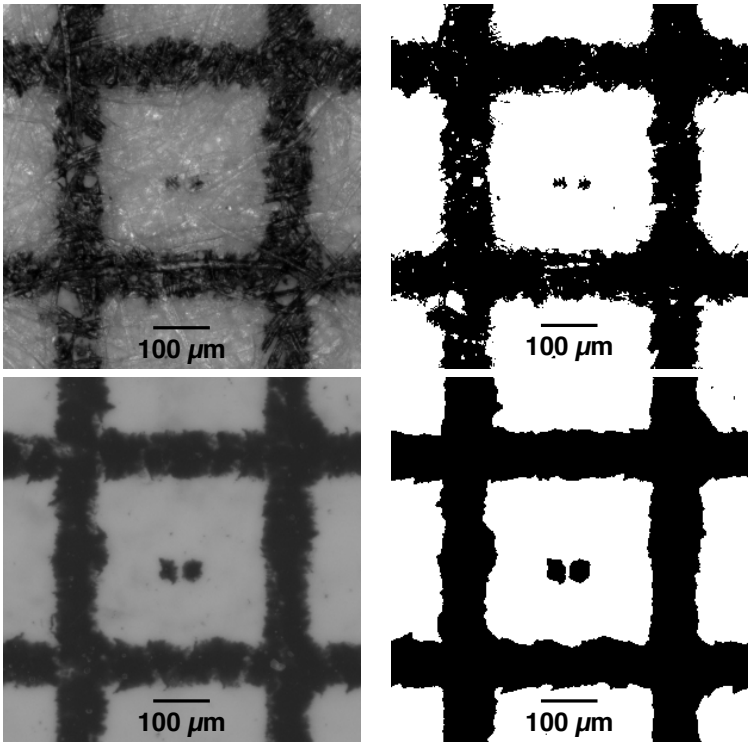

Figure 6: The conventional offset uncoated print (above) and waterless offset coated print (below) are binarised by maximum entropy method.

Wilcoxon nonparametric test [14] is more appropriate than t-test. In particular, for each set-up (conventional offset-uncoated paper, conventional offset-coated paper, waterless offset-coated paper, waterless offset-uncoated paper, laser-uncoated paper, laser-coated paper), we compare the similarity of different printing directions for each pattern. In most of the cases, different printing direction leads to different mean values, p-value < 0.01. However, there are some cases concluding that the samples are statistically from the same distribution. For example, the first and second patterns under the set-up conventional-offset-uncoated-paper, the indexes are statistically from the same distribution. It is similar in the case of waterless-offset-uncoated-paper. Therefore, the effect of the direction of printing is not reliable enough to be considered in source printer identification.

In addition, it is observed that the prints from waterless technology have statistically larger area than the prints from conventional technology. This is due to fundamental difference between the two processes as explained in Section 2.1. In Tables 7, 8, 9, 10, 11, 12, each image illustrates the frequency of appearance of black pixels at each position (profile), each profile is from 100 captured samples. As observed, the black pixels on coated paper are denser than on the uncoated one, this physical phenomenon is explained in Section 2.1.

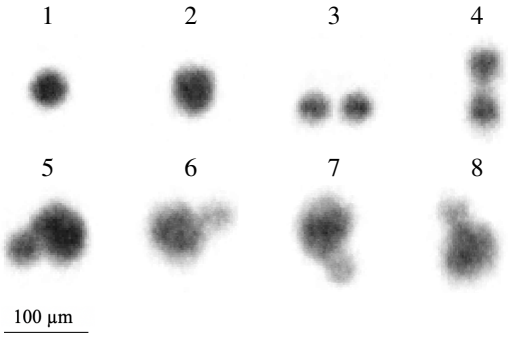

Figure 7: Observed patterns printed with conventional offset on uncoated paper, resolution 1200 dpi.

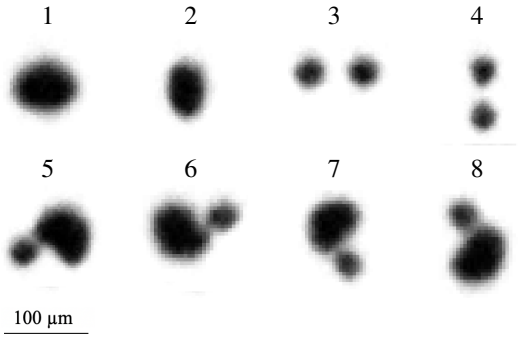

Figure 8: Observed patterns printed with conventional offset on coated paper, resolution 1200 dpi.

	pattern 1	pattern 2	pattern 3	pattern 4	pattern 5	pattern 6	pattern 7	pattern 8
				Conventional offset printing on uncoated paper				
Area (px^2)	229.35 (**61.08**)	214.78 (**54.89**)	178.59 (**56.99**)	186.80 (**67.15**)	456.43 (**100.10**)	552.30 (**171.33**)	596.64 (**156.83**)	572.15 (
Perimeter (px)	152.33 (**29.28**)	148.10 (**26.33**)	169.44 (**37.16**)	181.17 (**36.10**)	276.45 (**49.31**)	364.01 (**96.94**)	370.04 (**84.22**)	378.75 (
Solidity	0.63 (**0.16**)	0.63 (**0.13**)	0.37 (**0.14**)	0.43 (**0.14**)	0.59 (**0.12**)	0.44 (**0.11**)	0.49 (**0.10**)	0.46 (**0.12**)
Convexity	0.65 (**0.18**)	0.63 (**0.21**)	0.73 (**0.26**)	0.62 (**0.16**)	0.50 (**0.10**)	0.51 (**0.15**)	0.48 (**0.09**)	0.49 (**0.19**)
Circulation	0.36 (**0.08**)	0.36 (**0.07**)	0.28 (**0.05**)	0.27 (**0.04**)	0.28 (**0.05**)	0.24 (**0.05**)	0.24 (**0.04**)	0.23 (**0.05**)
				Conventional offset printing on coated paper				
Area	214.39 (**34.32**)	193.98 (**36.18**)	146.17 (**37.13**)	140.91 (**41.23**)	414.46 (**51.97**)	208.35 (**32.21**)	202.46 (**29.52**)	216.52 (
Perimeter	90.64 (**12.58**)	86.18 (**12.69**)	104.74 (**19.07**)	107.02 (**20.14**)	167.06 (**23.30**)	116.50 (**16.62**)	116.96 (**14.76**)	115.02 (
Solidity	0.83 (**0.07**)	0.83 (**0.09**)	0.45 (**0.07**)	0.49 (**0.08**)	0.69 (**0.07**)	0.69 (**0.08**)	0.70 (**0.06**)	0.70 (**0.08**)
Convexity	0.84 (**0.08**)	0.85 (**0.10**)	0.86 (**0.15**)	0.82 (**0.14**)	0.70 (**0.08**)	0.73 (**0.09**)	0.70 (**0.08**)	0.74 (**0.09**)
Circulation	0.58 (**0.08**)	0.58 (**0.08**)	0.41 (**0.05**)	0.39 (**0.05**)	0.44 (**0.07**)	0.45 (**0.07**)	0.44 (**0.06**)	0.46 (**0.06**)
				Waterless offset printing on uncoated paper				
Area	327.15 (**76.97**)	338.26 (**68.38**)	349.00 (**79.83**)	389.27 (**69.06**)	615.02 (**87.76**)	1273.69 (**185.94**)	1321.02 (**213.58**)	1348.35 (
Perimeter	138.64 (**29.06**)	142.04 (**26.27**)	189.58 (**34.76**)	198.68 (**31.49**)	240.58 (**42.31**)	422.38 (**81.12**)	422.20 (**80.47**)	442.68 (
Solidity	0.78 (**0.12**)	0.79 (**0.10**)	0.60 (**0.10**)	0.66 (**0.09**)	0.75 (**0.08**)	0.71 (**0.08**)	0.73 (**0.08**)	0.71 (**0.09**)
Convexity	0.71 (**0.14**)	0.70 (**0.12**)	0.62 (**0.12**)	0.60 (**0.09**)	0.57 (**0.10**)	0.49 (**0.09**)	0.49 (**0.09**)	0.48 (**0.10**)
Circulation	0.47 (**0.10**)	0.47 (**0.09**)	0.35 (**0.06**)	0.36 (**0.06**)	0.38 (**0.08**)	0.31 (**0.07**)	0.32 (**0.07**)	0.31 (**0.08**)
				Waterless offset printing on coated paper				
Area	518.63 (**47.29**)	508.31 (**67.82**)	581.58 (**49.60**)	634.39 (**54.73**)	899.94 (**98.13**)	1800.33 (**139.23**)	1793.46 (**152.01**)	1813.05 (
Perimeter	120.44 (**12.14**)	126.18 (**18.34**)	165.55 (**17.28**)	166.47 (**18.42**)	175.04 (**25.97**)	244.72 (**19.22**)	242.91 (**19.58**)	250.37 (
Solidity	0.75 (**0.20**)	0.61 (**0.25**)	0.57 (**0.20**)	0.51 (**0.20**)	0.68 (**0.19**)	0.82 (**0.13**)	0.79 (**0.14**)	0.77 (**0.17**)
Convexity	1.13 (**0.26**)	1.29 (**0.30**)	1.06 (**0.27**)	1.16 (**0.24**)	1.08 (**0.21**)	0.92 (**0.13**)	0.98 (**0.18**)	0.97 (**0.19**)
Circulation	0.67 (**0.05**)	0.64 (**0.07**)	0.52 (**0.06**)	0.54 (**0.05**)	0.62 (**0.08**)	0.62 (**0.04**)	0.62 (**0.04**)	0.61 (**0.05**)
				Electrophotography printing on uncoated paper				
Area	917.28 (**111.95**)	997.37 (**101.46**)	750.48 (**141.58**)	678.77 (**128.27**)	2063.29 (**181.08**)	1151.38 (**151.71**)	2061.84 (**152.86**)	2039.07 (
Perimeter	354.11 (**53.29**)	331.16 (**50.42**)	495.48 (**68.83**)	473.75 (**77.92**)	567.37 (**74.71**)	527.55 (**111.41**)	552.16 (**71.09**)	576.54 (
Solidity	0.45 (**0.17**)	0.55 (**0.15**)	0.29 (**0.10**)	0.25 (**0.10**)	0.56 (**0.12**)	0.46 (**0.13**)	0.55 (**0.12**)	0.53 (**0.11**)
Convexity	0.72 (**0.20**)	0.67 (**0.17**)	0.55 (**0.13**)	0.62 (**0.19**)	0.54 (**0.10**)	0.50 (**0.12**)	0.56 (**0.10**)	0.54 (**0.10**)
Circulation	0.31 (**0.05**)	0.35 (**0.06**)	0.20 (**0.03**)	0.20 (**0.04**)	0.29 (**0.04**)	0.24 (**0.07**)	0.30 (**0.04**)	0.28 (**0.04**)
				Electrophotography printing on coated paper				
Area	942.18 (**126.65**)	854.08 (**158.63**)	606.10 (**186.71**)	533.44 (**158.46**)	1668.36 (**214.90**)	1618.35 (**182.10**)	2379.54 (**164.42**)	2364.11 (
Perimeter	297.58 (**66.20**)	311.55 (**64.17**)	378.42 (**77.66**)	387.13 (**83.72**)	527.62 (**77.09**)	504.61 (**81.06**)	446.95 (**54.73**)	439.29 (
Solidity	0.48 (**0.18**)	0.41 (**0.16**)	0.24 (**0.11**)	0.20 (**0.09**)	0.47 (**0.11**)	0.54 (**0.11**)	0.61 (**0.13**)	0.59 (**0.13**)
Convexity	0.84 (**0.23**)	0.84 (**0.22**)	0.75 (**0.22**)	0.75 (**0.20**)	0.58 (**0.11**)	0.55 (**0.10**)	0.69 (**0.12**)	0.73 (**0.11**)
Circulation	0.38 (**0.08**)	0.34 (**0.07**)	0.23 (**0.04**)	0.22 (**0.04**)	0.28 (**0.04**)	0.29 (**0.05**)	0.39 (**0.05**)	0.40 (**0.05**)

Table 3: The mean values and standard deviation (in bold) of shape descriptor indexes of eight patterns for two offset printers and electrophotography on two types of paper - the unit pixel (px) depends on the resolution of the microscope, pixel size = 3.45 $\mu m \times 3.45\ \mu m$.

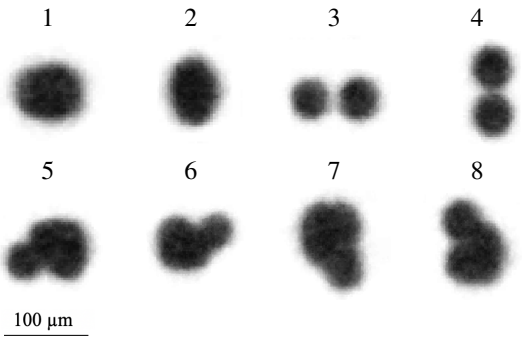

Figure 9: Observed patterns printed with waterless offset on uncoated paper, resolution 1200 dpi.

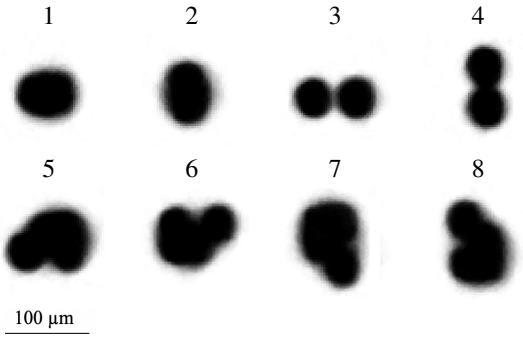

Figure 10: Observed patterns printed with waterless offset on coated paper, resolution 1200 dpi.

4. Printer Source Identifications

In this section, we apply machine learning using shape descriptor indexes to identify printing sources. Particularly, we employ Support Vector Machine and Random Forest algorithms, for the problem of multi-class classification using five aforementioned inputs features (*area, perimeter, solidity, convexity, circularity*).

4.1. Support Vector Machine multi-class for classification

The Support Vector Machine is a supervised learning algorithm applied frequently in classification and regression analysis. Giving a training data set $\{(x_1, y_1), \ldots, (x_n, y_n)\} \in X \times \mathbb{R}$, where X is the vector of input features. In the context of binary classification,

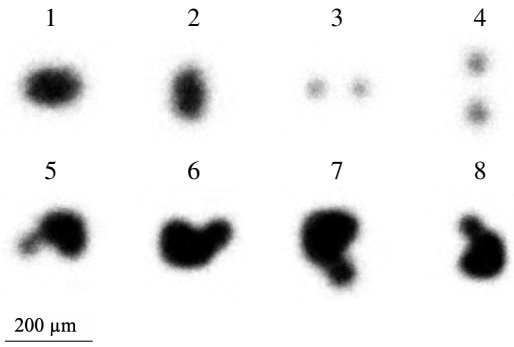

Figure 11: The observed patterns printed with electrophotographic printer on uncoated paper, resolution 600 dpi.

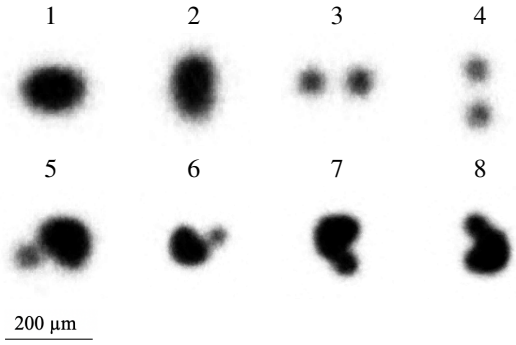

Figure 12: The observed patterns printed with electrophotographic printer on coated paper, resolution 600 dpi.

the outputs belong to $\{-1, 1\}$. The SVM algorithm finds a hyperplane with the maximum margin which separates two classes. For a non-linear SVM, the input data are mapped to higher dimensional space where the discriminating hyperplane can be linearly constructed. For the sake of presentation, a hyperplane is found by solving the

following optimisation problem:

$$\underset{\mathbf{w},\boldsymbol{\xi}}{\text{minimize}}\ \frac{1}{2}\left|\left|\mathbf{w}\right|\right|^2 + C\sum_{i=1}^{n}\xi_i \tag{4}$$

$$\text{subject to}\ \begin{cases} y_i(\langle \mathbf{w}, x_i\rangle + b) & \geq 1, \text{with } b \in \mathbb{R} \\ \xi_i & \geq 0, \end{cases}$$

where $\mathbf{w}$ is the slope of the hyperplane, $\langle ., . \rangle$ denotes the dot product in X. The slack variables ξ_i are introduced for the "soft margin" loss function. The constant $C > 0$ is a penalty factor. In the non-linear problem, a kernel function k replaces the dot product in the high dimensional space. The Gaussian or radical basis function (RBF) kernel is commonly used, $k\left(x_i, x_j\right) = \exp\left(-\gamma||x_i - x_j||^2\right)$, with $x_i, x_j \in X$. The SVM can also be extended to multi-class classification [30]. Particularly, the multi-class classifier is constructed by combining several binary classifiers. In practice, some algorithms were derived including the one-against-rest method, one-against-one method, DAG-SVM, ECOC-SVM, and SVM-BTA. We use the one-against-one method for classification procedure since it is practically suitable with similar performance. Here, LIBSVM [2] is used to implement the SVM classification algorithm in R.

4.2. Random Forest multi-class for classification

One of the most popular and applicable tree-based model for classification is Random Forest, which was discovered on top of the ideology of ensemble learning (known as a divide-and-conquer technique used to restrict the learning biases and push the performance of the prediction to the magnitude which is generally infeasible to common machine model). Basically, RF model is learned by stacking randomly multiple decision tree models on slightly different parts of data, then take advantage of these models' powers to perform the prediction. These sub-samples of the considered dataset are generated via the procedure of bootstrap resampling. Then, we train a decision tree model associated with each sample, and since it is learned only in the particular subset, it is a biased classifier which can capture distinct trends in the data. In the final step, a majority-voting process is applied in order to retrieve the final prediction output as the

highest average class probability of all the individual trees in the forest. We can imagine RF as a committee of experts of different fields and have some certain knowledge and understanding over the target topic (the classification problem) and they can only make decision on the problem based on voting. The package randomForest [31] is used to implement the RF classification algorithm in R.

4.3. Experimental setup

Label	Printing Technology	Printing Paper
I	Conventional Offset	Uncoated
II	Conventional Offset	Coated
III	Waterless Offset	Uncoated
IV	Waterless Offset	Coated
V	HP-600 M620	Uncoated
VI	HP-600 M620	Coated

Table 4: Printers and printing papers.

The proposed experiment is conducted with three printing technologies on two types of printing paper, which mean that there are totally six printing sources. Details are summarised in Table 4. For each pattern in eight patterns, Figure 5, we implement the classification model for six printing sources. For training and evaluating the SVM and RF classification, the ratio between the training dataset and testing dataset is 3:2. In other words, for each pattern, sizes of training and testing dataset are 320 and 240, respectively. For this dataset, we used the radical basis function (RBF) kernel, and the parameters are tuned with the 10-fold cross-validation. The optimal parameters are derived as $C = 10, \gamma = 0.1$. With the RF algorithm, the hyper-parameter is also tuned with the 10-fold cross-validation, the chosen value is ntree = 500. In order to evaluate the performance of the classifier, a commonly used metric is F_1 score, which is based on the Precision P and Recall R, the confusion matrices as defined in Table 5. The

formulas for P, R, and F_1 for each printing source are defined as:

$$P = \frac{TP}{TP+FP}, \tag{5}$$

$$R = \frac{TP}{TP+FN}, \tag{6}$$

$$F_1 = \frac{2}{1/P+1/R}, \tag{7}$$

where TP is denoted for the number of prints correctly assigned to the positive class, FP represents the number of prints incorrectly assigned to the positive class, and FN is the number of objects wrongly assigned to the negative class. To evaluate the classifier generally, we simply take the total accuracy A_{cc} and the harmonic F_h score as follow:

$$A_{cc} = \frac{\sum_{i=1}^{N_p} TP_i}{\sum_{i=1}^{N_p}(TP_i+FP_i)}, \tag{8}$$

$$F_h = \frac{1}{N_p}\sum_{i=1}^{N_p} F_{1,i}\,, \tag{9}$$

respectively, in which $F_{1,i}$ is the F_1 score of i-th printing source and N_p is the number of printing sources.

	predicted positives	predicted negatives
real positives	True positive (TP)	False negative (FN)
real negatives	False positive (FP)	True negative (TN)

Table 5: Confusion matrix.

4.4. Experimental results

We respectively report the confusion matrices corresponding to the classification performance of eight patterns using SVM, Tables 6 - 13, and Random Forest, Tables 14 - 21. In each table, "O" represents the predicted output of printing sources, while "A" is stand for the actual printing sources. As observed, in most of patterns, the printing source 4, Waterless Offset and Coated paper, shows a very good performance, better than the rest; while we observe the confusion matrices, many errors in the classification between printing source 6, HP-600 M620 electrophotographic printer and Coated

paper, and printing source 5, HP-600 M620 electrophotographic printer and Uncoated paper.

The overall report of classifier's accuracy A_{cc} and the harmonic mean F_h are shown in Table 22. The results indicate that the RF outperforms slightly the SVM in every case, up to 95.4%, compared to SVM with the accuracy and F1-scores are up to 94.2% In addition, as we can see, the patterns 5 to 8 (underlined numbers), especially for the patterns 6 to 8 (bold numbers), provides a very good classification, while the prediction results of patterns 1 to 4 are worse, in both classification algorithms. In other words, the more complex printing patterns we use (see Figure 5), the better source identification performance we can achieve.

PLEASE INSERT TABLES 6-22 HERE

4.5. Discussion

We can observe that, in this experiment, there are actually two groups of patterns, including "simple patterns", 1 to 4, and "complex patterns", 5 to 8, and the robustness of each group is clearly described in Table 22. This means that it provides a flexibility in designing the printing patterns with similar performances. For example, looking into the "complex patterns", 5 to 8, in Figure 5, they are simply the rotation of one pattern. Comparing the performance between the SVM and the RF, the latter algorithm is slightly better than the former with all patterns. Yet, as pointed out in [20], the SVM has a substantial better computational cost than the RF, which is more appropriate to implement in the portable devices. In addition, most of the errors observed in the confusion matrices are from electrophotography printing with the two kind of papers.

The experiment indicates that a good design of printing patterns and printing source can be developed for authentication application, especially for forgery detection of printed documents. The size of this dataset is relatively small and the pattern form is simple, yet the classification algorithm still performs well. Moreover, thanks to the simplicity and efficiency of SVM, the classification model is light and easy to implement in commercial production, hence our proposed study is expected to develop a lightweight model which can be integrated to IoT devices, smart-phone, etc., for print authentication.

5. Concluding remarks

The study performs an analysis on microscopic printing with three common printing technologies, conventional offset, waterless offset, and electrophotography printing, on two types of substrate, uncoated and coated paper. Different patterns printed under microscopic scale with various printing directions are statistically analysed, showing an insignificant effect of printing direction on shape descriptor indexes of the prints. The analysis from the simple patterns to the complex ones shows an ability to well perform authentication using microscopic printing.

Specifically, we perform the source printing identification on the combination of source printers and printing papers. We employ SVM and RF for multi-class classification on five shape descriptor indexes of the micro-printing patterns. Experimental results and corresponding analyses, carried out by considering eight printing patterns, from simple patterns to complex ones, show the similar performance of both SVM and RF in multi-class classification. Particularly, the RF gives the better classification in all cases with the accuracy and F1-scores up to 95.4%, while the results from SVM are up to 94.2%. However, the SVM shows a significant advantage in computational cost, which is a benefit in practical applications. Moreover, under microscopic scale, the study indicates that the complex geometric patterns improve notably the performance of the printing source authentication.

Finally, thanks to the lightweight of SVM, it is practical to implement in the IoT or portable devices for printing source tracking. In future study, more different types of printer as well as printing material should be considered. A further research on the pattern structure should be addressed to find the optimal way to generate pattern automatically. Moreover, we could consider an approach using deep learning such as Conventional Neural Network for feature extraction directly from the colour images.

References

[1] Ali, G. N., Mikkilineni, A. K., Allebach, J. P., Delp, E. J., Chiang, P.-J., Chiu, G. T., 2003. Intrinsic and extrinsic signatures for information hiding and secure

printing with electrophotographic devices. In: NIP & Digital Fabrication Conference. Vol. 2003. Society for Imaging Science and Technology, pp. 511–515.

[2] Chang, C.-C., Lin, C.-J., 2011. Libsvm: A library for support vector machines. ACM transactions on intelligent systems and technology (TIST) 2 (3), 1–27.

[3] Chiang, P.-J., Khanna, N., Mikkilineni, A. K., Segovia, M. V. O., Suh, S., Allebach, J. P., Chiu, G. T.-C., Delp, E. J., 2009. Printer and scanner forensics. Signal Processing Magazine, IEEE 26 (2), 72–83.

[4] Darwish, S. M., ELgohary, H. M., 2021. Building an expert system for printer forensics: A new printer identification model based on niching genetic algorithm. Expert Systems 38 (2), e12624.

[5] Ferreira, A., Bondi, L., Baroffio, L., Bestagini, P., Huang, J., Dos Santos, J. A., Tubaro, S., Rocha, A., 2017. Data-driven feature characterization techniques for laser printer attribution. IEEE Transactions on Information Forensics and Security 12 (8), 1860–1873.

[6] Joshi, S., Khanna, N., 2017. Single classifier-based passive system for source printer classification using local texture features. IEEE Transactions on Information Forensics and Security 13 (7), 1603–1614.

[7] Kapur, J. N., Sahoo, P. K., Wong, A. K., 1985. A new method for gray-level picture thresholding using the entropy of the histogram. Computer vision, graphics, and image processing 29 (3), 273–285.

[8] Khanna, N., Mikkilineni, A. K., Chiu, G. T. C., Allebach, J. P., Delp, E. J., 2008. Survey of scanner and printer forensics at purdue university. In: Srihari, S. N., Franke, K. (Eds.), International Workshop on Computational Forensics. Springer, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 22–34.

[9] Kipphan, H., 2001. Handbook of print media: technologies and production methods. Springer Science & Business Media.

[10] Kittler, J., Illingworth, J., 1986. Minimum error thresholding. Pattern recognition 19 (1), 41–47.

[11] Krishnan, R., Yamat, M. C., Babij, H., Palmatier, R. T., Murray, R. R., Jul. 14 1998. Offset lithographic printing process with a water based ink. US Patent 5,778,789.

[12] Li, C., Tam, P. K.-S., 1998. An iterative algorithm for minimum cross entropy thresholding. Pattern recognition letters 19 (8), 771–776.

[13] Mai Hoang, B. A., Sawaya, W., Bas, P., 2014. Image model and printed document authentication: A theoretical analysis. In: 2014 IEEE International Conference on Image Processing (ICIP). IEEE, pp. 5377–5381.

[14] Mann, H. B., Whitney, D. R., 1947. On a test of whether one of two random variables is stochastically larger than the other. The annals of mathematical statistics, 50–60.

[15] Mikkilineni, A. K., Ali, G. N., Chiang, P.-J., Chiu, G. T., Allebach, J. P., Delp, E. J., 2004. Signature-embedding in printed documents for security and forensic applications. In: Electronic Imaging 2004. International Society for Optics and Photonics, pp. 455–466.

[16] Nguyen, Q. T., Delignon, Y., Chagas, L., Septier, F., 2014. Printer identification from micro-metric scale printing. In: Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on. IEEE, pp. 6236–6239.

[17] Nguyen, Q.-T., Delignon, Y., Chagas, L., Septier, F., 2014. Printer technology authentication from micrometric scan of a single printed dot. In: Alattar, A. M., Memon, N. D., Heitzenrater, C. D. (Eds.), Media Watermarking, Security, and Forensics 2014. Vol. 9028. International Society for Optics and Photonics, SPIE, pp. 306 – 312.
URL https://doi.org/10.1117/12.2039989

[18] Nguyen, Q. T., Delignon, Y., Chagas, L., Septier, F., 2015. Modélisation de points imprimés à l'échelle micro-métrique. In: XXVème Colloque GRETSI.

[19] Nguyen, Q. T., Delignon, Y., Septier, F., Phan-Ho, A. T., 2018. Probabilistic modelling of printed dots at the microscopic scale. Signal Processing: Image Communication 62, 129–138.
URL https://www.sciencedirect.com/science/article/pii/S0923596518300249

[20] Nguyen, Q. T., Fouchereau, R., Frenod, E., Gerard, C., Sincholle, V., 2020. Comparison of forecast models of production of dairy cows combining animal and diet parameters. Computers and Electronics in Agriculture 170, 105258.

[21] Norris, M., Barney Smith, E. H., 2004. Printer modeling for document imaging. In: Proceedings of the 2004 International Conference on Imaging Science, Systems, and Technology (CISST'04). pp. 1–7.

[22] Oliver, J., Chen, J., 2002. Use of signature analysis to discriminate digital printing technologies. In: NIP & Digital Fabrication Conference. Vol. 2002. Society for Imaging Science and Technology, pp. 218–222.

[23] Olson, E., 2011. Particle shape factors and their use in image analysis-part 1: Theory. Journal of GXP Compliance 15 (3), 85.

[24] Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics 9 (1), 62–66.

[25] Peng, F., Yang, J., Lin, Z.-X., Long, M., 2019. Source identification of 3d printed objects based on inherent equipment distortion. Computers & Security 82, 173–183.

[26] Peng, F., Yang, J., Long, M., 2018. 3-d printed object authentication based on printing noise and digital signature. IEEE Transactions on Reliability 68 (1), 342–353.

[27] Phan Ho, A., Mai Hoang, B., Sawaya, W., Bas, P., 2014. Authentication using graphical codes: Optimisation of the print and scan channels. In: 2014 22nd European Signal Processing Conference (EUSIPCO). IEEE, pp. 800–804.

[28] Phan Ho, A. T., Mai, B. A. H., Sawaya, W., Bas, P., 2015. Authentification par codes graphiques: optimisation du canal d'impression-acquisition. In: XXVème Colloque GRETSI.

[29] Phan Ho, A. T., Mai Hoang, B. A., Sawaya, W., Bas, P., 2014. Document authentication using graphical codes: Reliable performance analysis and channel optimization. EURASIP Journal on Information Security 2014 (1), 9.

[30] Qian, H., Mao, Y., Xiang, W., Wang, Z., 2010. Recognition of human activities using svm multi-class classifier. Pattern Recognition Letters 31 (2), 100–111.

[31] RColorBrewer, S., Liaw, M. A., 2018. Package randomforest. University of California, Berkeley: Berkeley, CA, USA.

[32] Richter, T., Escher, S., Schönfeld, D., Strufe, T., 2018. Forensic analysis and anonymisation of printed documents. In: Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security. pp. 127–138.

[33] Ryu, S.-J., Lee, H.-Y., Im, D.-H., Choi, J.-H., Lee, H.-K., 2010. Electrophotographic printer identification by halftone texture analysis. In: Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on. IEEE, pp. 1846–1849.

[34] Shang, S., Memon, N., Kong, X., 2014. Detecting documents forged by printing and copying. EURASIP Journal on Advances in Signal Processing 2014 (1), 140.

[35] Sharma, A., Subramanian, L., Brewer, E. A., 2011. Paperspeckle: microscopic fingerprinting of paper. In: Proceedings of the 18th ACM conference on Computer and communications security. ACM, pp. 99–110.

[36] Tsai, M., Hsu, C., Yin, J., Yuadi, I., 2016. Digital forensics for printed character source identification. In: 2016 IEEE International Conference on Multimedia and Expo (ICME). pp. 1–6.

[37] Tsai, M.-J., Hsu, C.-L., Yin, J.-S., Yuadi, I., 2015. Japanese character based printed source identification. In: Circuits and Systems (ISCAS), 2015 IEEE International Symposium on. IEEE, pp. 2800–2803.

[38] Tsai, M.-J., Tao, Y.-H., Yuadi, I., 2019. Deep learning for printed document source identification. Signal Processing: Image Communication 70, 184–198.

[39] Tsai, M.-J., Yuadi, I., 2017. Digital forensics of microscopic images for printed source identification. Multimedia Tools and Applications, 1–30.

[40] Ulichney, R. A., 1999. Review of halftoning techniques. In: Electronic Imaging. International Society for Optics and Photonics, pp. 378–391.

[41] Vallat-Evrard, L., Chagas, L., Passas, R., Reverdy-Bruas, N., 2019. High dynamic range capture and sensor calibration to improve microscale halftone ink and paper surface quantification. Review of Scientific Instruments 90 (8), 083706.

[42] Voloshynovskiy, S., Diephuis, M., Beekhof, F., Koval, O., Keel, B., 2012. Towards reproducible results in authentication based on physical non-cloneable functions: The forensic authentication microstructure optical set (famos). In: Information Forensics and Security (WIFS), 2012 IEEE International Workshop on. IEEE, pp. 43–48.

[43] Zhu, B., Wu, J., Kankanhalli, M. S., 2003. Print signatures for document authentication. In: Proceedings of the 10th ACM conference on Computer and communications security. ACM, pp. 145–154.

O/A	I	II	III	IV	V	VI	P
I	31	2	4	0	0	0	0.838
II	5	38	5	0	0	0	0.792
III	4	0	31	0	0	0	0.886
IV	0	0	0	40	0	1	0.976
V	0	0	0	0	34	16	0.680
VI	0	0	0	0	6	23	0.793
R	0.775	0.95	0.775	1.00	0.85	0.575	

Table 6: SVM Classification results with pattern 1.

O/A	I	II	III	IV	V	VI	P
I	30	3	8	0	0	0	0.732
II	3	37	2	0	0	0	0.881
III	7	0	29	0	0	0	0.806
IV	0	0	1	40	0	1	0.952
V	0	0	0	0	30	8	0.790
VI	0	0	0	0	10	31	0.756
R	0.75	0.925	0.725	1.00	0.75	0.775	

Table 7: SVM Classification results with pattern 2.

O/A	I	II	III	IV	V	VI	P
I	37	3	3	0	0	2	0.822
II	1	37	2	0	0	0	0.925
III	2	0	33	0	0	0	0.943
IV	0	0	2	40	0	2	0.909
V	0	0	0	0	26	6	0.813
VI	0	0	0	0	14	30	0.682
R	0.925	0.925	0.825	1.00	0.65	0.75	

Table 8: SVM Classification results with pattern 3.

O/A	I	II	III	IV	V	VI	P
I	35	5	3	0	0	0	0.814
II	3	35	0	0	0	0	0.921
III	1	0	35	0	0	0	0.972
IV	0	0	2	40	0	0	0.952
V	0	0	0	0	32	15	0.681
VI	1	0	0	0	8	25	0.735
R	0.875	0.875	0.875	1.00	0.80	0.625	

Table 9: SVM Classification results with pattern 4.

O/A	I	II	III	IV	V	VI	P
I	32	1	4	2	0	0	0.821
II	2	39	3	0	0	0	0.886
III	6	0	32	0	0	0	0.842
IV	0	0	1	38	0	0	0.974
V	0	0	0	0	33	6	0.846
VI	0	0	0	0	7	34	0.829
R	0.80	0.975	0.80	0.95	0.825	0.85	

Table 10: SVM Classification results with pattern 5.

O/A	I	II	III	IV	V	VI	P
I	40	0	0	0	1	0	0.976
II	0	40	0	0	0	0	1.00
III	0	0	35	0	1	0	0.972
IV	0	0	0	40	0	0	1.00
V	0	0	5	0	35	4	0.80
VI	0	0	0	0	3	36	0.923
R	1.00	1.00	0.875	1.00	0.875	0.90	

Table 11: SVM Classification results with pattern 6.

O/A	I	II	III	IV	V	VI	P
I	40	1	6	0	0	0	0.851
II	0	39	0	0	0	0	1.00
III	0	0	33	0	0	0	1.00
IV	0	0	1	40	0	0	0.976
V	0	0	0	0	35	5	0.875
VI	0	0	0	0	5	35	0.875
R	1.00	0.975	0.825	1.00	0.875	0.875	

Table 12: SVM Classification results with pattern 7.

O/A	I	II	III	IV	V	VI	P
I	39	1	2	0	0	0	0.929
II	1	39	0	0	0	0	0.975
III	0	0	37	0	1	0	0.974
IV	0	0	1	39	0	0	0.975
V	0	0	0	0	34	2	0.944
VI	0	0	0	1	5	38	0.864
R	0.975	0.975	0.925	0.975	0.85	0.95	

Table 13: SVM Classification results with pattern 8.

O/A	I	II	III	IV	V	VI	P
I	31	0	9	0	0	0	0.775
II	5	36	2	0	0	0	0.837
III	4	4	29	1	0	0	0.763
IV	0	0	0	39	0	0	1.00
V	0	0	0	0	30	13	0.70
VI	0	0	0	0	10	27	0.73
R	0.775	0.90	0.725	0.975	0.75	0.675	

Table 14: Random forest classification results with pattern 1.

O/A	I	II	III	IV	V	VI	P
I	31	3	11	0	0	0	0.689
II	3	37	2	0	0	0	0.881
III	5	0	26	1	0	0	0.8125
IV	1	0	1	39	0	1	0.929
V	0	0	0	0	32	9	0.769
VI	0	0	0	0	10	30	0.75
R	0.775	0.9250	0.65	0.975	0.75	0.75	

Table 15: Random forest classification results with pattern 2.

O/A	I	II	III	IV	V	VI	P
I	38	2	4	0	0	2	0.8261
II	1	37	1	0	0	0	0.9487
III	1	1	33	0	0	0	0.9429
IV	0	0	2	40	0	1	0.9302
V	0	0	0	0	25	6	0.8065
VI	0	0	0	0	15	31	0.6739
R	0.95	0.925	0.825	1.00	0.625	0.775	

Table 16: Random forest classification results with pattern 3.

O/A	I	II	III	IV	V	VI	P
I	31	3	2	0	0	2	0.816
II	3	37	0	0	0	0	0.925
III	5	0	35	0	0	1	0.854
IV	0	0	3	40	0	1	0.909
V	0	0	0	0	32	16	0.667
VI	1	0	0	0	8	20	0.69
R	0.775	0.925	0.875	1.00	0.8	0.5	

Table 17: Random forest classification results with pattern 4.

O/A	I	II	III	IV	V	VI	P
I	34	1	5	2	0	0	0.81
II	1	39	6	0	0	0	0.848
III	5	0	29	0	0	1	0.829
IV	0	0	0	38	0	0	1.00
V	0	0	0	0	34	6	0.85
VI	0	0	0	0	6	33	0.8462
R	0.85	0.975	0.725	0.95	0.85	0.825	

Table 18: Random forest classification results with pattern 5.

O/A	I	II	III	IV	V	VI	P
I	38	0	0	0	1	0	0.9744
II	0	40	0	0	0	0	1.0000
III	0	0	33	0	1	0	0.9706
IV	0	0	0	40	0	0	1.0000
V	2	0	6	0	36	3	0.7660
VI	0	0	1	0	2	37	0.9250
R	0.95	1.00	0.825	1.00	0.90	0.925	

Table 19: Random forest classification results with pattern 6.

O/A	I	II	III	IV	V	VI	P
I	40	0	5	0	0	0	0.89
II	0	40	0	0	0	0	1.00
III	0	0	35	0	0	0	1.00
IV	0	0	0	40	0	0	1.00
V	0	0	0	0	35	5	0.875
VI	0	0	0	0	5	35	0.875
R	1.00	1.00	0.875	1.00	0.875	0.875	

Table 20: Random forest classification results with pattern 7.

O/A	I	II	III	IV	V	VI	P
I	39	0	1	0	0	0	0.975
II	1	40	0	0	0	0	0.976
III	0	0	38	1	1	0	0.95
IV	0	0	1	38	0	0	0.974
V	0	0	0	0	36	2	0.947
VI	0	0	0	1	3	38	0.905
R	0.975	1.00	0.95	0.95	0.90	0.95	

Table 21: Random forest classification results with pattern 8.

	SVM		Random Forest	
pattern	A_{cc}	F_h	A_{cc}	F_h
1	0.821	0.818	0.80	0.80
2	0.821	0.819	0.804	0.802
3	0.846	0.844	0.850	0.848
4	0.842	0.841	0.813	0.807
5	0.867	0.866	0.863	0.862
6	**0.942**	**0.943**	**0.933**	**0.934**
7	**0.925**	**0.926**	**0.937**	**0.938**
8	**0.942**	**0.942**	**0.954**	**0.954**

Table 22: The overall accuracy A_{cc} and harmonic mean F_h of each printing pattern using SVM and random forest, respectively, for classification.

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.

Credit Author Statement

Quoc-Thông Nguyen: Conceptualization, Methodology, Software, Validation, Data curation, Visualization, Writing - original draft, review & editing. An Mai: Writing - review & editing. Lionel Chagas: Conceptualization, Visualization, Writing - review & editing. Nadège Reverdy-Bruas: Validation, Writing - review & editing.

Orléans, Orléans, France, in 2011, and Ph.D. degree in Statistical Signal Processing from the University Lille 1, Lille, France, in 2015. He was a research fellow Laboratoire de Mathématiques de Bretagne Atlantique, Université de Bretagne Sud, Vannes, France in 2015- 2020. Currently, he is a research fellow at Ecole Nationale Supérieure des Arts et Industries Textiles, Roubaix, France. His areas of interest include printed document forensics, statistical/machine/deep learning in classification and forecasting, data science, statistical process control.

An Mai received M.Sc. degree in Applied Mathematics from the University of Orléans, Orléans, France, in 2010, and Ph.D degree in Computer Science, majoring Statistical Signal Processing at Ecole Centrale Lille in 2014. He is currently Lecturer at the International University, Ho Chi Minh city, Vietnam. He is responsible for all the tasks of seeding and growing the future data scientists from students, doing partnership and carrying out R&D with tech industries to bring the scientific research to real-life applications. Recently, he focuses on machine learning application in different areas of image forensics, quantitative trading, fraud detection, predictive maintenance.

Lionel Chagas obtained his engineering degree from Pagora in 1989 and his doctorate in process engineering from the National Polytechnic Institute of Grenoble, France, in 1997. He is currently a research engineer at the International School of Paper, Media prints and biomaterials from the Université Grenoble Alpes in Grenoble, France. His research interests are focused on colorimetry, color management and security printing. He is responsible for the engineering option of printed communication at the Pagora engineering school.

Nadège Reverdy-Bruas received her Master in Mechanics of geophysical environments from the University Joseph Fourier, Grenoble, France in 1996 and the PhD degree in Process Engineering from the National Polytechnical Institute of Grenoble, France, in 2002. She is, currently, Associate Professor in the International School of Paper, Print Media and Biomaterials from University Grenoble Alpes in Grenoble, France. She obtained her degree 'Habilitation Diriger des Recherches' in 2016. Her research interests are focused on printed electronics, paper-ink interactions and, since a few years on plastronic innovations. In this context, she holds a Chair of Excellence dedicated to printed and molded electronics with Schneider Electric acting as a spon-

sor. The aim is to develop processes to design and manufacture electronic devices, adapted to IoT, on 3D thermoplastics by additive technics.