

# Identification of Fake Printed Medicine Packaging from a SVM Approach and Dots Shape Features

Indrama Das  
Project PackMark, Printing  
Engineering Department  
Jadavpur University Kolkata, India

Swati Bandyopadhyay  
Associate Professor Printing  
Engineering Department  
Jadavpur University Kolkata, India

Alain Trémeau  
Professor, Faculty of Sciences  
and Technologies  
Université de Lyon, Université  
Jean Monnet,  
CNRS, Hubert Curien Laboratory,  
Saint Etienne, France

## ABSTRACT

Medicine package printing becomes critical nowadays as fake medicines can be easily packed and sold if printed with the proper color combination. Hence, it is important to identify whether a printed package is printed by original medicine manufacturers or their authorized printers or it is printed by counterfeiters. Scanning or photographing the package and reprinting it is one approach for forging an authentic package sample. In this work, the microscopic examination of printed foils has been carried out to verify whether it is original or not. Blister foil is widely used in the pharmaceutical packaging sector and hence it is chosen as the substrate. The microscopic dot pattern can be viewed as a distinct signature of printing processes. In this study, the reference target chart IT 8.7/3 is printed using three different gravure printers (P1, P2, P3). The images of the print samples are then captured using camera. The images are printed again in the same three printers and then the samples are named as reprint samples (R1, R2, R3). Different shape descriptors index parameters such as dot area, perimeter, circularity, eccentricity, solidity, major and minor axis of dots have been used for classification between the print sample and scanned reprint sample. The study demonstrated that print and reprint dot shape descriptor index parameters might be utilized to distinguish at a microscopic scale. A multi-classification-based method Support Vector Machine have been utilized using shape descriptor index features for print and reprint source identification. The suggested method successfully classified the print and reprint samples at different dot percentages with a high rate of accuracy.

## General Terms

Identification, Counterfeiting, Medicine Packaging, Authentication.

## Keywords

Anti-counterfeiting, microscopic print analysis, dot shape, eccentricity, Support Vector Machine.

## 1. INTRODUCTION

In recent years, research on printing technology has aided in the effort against counterfeiting, a problem that threatens public safety and has an effect on the pharmaceutical packaging industries' bottom lines. An essential stage in the examination of alleged or suspected counterfeit pharmaceutical items is the identification of the packaging. It is possible to see the growth of fake package prints with a better look due to the development of printing technology. The original print may be counterfeited utilizing high-quality printers. The relevance of package

printing security has grown in recent years as a result of the accessibility of high-quality printing technology to avoid counterfeiting of pharmaceutical packaging, which is a great social problem.

This study will aid in detecting fake pharmaceutical packaging printing. Blister foil is a commonly used substrate in the pharmaceutical packaging sector due to its inertness, impermeability, and other qualities. Hence it is chosen as substrate for gravure printing technique in this study. A little amount of research has been done on the security concerns with the gravure printing method. Reproducing the text and images of the packages is the most common way to make counterfeit goods. Since it may be easier to copy the text part, the dot shape features of the images are taken into consideration in this study. The easiest way to counterfeit a package print is to take its image and reprint it with suitable color combination. However, the dot shape parameter values have been changed significantly when the image of the original print has been reprinted after imaging with different smart phones, cameras, or scanners. In this study, it has been seen that differences between the scanned reprint images and the original print dramatically increase when it has been printed even on the same printer and much higher when it has been printed on different printers.

A potential method to distinguish an original print document from a reprint document (which may be simulated as counterfeited) has been demonstrated in the previous [1, 2, 3]. This work has shown that a print may be distinguished from a reprint by analyzing the geometrical form of printed dots at a microscopic scale, much like a fingerprint.

## 1.1. Background

The medicine packaging market is constantly advancing and has experienced annual growth of at least five percent per annum in the past few years. Pharmaceutical counterfeiting is becoming a serious issue both in developed and developing countries. Since many years, the exponential growth of counterfeiting has required constant monitoring and always innovative secure techniques. The counterfeiters use printed material in their criminal activities, such as counterfeit documents or consumer goods used for purposes of identity and recording of transactions. Preventing and discouraging to use unauthorized printed materials is the expected achievement of the authentication methods. Therefore, the package printing is becoming important because its information and quality increases the marketability of the product. In the case of medicine packaging, it plays a primary role, as counterfeiters can sell fake medicine just by reprinting the copy of the original package print. Due to the technological advancements, the prints can be easily modified for malicious purposes. It is a major threat to mankind and pharmaceutical brands.

Counterfeiting is considered an old problem in the medicine, food, and beverage packaging industries. Therefore, the subject of counterfeiting of medicine packaging has become considerably more complex in recent years to distinguish between the genuine and fake medication packaging prints. A state-of-the-art quantitative analysis of global trends in counterfeit and pirated goods by OECD has been reported based on the largest available dataset to date, with an accompanying comprehensive detail [4]. The OECD has stated that 2.5% of the worldwide traded products have been counterfeited ones in 2013. For the European Union (EU), a remarkably higher value of 5% for counterfeited and imported products has been reported. In the case of medical products, counterfeits give rise to an economic loss and are considered moreover a potential threat to the consumer and patient health. The study has been conducted to improve patient safety stipulating an efficient anti-counterfeiting system. It has suggested serialization-based product authentication which uses a unique identifier (a 2D barcode) to monitor and identify each medical packages during transportation through the supply chain. However, the procedure is costly and can be vulnerable to being exploited by counterfeiters. Authentication of physical products such as documents, goods and medicine are generally verified by using the stochastic structure of either the materials that composes the product or of a printed package associated to it [5]. The study has suggested optical detection of random features in combination with digital signatures based on public key codes to recognize counterfeit objects without using expensive production techniques. This method is applied for protecting banknotes and identity verification of credit- or chip-card holders. In recent years, due to the technological advancement and the availability of high-quality printing and scanning devices, the number of counterfeiting or forgeries of documents and product packages has extensively increased. Therefore, the importance of various security elements has been proposed to prevent the images from counterfeiter [6, 7]. The study has described an intelligent system integrating texture features conveyed from the grey level co-occurrence matrix for the Arabic alphabet and niching genetic search for identifying printer source that is applicable for printer's fraud and forgery research in forensic science. The study has shown higher classification accuracy using k-nearest neighbors (KNN). Several studies have focused on classifying the source printer of the printed document using texture analysis [8-13]. Each printed and scanned set of dots (a dot being a binary element) have been affected by stochastic non-invertible noise which generates the difficulties during the reproduction of the original graphical code [14, 15]. These studies have focused on the micrometric scan of prints by a binary response model whose parameters include the location and shape of dots. A maximum likelihood estimation algorithm is used to estimate the location, darkness, scale, and shape parameters of the dot, which are relevant for designing an identification scheme for printer technologies. Furthermore, Nguyen et al. [16] have developed a probabilistic model that consists of vector parameters describing a spatial interaction binary model with an inhomogeneous Markov chain. This study has analyzed how those parameters have determined the location and described the diverse random structures of microscopic printed patterns. Q. Nguyen et al. [17] have presented a statistical analysis of microscopic printing for identifying the authentic printer source using micro-tags consisting of patterns of microscopic printed dots in the paper. This study has applied multi-class Support Vector Machine (SVM) and Random Forest (RF) on five shape descriptor indexes of the micro-printing patterns. The study has claimed that the complex geometric patterns show notably the performance of the printing source

authentication, especially for forgery detection of printed documents using a small dataset.

The use of intrinsic texture features of the packaging material has been reported by R. Schraml et al. [18]. This study has investigated the feasibility of a classification-based drug authentication system based on images of the cardboard packaging and top and bottom blister surface textures. Voloshynovskiy. S. et al. [19] have developed a forensic database FAMOS using microstructure images for authentication purposes.

Much of the previous research has focused on the security issues of the offset or digital prints. Little study has been conducted about the security issues of the gravure printing process. Few studies have focused on the process parameters to improve the gravure print quality [20]. Moreover, most of the studies have been performed on papers. Few works have been done on the blister foil. P. Kundu et. al. [21-23] have focused on the consistency of color and authentication of blister foil print samples printed with gravure printing techniques using spectral reflectance, color differences and ANN model. Also, the blister foil is widely considered as package substrate in the pharmaceutical industry. In this study blister foil has been taken as substrate due to its inertness, light weight and chemical resistance properties. This paper has developed an approach to identifying authentic prints using identifying parameters like Dot area, Perimeter, Circularity, Eccentricity, Solidity, Major and Minor axis of the microscopic printed dots in package prints. The features of these patterns of dots on printed samples and reprinted samples have been extracted and fed into a multi-class Support Vector Machine (SVM) classification model to identify the original print.

There have been pressing needs for effective print authentication technology to fight counterfeiting and forgery. This is because counterfeiting and forgeries are extremely widespread everywhere, causing tremendous damage to common people, industries, and societies. Development of authenticity of printed product from counterfeiting depends highly on the capacity to measure and control the properties of the print. With the improvement of modern digital measurement instruments, accurate measurement of printed dots at the micro-scale is becoming easy. The aim of this study is to distinguish between printed and reprinted dots using shape descriptor index parameters as print features such as dot area, perimeter, circularity, and eccentricity, solidity and major, minor axes of the samples. The reprint is created by the counterfeiter by scanning or capturing the original print and reproducing it. The main objective of this study is to classify the print and reprint dot shape features to determine the characteristics of dot shape size changes using a multi-class Support Vector Machine (SVM) classification model. It could be possible to determine whether a print is original or a replica that has been reproduced by reprinted by observed the certain factors properly.

This paper is organised as follows: Section 2 presents the experimental materials details and the methods of this study. Section 3, the proposed method, and in the Section 4 experimental results for identifying the print sample. Finally, in Section 5, a conclusion is drawn.

## **2. EXPERIMENTAL MATERIALS AND METHODS**

In this study, the gravure printing process has been used to print and reprint samples for the medicine packaging. It is considered as the major printing process used for medicine packaging due to its simplicity, productivity, color consistency and ability to print at a higher speed. The experiment has been carried out on

the blister foil substrate, which is extensively used in the medicine packaging industry. The standard color target chart IT8.7/3 has been taken to print with cyan, magenta, yellow, and black foil inks on the blister foil. For printing purpose, four gravure cylinders have been taken for each of the four colors (C, M, Y, K) and the standard chart (IT 8.7/3) has been engraved on the cylinders by electro-mechanical engraving process. The reference image (IT 8.7/3) has been shown in Fig. 1. The reference image has been engraved on the cylinders (size: 300x534 mm) at 150 LPI screen ruling and 130 stylus angles. Parameters that have been maintained during the printing process are as follows: thickness of the foil: 25  $\mu$ m, doctor's blade angle: 30°, gravure speed: 60 m/min, engraving process: electro-mechanical, pressure of rubber roller: 2.5 kg/cm<sup>2</sup> (for each unit), pressure of doctor's blade: 1 kg/cm<sup>2</sup> (for each unit), drying temperature: 50° C to 60° C, temperature: 17  $\pm$  3° C and humidity: 35  $\pm$  5%.

Using the same cylinders, the standard chart (artwork) has been printed with three different gravure printers named as P1, P2 and P3.

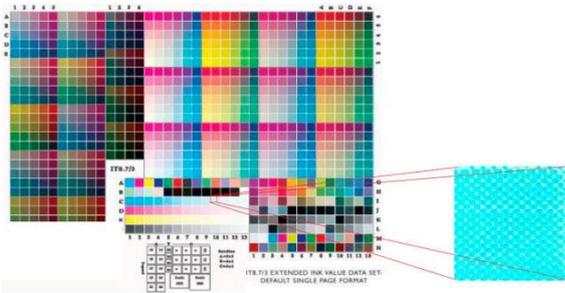
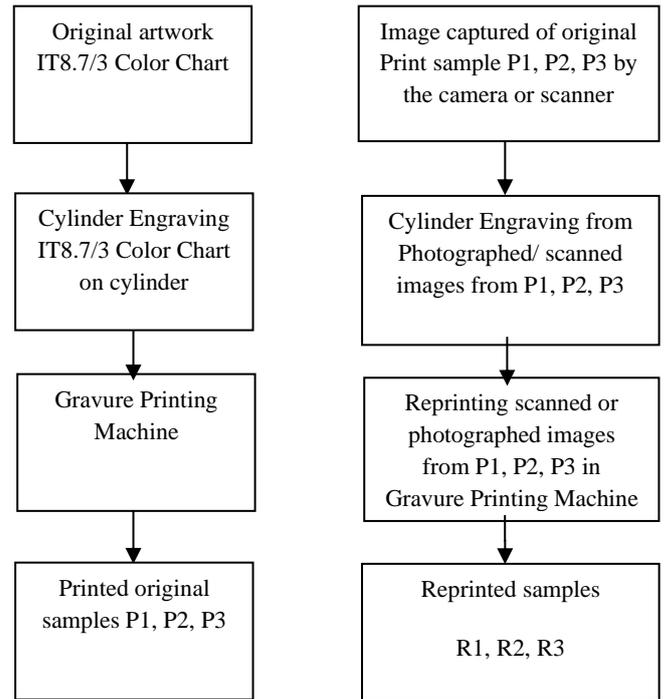


Fig. 1: Artwork IT8.7/3 Color Chart reference image

In this experiment, to simulate the counterfeiting process, the images of the original print samples collected from printers P1, P2, and P3 have been captured and then, new cylinders have been produced for each color with same engraving process and same parameters to print the simulated counterfeit samples denoted as reprint sample. All printing parameters, ink, substrate (blister foil) has been kept remain same to reprint the samples. The reprinting process has been carried out with the same three gravure printers, and the scanned reprint samples denoted as R1, R2 and R3. The simulation of the counterfeiting process has been shown in Fig 2.



(i)Flowchart of Printing Process (ii)Flowchart of Simulated Counterfeiting Process

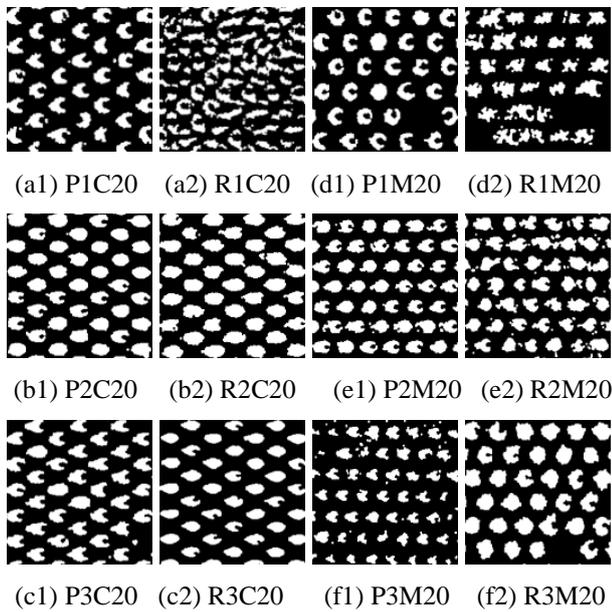
Fig. 2: (i) Original Print Process and (ii) Simulated Counterfeiting Process

## 2.1 Analytical process

The dots of printed (P1, P2, P3) and reprinted samples (R1, R2, R3) have been analyzed using a microscopic camera (model HDCE-X3) attached with Lawrence & Mayo microscope. It has been used to capture each color patch from print samples and reprint samples with 4x zoom magnification (step 1). The microscopic captured images have been processed with the ScopeImage 9.0 microscopy software. Different dot areas of cyan, magenta, yellow and black inks have been tested. In this paper, 20% dot areas of cyan and magenta color tints have been analyzed because it has been observed that the dot bridging and dot kissing increase between the dots when the dots percentage increases. So, with higher dots percentage it becomes difficult to separate the dots' structure properly.

To analyze the microscopic images of print and reprint samples for cyan and magenta tint, each image was first converted to a binary image (step 2), then the canny edge detection algorithm was applied on these binary images (step 3) using MATLAB R2018a generic functions. Edge detection was performed to determine the number of dots in print samples for a given percentage of dots. The edges of the dot sample images were computed from the intensity contrast between the foreground and background pixels. For the next stage of analysis, the contours of the edge dots found in the samples were extracted. The circular Hough transform, which is a modification of the conventional Hough transform, has been used to detect dot contours (step 4). This transform has been used to detect the dots irrespective of their shape. Then, seven dot shape descriptor indexes have been computed (step 5). The shape descriptor indexes of dot patterns have been computed using MATLAB generic functions. The dot descriptor indexes were then used as features vectors to classify the dots in the samples (step 6).

To distinguish between printed and reprinted samples, feature parameters such as dot area, parameter, circularity, eccentricity, solidity, major axis, and minor axis, have been used to classify the printed dots using a multi-class SVM classification model, with the objective to differentiate the printed dots from the reprinted dots. To check the repeatability of this study for statistical analysis, the measurements have been carried out 5 times for each press and 10 samples have been taken for each print and reprint sample. In this paper, only 20% of cyan and magenta color inks have been considered for analysis of print and reprint samples. In this study, Fig 3 shows some samples of dots structures used with 20% of cyan print (represented as P1C20), 20% of magenta print (represented as P1M20) printed using the printing press P1. Similarly, for two other gravure printing presses (P2 and P3) the print samples are represented as P2C20, P2M20 and P3C20, P3M20. The reprint samples are referred as R1C20, R1M20, R2C20, R2M20, R3C20, R3M20, for three printers respectively (R1, R2 and R3). It has been studies for all other dot percentages of print and reprint samples.



**Fig. 3: Binary images with several dot patterns corresponding to 20% of cyan (a, b, c) and magenta (d, e, f) tints (for Print samples - P1(a1, d1), P2(b1, e1), P3(c1, f1), for reprint samples - R1(a2, d2), R2(b2, e2), R3(c2, f2)).**

### 3. CLASSIFICATION PROCESS

This section presents the classification process that has been applied to detect and distinguish authentic print samples from reprint samples based on a machine learning method using several dot shape descriptor indexes. In this work, a Support Vector Machine (SVM) method has been applied as a multi-class classification process to differentiate original print samples from simulated counterfeit (scanned reprint samples), using seven 2D shape features such as area, perimeter, circularity, eccentricity, solidity, major axis and minor axis.

#### 3.1 Support Vector Machine multi-class for classification

A Support Vector Machine (SVM) is a supervised learning model based on the principle of structural risk minimization, as defined by Cortes and Vapnik in 1995 [24]. SVM is a supervised learning algorithm with associated learning algorithms that analyze data for classification and regression

analysis. The objective of the SVM algorithm is to transform input data into a higher dimensional space and to construct an optimal line or decision boundary that can divide a  $n$ -dimensional space into classes, allowing it to quickly classify new samples into the classification process. In order to enable the prediction of labels from one or more feature vectors, it seeks to construct a decision boundary between the output classes. Each hyperplane, corresponding to a decision boundary, is arranged in such a way that it is as far as possible from the nearest data point(s) from each class. Support vectors (SVM) labeled the training dataset explaining as:

$$\{(x_1, y_1), \dots, (x_n, y_n)\} \in X \times \mathbb{R} \quad (1)$$

where  $X$  represents the vector of input features. The output class labels ( $y_i$ ) belong to  $\{-1, 1\}$ . The input data for a non-linear SVM are transferred to a higher-dimensional space where the discriminating hyperplane can be generated linearly. A hyperplane is determined by solving the following optimization problem:

$$\underset{w, \xi}{\text{minimize}} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad (2)$$

with

$$\begin{cases} y_i((w, x_i) + b) \geq 1 \text{ with } b \in \mathbb{R} \\ \xi_i \geq 0 \end{cases} \quad (3)$$

Where:

- $\xi_i$  denotes the distance to the correct margin (find the max-margin classifier that perfectly separates the training data) with  $\xi_i \geq 0, i=1, \dots, n$ .
- $C$  indicates a regularization parameter (controls the tradeoff of how well separated the data).
- $\|w\|^2$  represents the normal vector (hyperplane's slope).
- $x_i$  represents the transformed input space vector (dot product of training data).
- $b$  denotes a bias parameter.
- $y_i$  represents the  $i^{\text{th}}$  target value.

A kernel function might be used to add more dimensions to the raw data in a non-linear situation, making it linear in the resulting higher dimensional space. In the high-dimensional space, a kernel function  $k$  replaces the dot product in the nonlinear problem.

For kernels, like the Gaussian or Radial Basis Function (RBF) kernel ( $K_{\text{RBF}}(x_i, y_j) = \exp(-\gamma \|x_i - y_j\|^2)$ ) with  $x_i, y_j \in X$ , the corresponding feature vector is infinite dimensional. Gamma ( $\gamma$ ), which is determined by cross validation, scales the squared distance and it scales the influence on classification. In SVM the cross-validation consists to randomly split the set of observations into  $C$  number of groups of equal size as per the selected fold for cross-validation, in which the first fold treated as a testing and validation set and the remaining  $C-1$  groups being used to train the model.

#### 3.2 Experimental Setup

In the proposed experiment, two distinct types of samples - original print samples and -reprint (simulated counterfeited) samples have been classified using SVM. The training dataset and test dataset ratio has been set to 3:2 for each percentage of dots for print and reprint samples (i.e., for 20% of dots). Different shape descriptor index parameters such as dot area,

perimeter, circularity, eccentricity, solidity and major and minor axes of dots for different print and reprint options have been measured and used for classification between the print sample and scanned reprint sample. The Gaussian kernel function has been tested using 6-fold cross-validation for this experiment. The optimal parameters for cross-validation, experimentally defined correspond to  $C = 6$ , gamma set to  $\gamma=0.06$  and plug it in the dataset from the observations are relatively close to each other or far to each other. Figs. 4 and 5 represent the model prediction plot of the dots features for gravure printers P1, P2, P3 and for reprints R1, R2, R3, for cyan and magenta tint (20% dots). Then, the confusion matrix has been enabled to compare the real (actual) target values with those predicted values by the SVM model to evaluate the model's performance and prediction errors. For the  $F_1$  score, it has computed the harmonic mean between precision and recall, which both depend on false positives and false negatives values. The efficiency of the classifier was evaluated based on Precision P and Recall R.

The formulas for P, R, and  $F_1$  for each print samples are defined as:

$$P = \frac{TP}{TP+FP} \quad (4)$$

$$R = \frac{TP}{TP+FN} \quad (5)$$

$$F_1 = \frac{2}{1/P+1/R} \quad (6)$$

Where TP denotes the number of prints correctly assigned to the positive class, TN is defined as the number of prints correctly assigned to the negative class. FP represents the number of prints incorrectly assigned to the positive class (which is considered as a Type 1 error), and FN represents the number of prints wrongly assigned to the positive class (i.e a Type 2 error). The accuracy score  $A_{cc}$  corresponds to the ratio of true positives by the sum of true positives and false positives, for all data points. The harmonic score  $F_h$  corresponds to the average value of the  $F_1$  score, as defined below:

$$A_{cc} = \frac{\sum_{i=1}^{N_p} TP_i}{\sum_{i=1}^{N_p} (TP_i+FP_i)} \quad (7)$$

$$F_h = \frac{1}{N_p} \sum_{i=1}^{N_p} F_{1,i} \quad (8)$$

Where,  $F_{1,i}$  is the  $F_1$  score of the  $i^{th}$  print sample, and  $N_p$  is the number of print samples.

#### 4. EXPERIMENTAL RESULTS

This experiment has been done to identify and differentiate an original print from a reprint sample (simulated counterfeit), both printed IT 8.7/3 target chart using a gravure printing on a blister foil. Experiments have been done with different dot areas of cyan, magenta, yellow and black inks. In this paper, 20% cyan and magenta dots are taken in consideration. The dataset has been used for SVM classification in MATLAB using Median Gaussian kernel function with 6-fold cross validation. Figs. 4 (a) and (b) represent the scatter plot of the dots features for gravure printers P1, P2, P3 and for reprints R1, R2, R3, for 20% of cyan. Figs. 4 (c) and 4(d) represent the model predictions of dot features of the corresponding prints and reprints. Similarly, Figs. 5 (a) and (b) represent the scatter plot of the dots features for gravure printers P1, P2, P3 and for reprints R1, R2, R3, for 20% of magenta and Figs. 5 (c) and 5(d) represent the model predictions of dot features of the

corresponding prints and reprints. It has been studied for all other dot percentages of print and reprint samples and similar results have been observed.

The confusion matrices have been created corresponding to the classification performance of dot patterns in print and reprint samples using SVM. Table 1 (a), (b) show the confusion matrix of 20% dots percent of cyan and magenta tints for print and reprint samples. In this table, 'O' denotes the predicted output for print and reprint samples and 'A' represents the actual input samples for print and reprint samples. Table 1(a) shows the confusion matrix for 20% of cyan where the Precision value is 98% and Recall is 90%, which means that the print sample (from P3 printer) has been classified for 20% cyan dots from the rest of the print and reprint samples. In the same way, Table 1(b) shows the confusion matrix for 20% of magenta respectively, for all print and reprint samples. In this study, a similar representation of the confusion matrix has been observed for all dot percentages of cyan and magenta respectively, for all print and reprint samples.

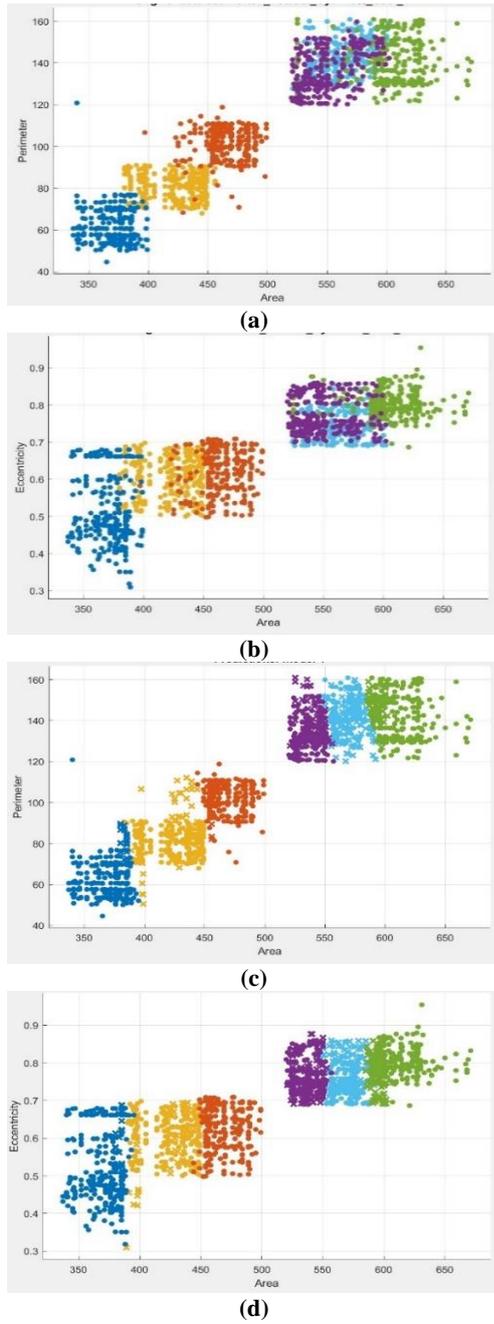
From the confusion matrix for 20% cyan, it has been observed that Type 1 and Type 2 errors have not occurred in the classification between original print samples (P3, P2, P1), whereas both Type 1 and Type 2 errors have occurred in the classification between reprint samples (R1, R2, R3). So, these results indicate that the classification model classified the print samples with higher precision and recall percentages compared to reprint samples. Similar results have been observed for the 20% magenta print and reprint samples.

**Table 1- (a) SVM Classification results for 20% cyan print and reprint**

O/A	P3 C20	P2 C20	P1 C20	R1 C20	R2 C20	R3 C20	Precision
P3 C20	<b>244</b>	9	17				<b>0.98</b>
P2 C20	6	<b>261</b>	3				0.94
P1 C20		9	<b>261</b>				0.93
R1 C20				<b>162</b>	66	42	0.63
R2 C20				49	<b>171</b>	50	0.58
R3 C20				46	60	<b>164</b>	0.64
Recall	0.90	<b>0.97</b>	<b>0.97</b>	0.60	0.63	0.61	

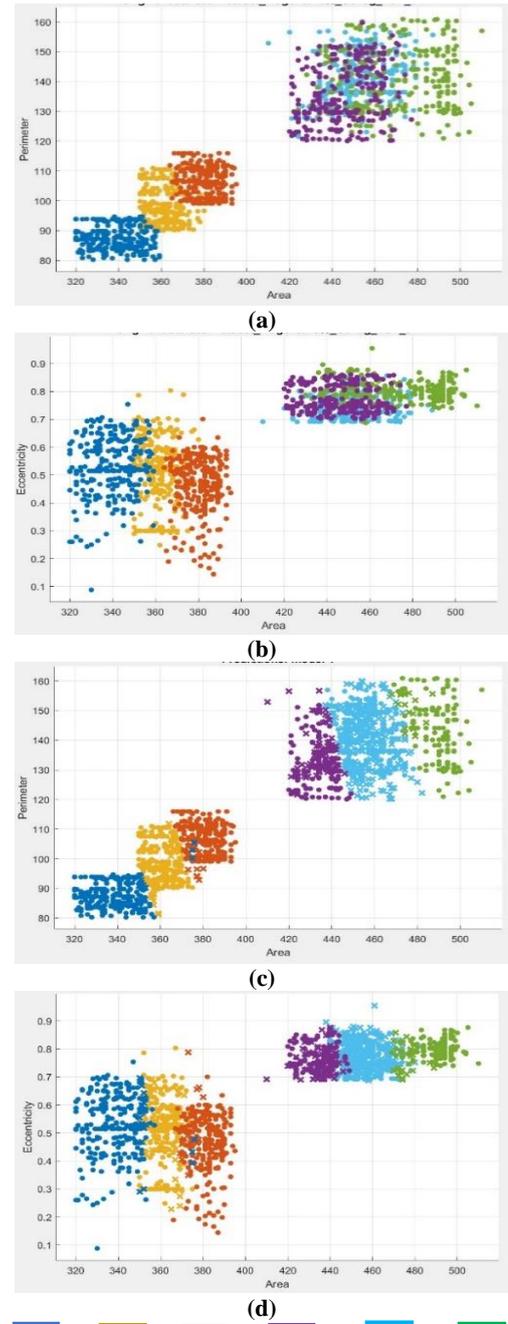
**Table 1- (b) SVM Classification results for 20% magenta print and reprint**

O/A	P3 M20	P2 M20	P1 M20	R1 M20	R2 M20	R3 M20	Precision
P3 M20	<b>259</b>	5	6				0.94
P2 M20	3	<b>258</b>	9				<b>0.97</b>
P1 M20	15	2	<b>253</b>				0.94
R1 M20				<b>220</b>	24	26	0.41
R2 M20				17	<b>224</b>	29	0.45
R3 M20				16	24	<b>230</b>	0.43
Recall	<b>0.96</b>	<b>0.96</b>	0.94	0.40	0.53	0.40	



■ P1 ■ P2 ■ P3 ■ R1 ■ R2 ■ R3

**Fig. 4:** Scatter plots for 20% of Cyan for Print sample (P1, P2, P3) and Reprint sample (R1, R2, R3) (a) Original data Area vs. Perimeter (b) Original data Area vs. Eccentricity, (c) Predicted data Area vs. Perimeter, (d) Predicted data Area vs. Eccentricity.



■ P1 ■ P2 ■ P3 ■ R1 ■ R2 ■ R3

**Fig. 5:** Scatter plots for 20% of Magenta for Print sample (P1, P2, P3) and Reprint sample (R1, R2, R3) (a) Original data Area vs. Perimeter (b) Original data Area vs. Eccentricity, (c) Predicted data Area vs. Perimeter, (d) Predicted data Area vs. Eccentricity.

The classifier's accuracy  $A_{cc}$  and the harmonic mean  $F_h$  have been shown in Table 2 (a), (b). The results indicate that the classification performs well, with a good accuracy and F1-scores for print samples P1, P2, P3, using the SVM model. The print samples provide a very good classification for 20% dots of cyan and magenta. It has also been observed from the Table 2 that the classifier's accuracy  $A_{cc}$  and the harmonic mean  $F_h$  are relatively lower for R1, R2 and R3 reprint samples. It shows similar results for all other dot percentages.

**Table 2(a) – the overall accuracy  $A_{cc}$  and harmonic mean  $F_h$  for each print and reprint using SVM for classification for 20% of Cyan.**

Samples	SVM	
	$A_{cc}$	$F_h$
P3	<b>0.98</b>	0.94
P2	<b>0.98</b>	<b>0.95</b>
P1	<b>0.98</b>	<b>0.95</b>
R1	0.87	0.61
R2	0.86	0.60
R3	0.88	0.62

**Table 2(b) – the overall accuracy  $A_{cc}$  and harmonic mean  $F_h$  for each print and reprint using SVM for classification for 20% of Magenta.**

Samples	SVM	
	$A_{cc}$	$F_h$
P3	0.98	0.95
P2	<b>0.99</b>	<b>0.96</b>
P1	0.98	0.94
R1	0.80	0.39
R2	0.80	0.48
R3	0.81	0.42

#### 4.1 Discussion

In this experimental study, "print" samples (P1, P2, P3) and "reprint" samples (R1, R2, R3) have been categorized based on their dot feature parameters.

Figs. 4 and 5 illustrate that the dot areas and perimeters for reprints are larger than those for prints. The eccentricity feature is also greater for reprints. It may be due to the additional dot gain while copying the print. Specifically, dot gain increases when a scanner or camera captures the original print image. Similar results have been observed for all other dot percentages of tints printed and reprinted, respectively. In, Table 2, the classification accuracy of printed dot parameters also indicates that the dot sizes and shapes for reprints have grown irregularly after imaging (simulated counterfeit).

Table 2 has been demonstrated the accuracy of the classification of printing dot patterns with 20% cyan and magenta color for print samples. The high precision and recall values for classification of P1, P2 and P3 show that the SVM model is quite effective for dot feature classification of prints. Relatively lower accuracy has been obtained with reprints. It shows that the dot sizes and shapes for reprints have increased irregularly after imaging. Table 2(a), (b) show the overall for prints and reprints using SVM classification for 20% cyan, 20% magenta respectively. From the table, it has been seen that there is little variation in accuracy and harmonic mean for different original prints for different dot areas. However, for the reprints, the variation of accuracy and harmonic means for different dot areas are wide apart and inevitable. Hence, SVM classification and its associated parameters may be effectively used for identifying counterfeited documents. Thus, printing sources may be established for authentication applications using this model, particularly for forgery detection of printed documents or medicine packages.

This classification model worked well with the dataset of color print and reprint samples using dot descriptor index parameters as features in the input dataset. The classification model is simple to implement for commercial production due to the simplicity and efficacy of Support vector Machine. It is possible to identify the original print for authentication and thus the original prints may be easily differentiated from the counterfeited ones in different packaging applications.

#### 5. CONCLUSION

In this study, analyzed the shape of microscopic dot structures in prints and reprints printed on blister foil medium, using different color inks and three different gravure machines. The images of original prints have been captured/scanned with a camera/scanner and printed again to simulate the counterfeited printing process and these reproduced samples have been mentioned as reprints. In this paper, experimental results related to 20% of cyan and magenta dots have been shown. This study has also worked on all other dot percentages. The experiments have been performed for yellow and black inks, which also show similar results.

The original prints and the reprints with 20% dot structure, have been statistically analyzed, showing a significant change on dot structure shape descriptor indexes between the prints and the reprints. In the Table 2, it has been clearly observed that print samples are classified very well using a SVM multi-class classification model. Relatively lower accuracy was obtained for reprints; this is caused by an optical dot gain due to the image acquisition of the original print. Since, the counterfeiting of a printed package with color images must involve a stage of imaging through a camera or scanner, the associated optical dot gains are unavoidable whatever may be the quality of the image capture device used.

The analysis of printed dot structure and of reprinted dot structure showed the ability of SVM classification to well perform authentication using microscopic printing. This study demonstrated that it is possible to identify an original print or a reprint sample (simulating a fake print) by comparing the source print and the reprint samples on foil. Here, SVM has been used to classify the seven-shape descriptor index of the micro-printing dot structure patterns into multiple classes. The accuracy and F1-scores of the experimental results of multi-class classification using SVM show that the classification is better in all cases of print samples. It has been observed that F1- scores for print samples can reach significant as 98%. However, the scores for reprint samples are relatively less as 63% in this study. Accuracy can reach large as 99% for print samples and can reach large as 88% for reprint samples.

Hence, SVM classification can be effectively utilized to differentiate between an authentic color image print sample from its counterfeited one as imaging is a necessary and unavoidable tool for counterfeiters which cannot avoid inherent dot gain.

In future work, it may be possible to produce an application that can be integrated into gadgets, smartphones, etc., for authentication of the printed documents or pharmaceutical/food packages.

#### 6. ACKNOWLEDGMENTS

The authors would like to acknowledge the Indo-French Centre for the Promotion of Advanced Research (IFCPAR/CEFIPRA) for funding the project. The authors also acknowledge Industry

Partner Sergusa Solutions Pvt. Ltd. for providing printing facilities to carry this work.

## 7. REFERENCES

- [1] Das I., Bandyopadhyay S., Trémeau A., “Characterization of prints based on microscale image analysis of dot patterns”, *Appl Sci*, 11 (14) (2021), p. 6634.
- [2] Das I., Bandyopadhyay S., Trémeau A., “Authenticity of a print for medicine packaging from its dot sizes and shape”, *International Conference on Innovations in Engineering and Technology (ICIET-2022)*, held on 15-17 September, 2022, proceeding page number 133.
- [3] Das I., Bandyopadhyay S., Trémeau A., "Identification of A Fake Medicine Packaging Print From its Dot Sizes and Shape", *International Journal of Engineering Research & Technology (IJERT)*, IJERTV12IS030199, ISSN: 2278-0181, Vol. 12 Issue 03, March-2023
- [4] OECD and EUIPO, “Trade in counterfeit and pirated goods,” OECD Publishing, p. 138, 2016
- [5] T Haist, HJ Tiziani, “Optical detection of random features for high-security applications”, *Optic. Comm.* 147(1–3), 173–179 (1998).
- [6] 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, 72–83. doi:10.1109/MSP.2008.931082.
- [7] 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, e12624.
- [8] Lee, H., & Choi, J. (2010). “Identifying color laser printer using noisy feature and support vector machine”. In *IEEE International Conference on Ubiquitous Information Technologies and Applications, China* (pp. 1–6).
- [9] Tsai MJ, Liu J (2013) “Digital forensics for printed source identification”. In *IEEE International Symposium on Circuits and Systems (ISCAS)*. May, pp. 2347–2350. doi: 10.1109/ISCAS.2013.6572349.
- [10] Tsai, M.J., Yuadi, I., 2017. “Digital forensics of microscopic images for printed source identification”. *Multimedia Tools and Applications*, 1–30doi:10.1007/s11042-017-4771-1
- [11] Tsai MJ, Yin JS, Yuadi I, Liu J (2014) Digital forensics of printed source identification for Chinese characters. *Multimedia Tools and Applications* 73:2129–2155. doi:10.1007/s11042-013- 1642-2
- [12] Tsai MJ, Hsu CL, Yin JS, Yuadi I (2015) “Japanese character based printed source identification”, *IEEE International Symposium on Circuits and Systems (ISCAS)*. May, Lisbon. pp. 2800-2803. doi: 10.1109/ISCAS.2015.7169268
- [13] 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, 1603–1614.
- [14] Q. T. Nguyen, Y. Delignon, L. Chagas, and F. Septier. “Printer identification from micro-metric scale printing” In *Proc. ICASSP*, pages 6277–6280, 2014.
- [15] Q. T. Nguyen, Y. Delignon, L. Chagas, and F. Septier. “Printer technology authentication from micrometric scan of a single printed dot” In *IS&T/SPIE Electronic Imaging*, pages 1–7, 2014.
- [16] 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.
- [17] Nguyen, Quoc-Thông, et al. "Microscopic printing analysis and application for classification of source printer." *Computers & Security* 108 (2021): 102320.
- [18] R. Schraml, L. Debiassi, C. Kauba, and A. Uhl, "On the feasibility of classification-based product package authentication," in *2017 IEEE Workshop on Information Forensics and Security (WIFS)*. IEEE, 2017, pp. 1–6.
- [19] 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. doi:10.1109/WIFS.2012.6412623.
- [20] Joshi A., Bandyopadhyay S. "Effect of gravure process variables on void area in shrink film". *Journal of Coatings Technology and Research*, v11 P 757-764 2014.
- [21] Paulomi Kundu, Swati Bandyopadhyay, Alain Tremeau “Authentication of a Gravure Printer from Color Values using an Artificial Neural Network”, *International Journal of Engineering Research & Technology (IJERT)* , Vol. 11 ,Issue 02, February-2022.
- [22] Paulomi Kundu, Swati Bandyopadhyay, Alain Tremeau,” Analysis of Spectral differences between Printers to detect the Counterfeit Medicine Packaging”, *Journal of Algebraic Statistics*, Volume 13, No. 2, p. 798 – 806, 2022.
- [23] Paulomi Kundu, Swati Bandyopadhyay, Alain Tremeau,” Study on Color Gamut to Authenticate the Gravure Printers Output”, *International Journal of Computer Applications (0975 – 8887)*, Volume 184– No.18, June 2022.
- [24] Cortes, C., Vapnik, V. “Support-vector networks”, *Machine Learning* 20, 273–297 (1995). <https://doi.org/10.1007/BF00994018>.