Text-Independent Source Identification of Printed Documents using Texture Features and CNN Model

Pushpalata Gonasagi[1](✉), Shivanand S. Rumma[1], and Mallikarjun Hangarge[2]

[1] Department of PG Studies and Research in Computer Science, Gulbarga University, Kalaburagi, Karnataka, India
gonasagi99@gmail.com

[2] Karnatak Arts, Science and Commerce College, Bidar, Karnataka, India

Abstract. Artificial Intelligence (AI) technologies have been used in digital forensic science to resolve disputed documents where one or more human experts would normally be contacted. The purpose of intelligent systems based on printer identification is determined printer created a specific document. Most solutions based on a text-dependent approach may be insufficient in certain scenarios. No study on text-independent based on various word images printed from various laser printer models has been done, as far as the researchers are knowledgeable. As a result, we classify the laser printer models based on the various gray scale word images. 40000-word images of four laser printer models are included in the collection. To classify the different laser printer models, the LBP (Local Binary Pattern) with KNN (K-Nearest Neighbors) and the cubic SVM (Support Vector Machine) classifiers are employed. The deep learning CNN (Convolution Neural Network) model is also used to determine the laser printer models. The experimental results of textural features and the CNN architecture are compared to recent work from a literature survey. We obtained high accuracy from K-NN and cubic SVM classifiers of 97.2% and 97.9%, respectively, and 94.3% accuracy in the CNN model.

Keywords: Printers; Forgery document; LBP · K-NN · SVM · CNN

1 Introduction

We use printed documents for security, instruction, and official work, such as land records, agreements, bills, etc., in our daily lives. The identification of the genuineness of these printed documents is essential. In the present digital era, identifying the authenticity of printed documents is challenging work. Because there is a high possibility of creating fake documents using software, printers, xerox machines etc. The usage of laser printers in day-to-day life is high because of their speed of printing documents, low cost and print quality. However, the use of laser printer devices to create forged documents is also high. Therefore, it is essential to identify the source of documents to authenticate their originality. Each laser printer model has unique printing quality, and this clue is utilized to determine the originality of the documents.

R. Manza et al. (Eds.): ACVAIT 2022, AISR 176, pp. 250–261, 2023.
https://doi.org/10.2991/978-94-6463-196-8_20

According to the literature survey, many approaches have been developed using atrifacts. Imperfections in the printing process lead to artefacts in the printed document, which are invisible to the eyes. These artefacts may be observed in the zoomed versions of printed documents. Experts inspected the documents' ink using chemical and physical analysis. The Raman Spectroscopy technique [1] examines the questionable documents in forensic document science. The drawback of this method is that it might be destroyed using chemicals or any physical devices, and it is time-consuming. The code words are embedded in documents for authenticity of documents. These types of methods are known as extrinsic signature methods [2], and there is no need to embed the additional information in the document. But these approaches are expensive and not followed by all industrialists.

Finally, many strategies have evolved to identify the printer devices based on documents. The laser printer models print distinctive noise due to manufacturer imperfection, printing technologies and slight differences in the different printer models. Each printer has unique printing quality. It can be observed at the margin of the characters in printed documents. In text-dependent approaches, various methods for identifying the source printer are used, and these features are used to simulate the printer-specific imperfection [3, 4]. This method is known as the intrinsic signature method or passive method. The intrinsic method exploits that specific signature that can be searched in printed documents using image processing learning techniques [5]. We present the process for identification of printers based on the printed documents at different types of word-level texture descriptors. It is considered a text-independent approach rather than fixing to a specific character or word image.

The remaining part of the paper is as follows. We present related work in Sect. 2. Section 3 describes the proposed technique, pre-processing, and feature extraction. The experimental results are presented in Sect. 4, and the conclusion is seen in Sect. 5.

2 Review of Related Studies

Significant studies have been taken to identify printer models based on documents. A brief literature review is presented below.

Shize et. al. [6] described a method to differentiate the document images produced by laser printers, inkjet printers, and xerox machines. They extracted the features like contour roughness, noise energy and average gradient from the individual letters in the documents. The accuracy is 90% using SVM classifier. Tsai et. al. [7] employed GLCM (Gray-Level Co-occurrence Matrix) and DWT (Discrete Wavelet Transform) to identify the printers based on documents in the Chinese language. They used SVM classifier to classify printer models, and an accuracy 98.64% has achieved. Elkasrawi et. al. [8] used a method to identify printers based on the noise produced by the printer on the documents. They have used statistical features like mean and contour. This system has classified 20 different models of printers and achieved a classification accuracy of 76.75%. Schreyer et. al. [9] detected a method to find the photocopy and described the recognition of printing technique based on machine learning algorithms. They have achieved the target by spatial and frequency domain analysis of the given document. Lambert et al. [10] developed a method to detect counterfeit documents generated using different printers such as laserjet based on the features like text edge roughness, texture, area difference and correlation coefficient. They carried out the classification of the documents using SVM classifier. Mikkilineni et al. [11] used texture features to identify the printer. They examined the font type, size and paper type of document to find the discernment features of documents. The classifier SVM is employed to classify the printers in printing the documents. Wu et. al. [12] described a method to identify intrinsic features of documents for recognizing the printers. Based on the intrinsic properties, they distinguished the documents produced by the type of printer. The geometric distortion and SVM classifier were used for classifying ten printers. Mikkilinen et al. [13] described a method for identifying the printers based on the GLCM texture features. Ten Electro Photographic (EP) printers were used to classify using KNN classifier. Devi et al. [14] proposed an algorithm for distinguished inkjet printers and photocopiers depending on the analysis of the skew and kurtosis of the histogram of text images. They selected five different printers and three different photocopiers to differentiate each other. Tsai et al. [15] presented the technique that analyses the microscopic printed character images for source identification. They have used SVM classifier to classify the printers using different descriptors such as GLCM, DWT, etc. Ferreira et al. [16] developed a method to identify the printers using character images. They have applied raw, median and average filters on character images to obtain features. The CNN is designed to train the data for this problem. They achieved an accuracy of 97.33% for the classification of ten printers. Jain et. al. [17] detected a method for classifying printers by applying geometric distortion techniques at a text-line level. They achieved 98.85% accuracy in classifying the printer models using SVM classifier for pages with different fonts and printers. Joshi et al. [18] proposed a method for document classification based on images of the letter 'e' that uses a single CNN model derived from the combination of letter images and their printer-specific noise residuals. They achieved an accuracy of 90.33% for their created dataset and 98.01% for the available dataset. Bibi et al. [19] presented a printer identification method for printed documents and developed a text-independent method using pre-trained CNN. They achieved 95.52% accuracy on 1200 documents from 20 different printers. Darwish et al. [20] presented a bio-inspired expert system for printer classification that uses the GLCM of the printed Arabic letter 'WOO' and a Niching genetic algorithm. They achieved a 91% accuracy rate.

3 Proposed Method

In recent years, most techniques have relied on a specific letter of document images, which may prove inadequate in real-world scenarios. In this paper, we present the classification of different laser printer models based on the texture analysis of the word of a document. We have segmented various word images from document images to identify the four laser printer models, and it is a text-independent approach. The textural characteristics are derived from the most widely used technique, LBP.

We also used the deep learning CNN model to classify word images from printed documents. Traditional printer recognition systems have depended on handcrafted characteristics and a significant portion of prior knowledge. CNN is the best method for determining printer models based on documents. In this paper, we examined the effect

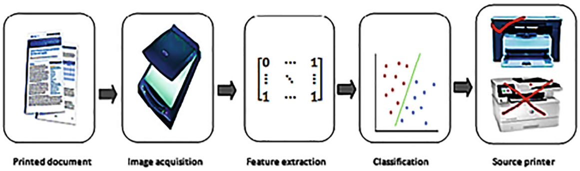

Fig. 1. Overview process of printer identification.

Table 1. Shows the number of word images of four laser printer models

Printer Model	Printed document	Word images	Randomly selected word images
CanoniR2270	100	38690	10000
Canonlbp3108	100	33756	10000
CanoniR7086	100	34540	10000
HPLaserJetM1136	100	30044	10000

of using activation functions ReLU (Rectified Linear Unit) for the inner CNN layer and softmax for the output layer in identifying laser printer models. Figure 1 depicts an overview of the printer identification process.

3.1 Data Collection

The standard dataset is not publicly available to evaluate the proposed method. Hence, we have created a dataset and gathered 100 printed document pages, including research articles. Then, 100 document pages are printed using four different laser printer models: CanoniR2270, CanonLbp3108, CanoniR7086, and HpLaserJetM1136. The word images are segmented from 400 printed document images of laser printer models and generate 137030-word images. To reduce computational complexity and time, we randomly selected 10000 segmented word images from each printer model, and a summary of the dataset is shown in Table 1. Figure 2 shows the samples of four laser printer models. Figure 3 depicts examples of segmented different word images. Figure 3(a) shows a sample of a scanned document image, Fig. 3(b) shows connected bounding box words, and Fig. 3(c) shows an example of segmented word images.

3.2 Pre-processing

The document images are scanned in grayscale and at a resolution of 300 dpi. The printed document image includes images, text, tables, graphs, and equations. The Otsu method [21] converts grayscale document images into binary document images. We used mathematical morphological operations to segment word images from the binary document images. We removed all objects containing fewer than 50 pixels with extraneous pixels along the border from the document images. These pixels have noise such as dots, special symbols, etc. The segmented components are saved as grayscale images to extract the

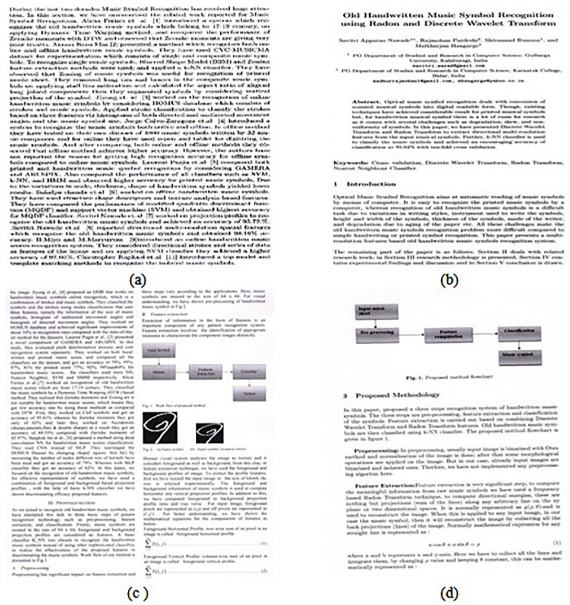

Fig. 2. Samples of document pages from four laser printers (a) CanoniR2270 (b)Canonlbp3108 (c) CanoniR7086 (d)HPLaserJetM1136

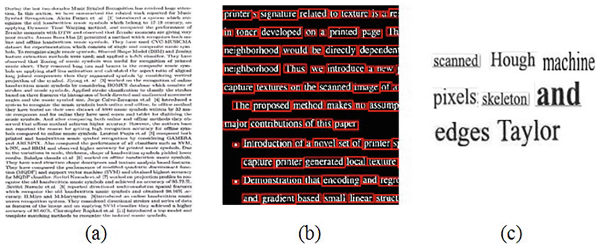

Fig. 3. (a) Document image (b) Bounding box of connected words (c) segmented word images

discriminated features [22]. These word images represent various scales and directions. The median filter approach is applied and effectively accomplished by removing the word images' background noise to improve the dataset's resilience. The word images are multidirectional, and those have multiple resolutions printed in the documents which affect the document images' texture.

This section explains the feature extraction methodology at the word level, using textural and deep learned methods.

3.3 Feature Extraction

This section explains the feature extraction methodology at the various word level, which uses the textural and deep learned method.

3.3.1 Textural Level Feature Extraction

Texture-based descriptors produce effective variations for distinguishing between segmented word images printed by different printers. LBP is an efficient technique to capture the texture of the images. LBP is used to recognize blur faces and the age of the persons most efficiently [23] and identify dating of historical documents most efficiently [24]. Hence, we use this to extract word image features from each set of printer's document images separately. LBP descriptor computes a binary code for each pixel in an image by thresholding circularly symmetric to a neighbouring pixel with the central pixel value. It occurs the different binary patterns by creating the histogram. The Eq. (1) of the LBP is given below.

$$LBP_{(P,R)}(Xc) = \sum\nolimits_{p=0}^{p-1} f(Xp - Xc).2^{p} \tag{1}$$

where

$$f(y) = \begin{cases} 1, y \geq 0 \\ 0, y < 0 \end{cases}$$

Xc indicates the centre pixel in the above equation, Xp indicates one of its p neighbours. To obtain the LBP labels, we assigned $P = 8$ neighbours spaced equally on a circle of radius ($R = 1$). Then, labelled binary patterns are used as the texture features. A total of 59 features are generated for a word image, and the resulting features are concatenated to form the final feature vector. Finally, this feature vector is fetched into a K-NN and SVM classifier to classify the source printer.

3.3.2 CNN Level Feature Extraction

The deep learning CNN architecture has been used for this problem, and its CNN architecture is shown in Fig. 4. It is highly effective in complex image classification [25]. The advantage of CNN over machine learning is that it reduces the number of parameters [26, 27]. Artifacts for source printer identification are limited to handcrafted methods. As a result, we analyze the CNN architecture for source printer classification. We identified solutions for detecting the pattern of word images segmented from document images of four laser printer models based on the CNN model. CNN is used on multiple representations of document images at a word level, allowing for better data discrimination. CNN extracts the necessary discriminating features directly from the document images. This CNN model learns the bias of characteristics from a group of training document images. We set the training phase with stochastic gradient descent, with an initial learning rate of 0.0001 and a maximum number of epochs of 20. The model created at the epoch with the lowest validation loss is the best choice for each CNN. The size of different word images is 40x227. The various word images are included in the dataset, as discussed in Sect. 3.1. The ReLU layer accelerates training and achieves rapid convergence in CNN training. A pooling layer is then used to summarize the data by sliding a window across the feature maps and performing a max nonlinear operation on the data. Classifiers are fully connected layers, and it is usually followed by a soft-max layer, which determines the input image class. The Soft-max layer is used to normalize input values, and its

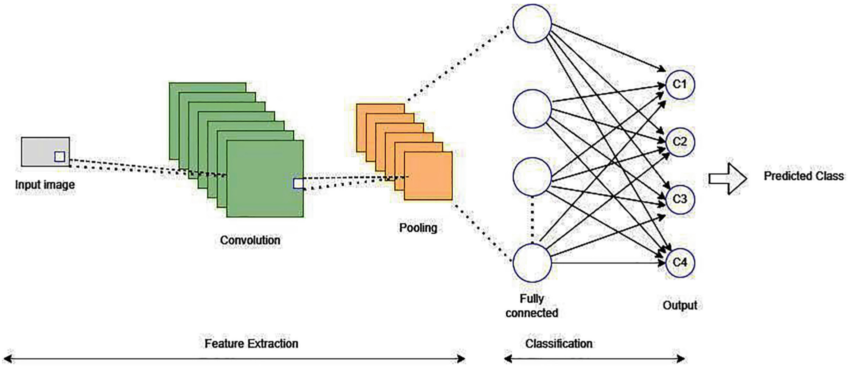

Fig. 4. Basic architecture of CNN model.

output can be interpreted as indicating the probability of a sample belonging to each class.

4 Experimental Results and Discussion

Two sets of experiments are carried out in this section. The experimental setup and results are analyzed in order to evaluate textural and deep learning features. Both experiments are carried out onMatlab software.

4.1 Performance of Textual Features

We used 10-fold cross-validation to classify document images because it is the most accurate and effective validation method and reports the average results of 10 iterations. We considered the randomly segmented various word images from the document images as discussed in Sect. 3.1. Different laser printer models have specific characteristics, and we capture the discriminated 59 features of word images by LBP as mentioned in Sect. 3.3. The features are fed into the K-NN [28] and cubic SVM [29] classifiers for classifying the four laser printer models. The experiment's outcome is measured in terms of overall average accuracy. Table 2 shows four laser printers' classification accuracy. We have attained the average accuracy of words images are 97.2% and 97.9% from classifiers K-NN and cubic SVM, respectively. The confusion matrices are shown in Table 3 and Table 4 for the K-NN and cubic SVM classifiers, respectively.

4.2 Deep Learning CNN Performance Measurement

In a data-driven approach, learn the discriminate features automatically from the readily collected data instead of an enormous training dataset. Extracting the meaningful discriminating data from a set of trained data is also essential for deep learning CNN. The single channel for various word images is used to prepare a CNN model. It works well with small patches of word images. It has been processed by including a Batch Normalization (BN) layer [27] to improve the faster learning of network layer ReLU parameters

Table 2. Classification accuracy of four laser printers using KNN and cubic SVM classifiers

Four Laser Printer models/Classes	KNN-Classifier		Cubic SVM-Classifier	
	Classification rate of words images (%)	Error rate (%)	Classification rate of words images (%)	Error rate (%)
CanoniR2270	95.81	4.19	98.39	1.61
Canonlbp3108	98.49	1.51	98.27	1.73
CanoniR7086	95.12	4.88	96.01	3.99
HplaserJet1136	99.22	0.78	98.73	1.27
Average Accuracy	97.2	2.8	97.9	2.1

and weights because it speeds up the training process. It introduced non-linearity, which improved classification accuracy. When designing the CNN architecture, the combination of BN and ReLU performed admirably in smaller image regions. As examined in Sect. 3.3.2, the CNN model is trained for 20 epochs before selecting the model with the lowest validation loss. The layer has 20 filters with 3X3 dimensions because the CNN's input is single-channel word images. To generate the features maps, a succession of fixed-size filters are applied to the word image. These filters draw attention to patterns useful for image identification, including edges, regular patterns, etc. The classification layer used to classify the CNN characterizes the four-laser printers. We have used 8000-word images for training and 2000-word images for testing to identify the printer models. The evaluation of the deep learning CNN model is estimated by building the CNN architecture. Finally, we achieved 94.3% accuracy in identifying the four laser printer models, promising results. The results are presented as a confusion matrix, as shown in Table 5. The precision, recall, and F1 Score from Eqs. (2), (3), and (4) are also used to evaluate the performance. Precision is measured by how many positives are properly identified out of all positives. The proportion of positive instances that are retrieved is known as recall. The F1 Score is used to calculate the classifier's classification capabilities. It is thought to be a better indicator of the model's performance than the usual accuracy metric.

$$Precision = TruePositive/(TruePositive + FalsePositive) \tag{2}$$

$$Recall = TruePositive/(FalseNegati + TruePositive \tag{3}$$

$$F1score = 2 * (Precision * Recall)/(Precision + Recall) \tag{4}$$

The estimated performance measures of the models are shown in Table 6

4.3 Comparison Analysis

We proposed two experiments in this paper that use handcrafted features and the deep leering CNN model. The developed system is the novel approach for source printer classification for various segmented word images, rather than focusing on a single letter or

Table 3. Confusion matrix using K-NN classifier of four laser printer models

Classes/Printer model	CanoniR2270	CanonLbp3108	CanoniR7086	HpLaserJetM1136
CanoniR2270	**1925**	36	15	24
CanonLbp3108	74	**1853**	24	49
CanoniR7086	66	48	**1660**	226
HpLaserJetM1136	13	5	12	**1970**

Table 4. Confusion matrix using Cubic SVM classifier of four laser printer models

Laser Printers/ Classes	CanoniR2270	Canonlbp3108	CanoniR7086	HplaserJet1136
CanoniR2270	**9839**	51	75	35
Canonlbp3108	17	**9827**	90	66
CanoniR7086	27	105	**9601**	267
HplaserJet1136	6	12	109	**9873**

Table 5. Confusion matrix using CNN model

Classes/printer model	CanoniR2270	Canonlbp3108	CanoniR7086	HplaserJet1136
CanoniR2270	**1955**	13	18	14
Canonlbp3108	74	**1853**	24	49
CanoniR7086	64	48	**1762**	126
HplaserJet1136	13	5	12	**1970**

Table 6. Performance measurement for identifying the four laser printer models using CNN architecture

Precision	Recall	F1 Score	Accuracy
0.9434	0.9440	0.9431	94.3%

word image. As a result, the image quality of laser printer models varies in printed documents. Based on the performance measures such as average precision, average recall, F1 Score, and accuracy obtained. The deep learning CNN model performed well with the train network and achieved a better recognition accuracy rate of 94.3%, lower than the textural approach. The CNN model necessitates massive data and a high computer configuration. Our proposed methodology classified printed documents more accurately and effectively. Table 7 shows the numeric comparative study of our experimental results to the most recent existing methods reported in the literature in [19, 20].

Table 7. Comparison with recent related work

Authors	Techniques	Features	Classifier	Accuracy (%)
Maryam et. al.[19]	Text-independent	Textural features (LBP)	SVM	93.25%
		Deep learning features (Resnet50)	CNN	95.52%
Darwish et. al. [20]	Text-dependent	Textural features (GLCM) and Niching Genetic Algorithm (NGA)	KNN	91%
Proposed Approach	Text-independent	Textural features (LBP)	KNN	**97.2%**
		Textural features (LBP)	Cubic SVM	**97.9%**
		Deep learning features (ReLU CNN)	CNN	**94.3%**

5 Conclusion

The developed textural method LBP and CNN model are the best explorers for determining the source printer. We proved that different printers printed the same document in different ways. This system automatically detects forgery documents based on document images at various word levels. The overall high accuracy classification is 97.9% using the cubic SVM classifier and 97.2% using KNN. We successfully trained the CNN model with a set of parameters to identify the source printer and achieved 94.3%. We plan to improve printer classification accuracy by using a more significant number of printer models and evaluating source mobiles attribution using flatbed and camera-based document images.

References

1. Braz, A., López-López, M., & García-Ruiz, C. (2013). Raman spectroscopy for forensic analysis of inks in questioned documents. *Forensic science international*, *232*(1-3), 206-212.
2. Ali, G. N., Mikkilineni, A. K., Allebach, J. P., Delp, E. J., Chiang, P. J., & Chiu, G. T. (2003, January). Intrinsic and extrinsic signatures for information hiding and secure printing with electrophotographic devices. In *NIP & Digital Fabrication Conference* (Vol. 2003, No. 2, pp. 511–515). Society for Imaging Science and Technology.
3. Gebhardt, J., Goldstein, M., Shafait, F., & Dengel, A. (2013, August). Document authentication using printing technique features and unsupervised anomaly detection. In *2013 12th International conference on document analysis and recognition* (pp. 479–483). IEEE.
4. Ferreira, A., Navarro, L. C., Pinheiro, G., dos Santos, J. A., & Rocha, A. (2015). Laser printer attribution: Exploring new features and beyond. *Forensic science international*, *247*, 105-125.

5. Khanna, N., Mikkilineni, A. K., Chiu, G. T. C., Allebach, J. P., & Delp, E. J. (2008, August). Survey of scanner and printer forensics at purdue university. In *International Workshop on Computational Forensics* (pp. 22–34). Springer, Berlin, Heidelberg.
6. Shang, S., Memon, N., & Kong, X. (2014). Detecting documents forged by printing and copying. *EURASIP Journal on Advances in Signal Processing*, *2014*(1), 1-13.
7. Tsai, M. J., & Liu, J. (2013, May). Digital forensics for printed source identification. In *2013 IEEE International Symposium on Circuits and Systems (ISCAS)* (pp. 2347–2350). IEEE.
8. Elkasrawi, S., &Shafait, F. (2014, April). Printer identification using supervised learning for document forgery detection. In *2014 11th IAPR International Workshop on Document Analysis Systems* (pp. 146–150). IEEE.
9. Schreyer, M., Schulze, C., Stahl, A., &Effelsberg, W. (2009, March). Intelligent Printing Technique Recognition and Photocopy Detection for Forensic Document Examination. In *Informatiktage* (Vol. 8, pp. 39–42).
10. Lampert, C. H., Mei, L., &Breuel, T. M. (2006, November). Printing technique classification for document counterfeit detection. In *2006 International Conference on Computational Intelligence and Security* (Vol. 1, pp. 639–644). IEEE.
11. Mikkilineni, A. K., Arslan, O., Chiang, P. J., Kumontoy, R. M., Allebach, J. P., Chiu, G. T. C., & Delp, E. J. (2005, January). Printer forensics using svm techniques. In *NIP & Digital Fabrication Conference* (Vol. 2005, No. 1, pp. 223–226). Society for Imaging Science and Technology.
12. Wu, Y., Kong, X., & Guo, Y. (2009, November). Printer forensics based on page document's geometric distortion. In *2009 16th IEEE International Conference on Image Processing (ICIP)* (pp. 2909–2912). IEEE.
13. Mikkilineni, A. K., Chiang, P. J., Ali, G. N., Chiu, G. T. C., Allebach, J. P., & Delp, E. J. (2004, January). Printer identification based on texture features. In *NIP & Digital Fabrication Conference* (Vol. 2004, No. 1, pp. 306–311). Society for Imaging Science and Technology.
14. Devi, M. U., Rao, C. R., & Jayaram, M. (2014). Statistical measures for differentiation of photocopy from print technology forensic perspective. *International Journal of Computer Applications*, *105*(15).
15. Tsai, M. J., &Yuadi, I. (2016, September). Printed source identification by microscopic images. In *2016 IEEE International Conference on Image Processing (ICIP)* (pp. 3927–3931). IEEE.
16. Ferreira, A., Bondi, L., Baroffio, L., Bestagini, P., Huang, J., Dos Santos, J. A., ... & Rocha, A. (2017). Data-driven feature characterization techniques for laser printer attribution. *IEEE Transactions on Information Forensics and Security*, *12*(8), 1860-1873.
17. Jain, H., Joshi, S., Gupta, G., & Khanna, N. (2020). Passive classification of source printer using text-line-level geometric distortion signatures from scanned images of printed documents. Multimedia Tools and Applications, 79(11), 7377-7400 https://doi.org/10.1007/s11042-019-08508-x
18. Joshi, Sharad, Suraj Saxena, and Nitin Khanna. "Source printer identification from document images acquired using smartphone." arXiv preprint arXiv:2003.12602 (2020).
19. Bibi, Maryam, Anmol Hamid, Momina Moetesum, and Imran Siddiqi. "Document Forgery Detection using Printer Source Identification—A Text-Independent Approach." In 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW), vol. 8, pp. 7–12. IEEE, 2019.
20. 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.
21. Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, *9*(1), 62-66.

22. Hangarge, M., Santosh, K.C., Doddamani, S., & Pardeshi, R. (2013). Statistical Texture Features based Handwritten and Printed Text Classification in South Indian Documents. *ArXiv, abs/1303.3087*.
23. Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, *24*(7), 971-987.
24. Gonasagi, P., Rumma, S.S., &Hangarge, M. (2020). Classification of Historical Documents Based on LBP and LPQ Techniques.
25. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.
26. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324.
27. Ioffe, S., & Szegedy, C. (2015, June). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning* (pp. 448–456). PMLR.
28. Fix, E., & Hodges, J. L. (1951). *Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties USAF School of Aviation Medicine, Randolph Field* (pp. 1–21). Texas, Tech. Report 4.
29. Cortes, C., &Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3), 273-297