Laser Printer Identification Using Convolutional Neural Network for Forensic Document Authentication

Dr. Pushpalata Gonasagi

Associate Professor, Department of Computer Science, Govt. First Grade College, Mahagaon Cross, Kalaburagi, India.

DOI: https://doi.org/10.51583/IJLTEMAS.2026.150400019

Received: 06 April 2026; 11 April 2026; Published: 02 May 2026

ABSTRACT

Document forgery has become easier with the advancement of printing technologies and image editing software. Identifying the source printer of a printed document is an important task in forensic document analysis. Traditional approaches rely on handcrafted texture features such as Local Binary Pattern (LBP), Local Directional Pattern (LDP), and Local Optimal Oriented Pattern (LOOP). However, these methods require manual feature extraction and often fail to capture complex intrinsic printer signatures effectively. This research proposes a deep learning-based approach using Convolutional Neural Networks (CNN) to automatically identify laser printer models based on texture patterns observed in printed documents. The CNN model learns discriminative features from character-level images without requiring handcrafted descriptors. The dataset consists of scanned document images printed from ten different laser printers, and character-level segmentation is applied to extract the character 'e' images. The proposed CNN-based method achieves high classification accuracy and demonstrates superior performance compared to traditional machine learning approaches such as SVM with handcrafted features. The results show that CNN can effectively capture intrinsic printer signatures and improve document authentication systems in forensic applications.

Keywords: Laser printer, CNN, RELU, Grayscale.

INTRODUCTION

Forgery detection plays an important role in forensic document examination, particularly in cases involving legal documents such as contracts, agreements, wills, ownership papers, and suicide notes [1]. With the rapid evolution of printing technologies, identifying whether a document is genuine or forged has become a challenging task. Modern printers produce high-quality outputs, making it difficult to distinguish printed documents visually. Each printer has a unique intrinsic signature caused by mechanical imperfections, toner distribution, and printing mechanisms [2]. These signatures appear as subtle texture variations in printed characters. Therefore, printer identification can be used as a reliable method for document authentication. Traditional approaches rely on handcrafted feature extraction methods such as LBP, LDP, GLCM, and DWT. However, these techniques require manual feature engineering and may fail to generalize well for complex texture patterns. Recently, deep learning techniques such as CNN have shown remarkable performance in image classification tasks due to their ability to automatically learn hierarchical features. In this research, we propose a CNN-based approach to identify ten laser printer models using character-level images extracted from scanned printed documents.

Related Work

Several studies have been conducted to identify printers based on texture and geometric distortions in printed documents. Elkasrawi et al. [3] proposed a supervised learning method using noise features produced by printers and achieved 76.75% accuracy. Tsai et al. [4] applied GLCM and DWT features with SVM classifier for printer identification and obtained 98.64% accuracy. Lampert et al [5] used text-line features such as edge roughness and correlation coefficients for identifying forged documents using SVM classifier. Mikkilineni et al. [6] analysed font characteristics, paper type, and document age for printer identification. Wu et al. [7] used geometric distortion features at the page level and achieved 100% classification accuracy using SVM. Ferreira

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue IV, April 2026

et al. [8] applied CNN to classify ten printers using character images and achieved 97.33% accuracy. Jain et al. [9] used text-line geometric distortion features and obtained 98.85% accuracy using SVM classifier. Shang et al. [10] extracted contour roughness and noise energy features to differentiate laser printers, inkjet printers, and photocopiers. Although traditional machine learning approaches perform well, they depend heavily on handcrafted features. CNN-based models can automatically extract discriminative features, improving classification performance.

Proposed Method

The proposed CNN-based printer identification system follows four main steps. First, data collection is performed by gathering scanned printed documents from different laser printer models. Second, preprocessing is applied to improve the quality of the extracted character images by resizing, converting to grayscale, removing noise, and normalizing pixel values. Third, feature learning using CNN is carried out, where the convolutional neural network automatically learns important patterns such as edges, textures, and microscopic printing characteristics from the character images. Finally, classification of printer models is performed using the trained CNN model to identify the source printer of the document.

Data Collection

The dataset used in this study consists of scanned printed documents obtained from the Figshare dataset [8]. The documents contain scientist biographies collected from Wikipedia and printed using ten different laser printer models, including Brother HL-4070CDW, Canon D1150, Canon MF3240, Canon MF4370DN, HP CP1518, HP CP2025A, HP CP2025B, Lexmark E260DN, OKI Data C330DN, and Samsung CLP315. A total of 600 pages were printed and scanned at a resolution of 600 dpi to ensure high-quality image capture. The character ‘e’ was selected for analysis because it appears frequently in English text, providing a large number of samples for training the model. Character segmentation was applied to extract individual character images from the scanned pages. The dataset contains 600 scanned pages and approximately 100,000 images of the character ‘e’, out of which 10,000 images were selected for the experiment, with 1,000 images collected from each printer model.

Preprocessing

Character images often contain variations in size, orientation, and noise due to differences in printing and scanning conditions. Therefore, preprocessing is applied to enhance the image quality before training the CNN model. In this process, all character images are first resized to 28 × 28 pixels to maintain a uniform input size for the network. The images are then converted into grayscale to reduce computational complexity while preserving important structural information. A median filtering technique with a 3 × 3 kernel is applied to remove noise generated during scanning, while preserving important edge details of the characters. Finally, the pixel values are normalized between 0 and 1, which helps in faster convergence and improves the learning performance of the CNN model. Median filtering plays an important role in maintaining edge information while effectively reducing scanning noise, thereby improving the feature learning capability of the CNN.

CNN Architecture

The CNN is used to automatically learn discriminative features from the segmented character images without manual feature extraction. The input to the network is a 28 × 28 grayscale image of the character ‘e’. The first convolution layer applies multiple filters to detect low-level features such as edges and curves, followed by a ReLU activation function to introduce non-linearity. A max-pooling layer reduces the spatial dimensions and helps in extracting dominant features while reducing computational complexity. The second convolution layer learns higher-level patterns such as textures and micro-printing characteristics that are unique to each printer. Another max-pooling layer is applied to further reduce dimensionality. The extracted feature maps are then flattened into a one-dimensional vector and passed to fully connected layers, which perform classification based on learned patterns. A softmax output layer is used to classify the input character image into one of the ten printer models [11]. The CNN architecture effectively captures subtle printing artifacts such as toner distribution, microscopic distortions, and texture differences, enabling accurate identification of the source printer. CNN automatically extracts hierarchical features such as edges, shapes, and textures from character images. The proposed CNN architecture and image as shown in the Table 1 and Figure 1 respectively.

Table 1: CNN architecture

Layer No.	Layer Name	Configuration	Output Size	Purpose
1	Input Layer	28 × 28 grayscale image	28 × 28 × 1	Takes character image as input
2	Convolution Layer 1	32 filters, kernel size (3 × 3), Activation: ReLU	26 × 26 × 32	Extracts low-level features such as edges and textures
3	Max Pooling Layer 1	Pool size (2 × 2)	13 × 13 × 32	Reduces spatial dimensions and computation
4	Convolution Layer 2	64 filters, kernel size (3 × 3), Activation: ReLU	11 × 11 × 64	Extracts complex patterns and shapes
5	Max Pooling Layer 2	Pool size (2 × 2)	5 × 5 × 64	Further reduces feature map size
6	Flatten Layer	Converts 2D feature maps to 1D vector	1600	Prepares data for fully connected layer
7	Fully Connected Layer	128 neurons, Activation: ReLU	128	Learns high-level feature representation
8	Output Layer	10 neurons, Activation: Softmax	10	Classifies input into 10 printer classes

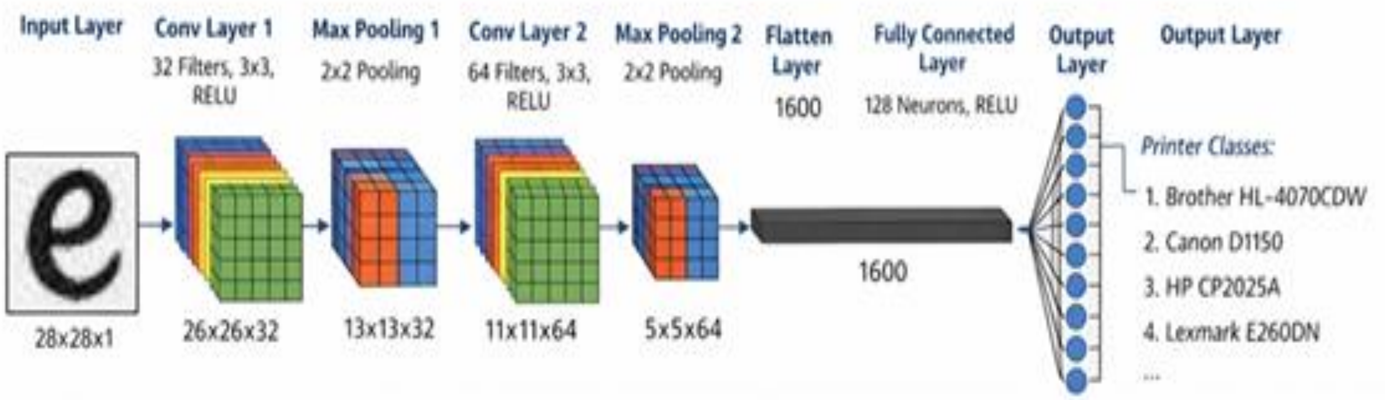

Figure 1: CNN architecture of printer identification.

Training Parameters

The CNN model was trained using carefully selected parameters to ensure effective learning and accurate classification of printer models. The input images used for training were 28 × 28 pixels in size and in grayscale format, which reduces computational complexity while preserving important texture and structural information necessary for printer identification. The categorical cross-entropy loss function was used because the problem involves multi-class classification with 10 different printer models. The Adam optimizer was applied to efficiently update network weights and achieve faster convergence during training [12][13]. Model performance was evaluated using standard metrics including accuracy, precision, recall, and F1-score, which provide a comprehensive assessment of classification performance.

INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING,
MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)
ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue IV, April 2026

EXPERIMENTAL RESULTS AND DISCUSSION

The experiment was conducted using Python with the TensorFlow deep learning framework to implement and evaluate the proposed CNN model. The dataset was divided into training and testing sets, where 80% of the data (8000 images) was used for training the model and 20% (2000 images) was used for testing its performance. The confusion matrix is shown in the Table [2]. The experimental results show that the CNN model achieved an accuracy of 99.3% as shown in the Table [3], while the SVM classifier with LOOP feature extraction achieved 99.8% accuracy [14].

The comparison result is shown in the Table [4]. The CNN model effectively captures subtle texture variations caused by toner distribution and mechanical differences among printers. The experimental result is slightly lower than the existing method due to limited system resources. The experimental implementation was carried out on a system with limited computational resources (hardware constraints), which influenced the training performance of the CNN model. But the major advantages of the CNN-based approach include automatic feature extraction, high classification accuracy, reduced manual effort, robustness to noise variations, and scalability for large datasets, making it suitable for printer identification tasks in digital forensics.

Table 2. Confusion matrix

Ten Laser Printer models/Classes	**Brother --HL- 4070CDW**	**Canon --D1150**	**Canon-- MF3240**	**Canon--MF4370DN**	**Hewlett.PackardCP15 18**	**Hewlett.packard- CP2025A**	**Hewlett ackard- CP2025B**	**Lexmark-E260DN**	**OKI Data-C330DN**	**Samsung- CLP315**
Brother- -HL-4070CDW	994	1	--	2	--	--	2	--	1	--
Canon --D1150	--	994	1	1	1	1	2	--	--	--
Canon-- MF3240	--	--	997	1	1	--	--	--	--	1
Canon-MF4370DN	1	1	2	992	1	2	--	1	--	--
Hewlett Packard-CP1518	--	--	--	--	999	--	1	--	--	--
Hewlett Packard-CP2025A	--	--	--	--	--	1000	--	--	--	--
Hewlett Packard-CP2025B	--	--	--	4	--	--	994	1	1	--
Lexmark-E260DN	1	3	--	1	2	--	--	1000	--	--
OKI Data-C330DN		4	--	--	--	--	--	--	1000	--
Samsung- CLP315	4	--	2	--	--	1	--	--	--	993

Table 3. Classification accuracy using CNN with ReLU activation

Ten Laser Printer models/classes	**Classification rate (%)**	**Error Rate (%)**
Brother -HL- 4070CDW	99.4	0.6
Canon -D1150	99.4	0.6
Canon- MF3240	97.0	3.0
Canon-MF4370DN	99.1	0.9
Hewlett Packard- CP1518	98.8	1.2
Hewlett Packard- CP2025A	100	0.0
Hewlett Packard- CP2025B	99.4	0.6
Lexmark-E260DN	99.2	0.8

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue IV, April 2026

OKI Data-C330DN	100	0
Samsung- CLP315	100	0
Average Accuracy	99.3	0.7

Table 4. Comparison performance of the proposed method.

Authors	Classifier	Accuracy (%)
Gonasagi et al. [14]	Linear SVM	99.2
	Quadratic SVM	99.8
Proposed Approach	CNN + ReLU activation	99.3

CONCLUSION

This research presented a CNN-based approach for identifying laser printer models using printed character images. The experimental results demonstrate that the CNN model can effectively capture intrinsic printer signatures, such as texture patterns and toner distribution characteristics, and classify printers with high accuracy. The proposed method enhances forensic document authentication by providing an automated and reliable system for identifying the source printer of a document. For future work, the study can be extended by including inkjet and dot-matrix printers, utilizing word-level and line-level features, and increasing the dataset size to improve model generalization. In addition, advanced transfer learning models such as ResNet and VGG can be applied to further improve performance, while efforts can be made to reduce computational complexity for faster processing. Overall, CNN-based printer identification systems can assist forensic experts in efficiently detecting forged documents and determining document ownership, making them valuable tools in the field of digital forensics.

REFERENCES

1. Hilton, O. (1992). Scientific examination of questioned documents. CRC press.
2. 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).
3. 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.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue IV, April 2026

11. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.
12. 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.
13. 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.
14. Gonasagi, Pushpalata, and Mallikarjun Hangarge. "Source Identification of Documents Based on LOOP Features." In Futuristic Trends for Sustainable Development and Sustainable Ecosystems, pp. 237-248. IGI Global Scientific Publishing, 2022.