*Research Journal of Recent Sciences* ISSN 2277-2502
Vol. **15(3),** 81-85, July **(2026)** *Res. J. Recent Sci.*

# Deep Attention Models for Identification of Laser Printed Document Origins

**Pushpalata Gonasagi**
Department of Computer Science, Govt. First Grade College Mahagaon Cross, Kalaburagi, Karnataka, India
gonasagi99@gmail.com

**Available online at: www.isca.in, www.isca.me**
Received 10th April 2026, revised 13th May 2026, accepted 13th June 2026

## Abstract

*In the field of digital forensic science, artificial intelligence (AI) technologies are being used more and more to settle document-related disputes that have historically required the knowledge of human experts. Finding the precise printer that generated a given document is one of the main goals of intelligent systems based on printer identification. Nevertheless, a lot of current methods rely on text-dependent approaches, which might not work in some forensic situations. These drawbacks have spurred research into text-independent methods using word images from various laser printer models. In this work, laser printer models are classified using grayscale word images. The constructed dataset consists of 100,000-word images collected from five distinct laser printer models. A deep learning–based Convolutional Neural Network (CNN) is employed to identify the source laser printer model. The performance of the proposed CNN architecture is evaluated and compared with recent studies reported in the literature, particularly those based on textural feature analysis. Experimental results demonstrate that the proposed CNN model achieves a high classification accuracy of 98.8%, outperforming several existing methods.*

**Keywords:** Printers, Deep Learning, document, CNN, ReLU.

## Introduction

Examples of printed documents that are regularly used in daily life for official transactions, security, and educational purposes include land records, agreements, bills, and other legal documents. Therefore, validating the authenticity of printed records is of significant importance. Determining the authenticity of documents in the modern digital era is a challenging task, as printers, photocopying machines, and software tools have great potential for document forgery. The laser printers are famous because of their minimum prize, printing speed, and also print quality. However, the manufacturing of fake documents has also been made easier by the growing usage of laser printers. Consequently, to ensure that documents are not copied or forged, it is essential to identify their source of origin. A document's originality can be ascertained by utilizing the distinctive printing properties that each laser printer model possesses.

The literature review indicates that printing artifacts have been utilized in the development of various identification techniques. Due to flaws in the printing process, artifact patterns like dot inconsistencies in printed papers are invisible to the human eye. However, these artifacts can be observed when printed documents are magnified. Physical and chemical analysis techniques have been used in a number of studies to investigate the characteristics of ink in documents. In recent years, a number of document-based printer identification methods have evolved. Because of manufacturing flaws, printing processes, and slight variances between printer types, each laser printer has unique characteristics that affect output quality. These differences are often seen in printed character margins and edges[1].

The intrinsic attributes are used in text-dependent algorithms to determine the printing origin[2]. Image processing–based learning methods are used to detect printer-specific signatures present in printed media[3]. In this work, we present a method for analyzing printer sources based on printed texts using different word-level texture descriptors. The suggested approach is text-independent and does not rely on particular letters or word images. The remainder of the paper is organized as follows. Section II presents related work. Section III obtained suggested approach, which includes preprocessing and feature extraction. Section IV presents the experimental results, and Section V concludes the paper.

**Literature Review:** Considerable research efforts have been devoted to identifying printer models based on printed documents. A summary of the relevant literature is presented below. Elkasrawi et al.[4] proposed a method for printer identification by analyzing noise patterns produced during the printing process. Statistical data such as contour information and mean values were recovered, allowing for the categorization of twenty different printer models with a 76.75% accuracy.

Shang et al.[5] introduced a method to visualize differences among document images generated by laser printers, inkjet printers, and photocopiers. By means of individual characters from text, they extracted features. An accuracy of 90% was achieved using a Support Vector Machine classifier. The printer devices were categorized using an SVM classifier, achieving an accuracy of 98.64% in other methods. Schreyer et al.[6] investigated methods for identifying photocopied documents and proposed a machine learning–based framework to determine the printing technique. This method applied for spatial and frequency domain studies.

Lambert et al.[7] developed a method to detect forged documents produced by different printers, including laser printers, by analyzing attributes. Wu et al.[8] reported a method for detecting intrinsic document properties to identify printers. Documents printed using different printer types were distinguished based on defining characteristics, and geometric distortion features combined with an SVM classifier were used to classify ten printers. Devi et al.[9] presented a technique based on skewness and kurtosis analysis of text image histograms to differentiate between inkjet printers and photocopiers. Their study considered three manufacturers, photocopiers, and printers. Tsai et al.[10] developed a source identification technique based on the analysis of small printed character images. Various descriptors, including GLCM and Discrete Wavelet Transform (DWT), were extracted, and an SVM classifier was used for printer classification. Ferreira et al.[11] proposed a printer recognition method using character images. Raw, median, and average filtering approaches were used to extract features, while a Convolutional Neural Network (CNN) was trained for classification. Their technique had a classification accuracy of 97.33% across ten printer models.

Jain et al.[12] proposed a printer classification approach based on text-line geometric distortion data. Using an SVM classifier, they correctly detected printer models across various fonts and printer kinds, with an accuracy of 98.85%. Joshi et al.[13] proposed a single CNN model that combines character images and printer-specific noise residuals for document classification, using images of the letter ‘e’. Their method achieved a precision of 98.01% on standard dataset and 90.33% on a self-created dataset. Bibi et al.[14] employed a pre-trained CNN in a text-independent printer identification framework. This study involved twenty different printers and 1200 printed document pages, attaining an correctness of 95.52%. Darwish et al.[15] proposed an expert system for printer classification using a bio-inspired genetic algorithm combined with GLCM features extracted from printed Arabic letter “WOO.” Their approach achieved an accuracy of 91%.

## Methodology

In recent years, existing approaches have depended on specific characters in document images, which may be not effective in real-world circumstances. This research describes the classification of several laser printer models using word-level textural data taken from printed documents. Different word images are extracted from document images to identify five laser printer models using a text-independent approach. To categorize extracted word pictures from printed documents, a deep learning-based Convolutional Neural Network (CNN) is used. Traditional printer recognition algorithms usually rely on handcrafted features and need significant prior knowledge. In contrast, CNN-based approaches offer a more efficient and automated option for identifying printer models.

This paper also discusses the impact of activation functions, specifically the Rectified Linear Unit (ReLU) in the hidden layers and the SoftMax function in the output layer, on the classification of laser printer models. Figure-1 represents the complete system for the proposed printer identification method. This is the extension work of the study of reference[22].

**Data Collection:** The standard dataset used to evaluate the proposed method is not publicly available. Therefore, we collected 100 pages of written documents. These documents were printed on five laser printer models: Canon iR2270, Canon LBP3108, Canon iR7086, HP LaserJet M1136, and HP LaserJet Pro MFP M126a. A total of 500 printed pages from these printers were scanned, giving one lakh word pictures, which were randomly chosen from each printer model for training and assessment. Table-1 provides a detailed description of the dataset and Figure-2 displays the samples of segmented word images.

**Pre-processing:** The document pictures were scanned in grayscale at 300 dpi. Individually scanned document images contain variety of text, images, tables, equations and graphs, The Otsu method[16] was employed to convert grayscale document images into binary images. Word images were then extracted from the binary documents using mathematical morphological operations. This technique deleted items with fewer than 50 pixels, as well as stray pixels around the boundaries. These little components typically transmit noise in the form of incorrect characters, dots, or artifacts. To assist the extraction of discriminative characteristics, individual word pictures were preserved as grayscale images after segmentation[17]. The repossessed word images vary in size and orientation. A median filter was used to decrease background noise, which improved the dataset's quality and consistency. Dissimilarities in word orientation and print resolution have an influence on the texture features derived from these photos. Deep learning algorithms were then used for feature extraction at the word level, as explained in the next section.

**CNN level feature extraction:** This section defines an approach for deep learning-based feature extraction at the word level. CNNs are extremely effective at complex image classification tasks[18,19]. CNNs' primary benefit over standard machine learning approaches is their capacity to automatically train discriminative features, which decreases the requirement for manual feature engineering and the amount of parameters[20,21]. Contrasting other printer identification methods that rely on handcrafted features, CNNs can learn patterns from data, making them ideal for source printer classification. Using the CNN model, patterns in segmented word pictures from documents generated by five laser printer models were identified. The CNN uses several representations of word-level document pictures, allowing for precise feature discrimination. The model automatically derives the relevant discriminative characteristics from the input images. The CNN was trained with stochastic gradient descent for up to 20 epochs with a learning rate of 0.0001.

The optimum model was chosen based on the smallest validation loss observed during training. The word images were standardized at 40×200 pixels. As stated, the collection includes word images retrieved from scanned document images. To speed up training and convergence, the hidden layers employed the ReLU activation function. Pooling layers were used to summarize feature maps by sliding a window across them and choosing the most often occurring features with the max function.

The SoftMax layer at the output determines the probability distribution across classes and generates the printer model for each input word image, while the fully linked layers function as classifiers. The output values, which can then be read as the probability that the input belongs to each class, are normalized by the SoftMax function.

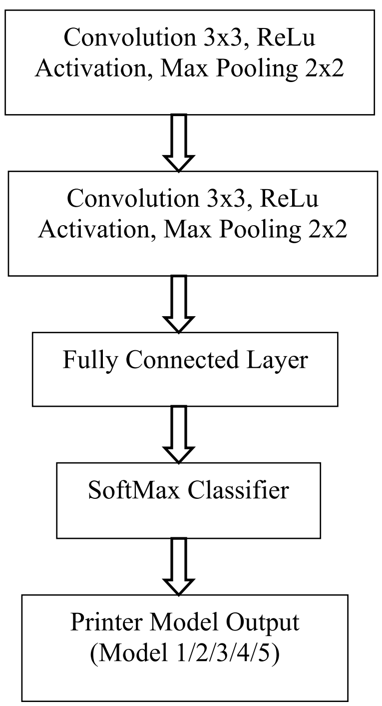


**Figure-1:** Basic architecture of CNN model.

**Table-1:** Number of word images of five laser printer models.

| Printer Model | Printed document | Word images |
|---|---|---|
| CanoniR2270 | 100 | 20000 |
| Canonlbp3108 | 100 | 20000 |
| CanoniR7086 | 100 | 20000 |
| HPLaserJetM1136 | 100 | 20000 |
| HPLaserJet Pro MFP M126a | 100 | 20000 |


**Figure-2:** Samples of segmented word images.

## Results and Discussion

The experimental setup is explained in this section. The design of the tests and the analysis of the results acquired serve as the foundation for the evaluation of textural aspects. Matlab software was used for all of the experiments.

**Deep learning CNN performance measurement:** A data-driven method does not require a large training set because discriminative features are automatically learned from the existing data. The extraction of significant discriminative features from a batch of training data is a crucial prerequisite for a deep learning CNN. In this study, single-channel word images are used to train the CNN model, which performs well even with brief word image segments. A Batch Normalization (BN) layer, which speeds up the training of network parameters and weights in the ReLU layers, was used to improve the learning process. Additionally, BN adds non-linearity, which increases the accuracy of classification. BN and ReLU worked well together to handle small word image patches in the CNN architecture that was designed. The CNN model was trained for 20 epochs, and the model with the lowest validation loss was chosen. Single-channel word images are used as input for the 20 convolutional filters in the first layer, which have dimensions of 3×3. To create feature maps, a series of fixed-size filters are applied to the word images. Patterns like edges and repeating structures, which are crucial for differentiating images, are captured by these filters. The word images are divided into five laser printer models using the CNN classification layer. 2000-word images were utilized for assessment, and 18000-word images were used for training.

Our suggested methodology offers a successful and economical way to categorize printed documents, despite the fact that CNN designs usually require big datasets and costly processing resources. F1-score, precision, and recall were used in the analysis. As indicated in Table-2, the suggested CNN identified the five laser printer models with a promising classification accuracy of 98.8%. Comparative Analysis: With a recognition accuracy of 98.8%, the deep CNN model demonstrated remarkable efficacy. Our suggested methodology offers a successful and economical way to categorize printed documents, despite the fact that CNN designs usually require big datasets and costly processing resources. A quantitative comparison of our experimental results with contemporary techniques documented in the literature[14,22] is shown in Table-3, which emphasizes the enhanced performance of our method.

**Table-2:** Accuracy rate of five laser printers using F1 Score.

| Five Laser Printer | CNN (ReLU Activation Function) | |
|---|---|---|
| | Classification rate of words images (%) | Error rate (%) |
| CanoniR2270 | 99.3 | 0.7 |
| Canonlbp3108 | 96.3 | 3.7 |
| CanoniR7086 | 98.5 | 1.5 |
| HplaserJet1136 | 99.8 | 0.2 |
| HPLaserJet Pro MFP M126a | 99.7 | 0.3 |
| Average Accuracy | 98.8 | 1.2 |

**Table-3:** Comparison analysis with recent work.

| References | Techniques | Features | Classifier | Accuracy (%) |
|---|---|---|---|---|
| 14 | Text-independent | Textural features (LBP) | SVM | 93.25% |
| | | Deep learning features (Resnet50) | CNN | 95.52% |
| 22 | Text-independent | Textural features (LBP) | KNN | 97.2% |
| | | | Cubic-SVM | 97.9% |
| | | Deep learning features (ReLU- CNN) | CNN | 94.3% |
| Proposed Approach | Text-independent | Deep learning features (ReLU -CNN) | CNN | 98.8% |

## Conclusion

This work offered a CNN-based technique on word images to detect laser printers, and leading to text-independent solution for forensic document analysis. The suggested model beat the current feature-based methods with an accuracy of 98.8%. The results demonstrate that word-level, discriminative printer-specific information may be effectively extracted from textures using deep learning. Future research will concentrate on adding more number printer kinds to the dataset in order to improve the method's robustness and use.

## References


1. Gebhardt, J., Goldstein, M., Shafait, F., & Dengel, A. (2013). Document authentication using printing technique features and unsupervised anomaly detection. In *2013 12th International conference on document analysis and recognition* (pp. 479-483). IEEE.
2. 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 *International Workshop on Computational Forensics* (pp. 22-34). Berlin, Heidelberg: Springer Berlin Heidelberg.
3. 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.
4. Elkasrawi, S., & Shafait, F. (2014). Printer identification using supervised learning for document forgery detection. In *2014 11th IAPR International Workshop on Document Analysis Systems* (pp. 146-150). IEEE.
5. Shang, S., Memon, N., & Kong, X. (2014). Detecting documents forged by printing and copying. *EURASIP Journal on Advances in Signal Processing*, *2014*(1), 140.
6. Schreyer, M., Schulze, C., Stahl, A., & Effelsberg, W. (2009). Intelligent Printing Technique Recognition and Photocopy Detection for Forensic Document Examination. In *Informatiktage* (Vol. 8, pp. 39-42).
7. Lampert, C. H., Mei, L., & Breuel, T. M. (2006). Printing technique classification for document counterfeit detection. In *2006 International Conference on Computational Intelligence and Security* (Vol. 1, pp. 639-644). IEEE.
8. Wu, Y., Kong, X., You, X. G., & Guo, Y. (2009). Printer forensics based on page document's geometric distortion.

In *2009 16th IEEE International Conference on Image Processing (ICIP)* (pp. 2909-2912). IEEE.

**9.** 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).

**10.** Gonasagi, P., Rumma, S. S., & Hangarge, M. (2023). Text-independent source identification of printed documents using texture features and CNN model. In *First International Conference on Advances in Computer Vision and Artificial Intelligence Technologies (ACVAIT 2022)* (pp. 250-261). Atlantis Press.

**11.** 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.

**12.** 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.

**13.** Joshi, S., Saxena, S., & Khanna, N. (2020). Source printer identification from document images acquired using smartphone. *arXiv preprint arXiv:2003.12602*.

**14.** Bibi, M., Hamid, A., Moetesum, M., & Siddiqi, I. (2019). 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.

**15.** 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.

**16.** Otsu, N. (1979). A threshold selection method from gray-level histograms. *Automatica*, 11(285-296).

**17.** Hangarge, M., Santosh, K. C., Doddamani, S., & Pardeshi, R. (2013). Statistical texture features based handwritten and printed text classification in south Indian documents. *arXiv preprint arXiv:*1303.3087.

**18.** Gonasagi, P., Rumma, S. S., & Hangarge, M. (2020). Classification of historical documents based on LBP and LPQ techniques. *Int J Innov Technol Exploring Eng* , *9*(3).

**19.** Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.

**20.** Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning* (pp. 448-456). pmlr.

**21.** Fix, E. (1985). *Discriminatory analysis: nonparametric discrimination, consistency properties* (Vol. 1). USAF school of Aviation Medicine.

**22.** Gonasagi, P., Rumma, S. S., & Hangarge, M. (2023, August). Text-independent source identification of printed documents using texture features and CNN model. In *First International Conference on Advances in Computer Vision and Artificial Intelligence Technologies (ACVAIT 2022)* (pp. 250-261). Atlantis Press.