International Journal of Medical Toxicology & Forensic Medicine
2026, Volume 16, E53598

Research Paper

Forensic Source Attribution of Color Laser Printers from Machine Identification Code Patterns: Impact of Class Imbalance on Machine Learning Performance

Shreya Arora[1], **Rajendra Kumar Sarin**[2], **Pooja Puri**[1*]

1. *Amity Institute of Forensic Sciences, Amity University, Noida, Uttar Pradesh, India.*
2. *National Forensic Sciences University, Delhi Campus, New Delhi, India.*

Citation Arora S, Sarin RK, Puri P. Forensic Source Attribution of Color Laser Printers from Machine Identification Code Patterns: Impact of Class Imbalance on Machine Learning Performance. International Journal of Medical Toxicology and Forensic Medicine. 2026; 16:E53598.

https://doi.org/10.22037/ijmtfm.v16.53598

Article info:

Received: 28 September, 2026

Accepted: 14 October, 2026

Published: 09 November, 2026

Keywords:

Class Imbalance, Ensemble Learning, Document Authentication, Gray Level Co-occurrence Matrix, Machine Identification Code, Printer Forensics, Texture Analysis

ABSTRACT

Background: Printer source attribution cases are among the most critically benchmarked on artificially balanced datasets, a practice that masks the classifiers. This further complicates the study when one manufacturer dominates the evidence pool. This research study further frames the source identification as a class-imbalanced problem and uses features like geometric and textural structures of the Machine Identification Code (MIC) dot patterns from the samples collected across six major color laser printer manufacturers under the skewed, uncurated class distribution of the printed outputs.

Methods: Nine interpretable descriptors spanning geometric complexity (Fractal Dimension, Laplacian Variance, and Standard Deviation) and Gray-Level Co-occurrence Matrix texture (Contrast, Correlation, Homogeneity, Energy, Entropy) were extracted and evaluated first through unsupervised clustering and then through five supervised classifiers. The unsupervised clustering failed to recover brand identity from the raw feature space, confirming that the proximity- based grouping alone cannot resolve the manufacturer overlap.

Results: Among the supervised models, Random Forest achieved the strongest balanced performance and the highest accuracy, consistently outperforming the distance-based and single-tree baselines. Furthermore, per-class analysis exposed a pronounced imbalance vulnerability, while the majority class and texturally distinctive brands, including Konica, Ricoh, and HP were identified with near-perfect recall. The minority- class brand Toshiba was misclassified in every model (recall = 0), and Xerox recall remained below 33% throughout.

Conclusion: These results indicate that the reported aggregate accuracy substantially overstates the real- world reliability for the underrepresented printer sources, and they motivate the imbalance training strategies as a necessary next step for the forensic-grade printer attribution systems rather than an optional refinement.

*** Corresponding Author:**
Pooja Puri, PhD
Amity Institute of Forensic Science, Amity University, Noida, Uttar Pradesh, India.
E-mail: *pmalik1@amity.edu*

Copyright © 2026 The Author(s);
This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0: https://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

The widespread availability of high-resolution digital printers has unintentionally contributed to an increase in document forgery, counterfeiting, and anonymous threats [1, 2]. In document examination, establishing a link between a printed document and its source printer remains a vital goal [3]. Traditional forensic methods rely on macroscopic defects such as the banding artifacts described by Mikkilineni et al. [4] or the geometric distortion of text lines proposed by Orhan et al. [5]. These techniques often fail when the questioned document contains limited text or is printed at high quality, where electromechanical fluctuations are also minimized [6, 7]. As a result, modern research has shifted toward analyzing extrinsic signatures embedded by manufacturers, most notably the MIC, commonly known as yellow dots [2, 8].

MICs are microscopic yellow dot patterns invisible to the naked eye, embedded by color laser printers to track the device's origin [8–10]. Recent problems related to the authenticity and source identification of printed documents have shifted the focus toward analyzing the morphological and textural properties of the dots themselves. Studies by Ferreira et al. [11]and Choi et al. [12] demonstrated that the electrophotographic process leaves discrete noise signatures using the toner particles [13–16]. Furthermore, existing feature-based methods often struggle to discriminate between brands that utilize similar frequency-modulated (FM) screening techniques, creating a need for more rigorous geometric and textural descriptors [14, 17, 18].

In practice, printer source attribution is shaped by the devices actually in circulation. Brands like Ricoh, Konica Minolta, Xerox, and Toshiba are mostly used as multifunctional office printers, and public reports do not itemize vendor-level shipment figures. Evidence pools drawn from casework are therefore expected to be dominated by a few brands, with the remaining manufacturers appearing in rare cases [19].

This study addresses three major research gaps: (1) MIC research has largely focused on visually comparing the dot matrix of one brand at a time using manual image-editing workflows [8–10, 20], while automated, cross-brand discrimination based on physical appearance has received little attention. (2) In a recent review [2], it was observed that the learning-based printer-attribution studies are mostly reported on curated, class-balanced datasets, so the per-class cost of imbalance is rarely quantified. (3) It is occasionally examined how an MIC-based classifier behaves when the questioned document comes from a brand that is scarce in the reference data, which is one of the major problems faced by the examiners.

The key question addressed in this study is whether the physical appearance of MIC dots attributes a questioned color laser printout to its manufacturer even when the reference data are imbalanced. The study fills these gaps by providing an automated, cross-brand benchmark of MIC dot features under uncurated class proportions, with per-class reporting of where and why imbalance degrades attribution.

This study makes three contributions. First, it assesses whether physical texture traits of the MIC alone are sufficient for printer identification, without decoding the embedded binary code. Second, it characterizes which geometric and textural descriptors are most discriminative across six color printer brands, using both unsupervised and supervised analysis. Third, and most centrally, it departs from the balanced-dataset convention followed in prior printer-attribution literature and instead evaluates classifier performance under a naturally occurring, uncurated class distribution. This third contribution reframes printer source attribution as a class-imbalanced learning problem, and the resulting per-class breakdown quantifies a vulnerability that balanced-dataset benchmarks cannot reveal: standard classifiers trained on skewed forensic data systematically underperform on minority-class printer brands even when aggregate accuracy appears acceptable.

Materials and Methods

Sample Acquisition and Dataset Construction

This study uses a dataset of 420 printed documents produced by six color laser printer manufacturers: Canon, HP, Konica Minolta, Ricoh, Toshiba, and Xerox. These manufacturers were chosen to capture both within-brand variation and between-brand differences in the observed MIC patterns. We sampled multiple printer models from each manufacturer, sourced through a range of channels across the Delhi–NCR region in India. Each page was printed using standard A4 white paper with 75 GSM (JK Copier) with default color settings, trying to keep dot density and intensity comparable across the datasets collected [19].

Image Pre-processing with Enhancement

Documents were scanned at 600 dpi (dots per inch) using an HP flatbed scanner to extract microscopic yellow tracking dots visible, and images were saved in uncompressed form, i.e., TIF (Tagged Image File

Format), to avoid compression-related artifacts and preserve fidelity [19]. Because the yellow dots show very low contrast against white paper, an enhancement workflow was built in GIMP 2.10.34 (GNU Image Manipulation Program). The non-linear adjustments on the blue channel gave clear segmentation, which included setting brightness to −108 and contrast to 127 to suppress paper texture while making the dot visible [8, 20].

Feature Extraction

A pipeline with Python was built to quantify the geometric and textural appearances of the detected dot [21]. The following features fell into two broad categories that include:

Geometric Features

The fractal dimension was designed to capture the self-similarity and complexity observed in the spatial arrangement of dots. The Laplacian Variance served as a proxy for sharpness, edge strength, Standard Deviation, and overall variation in pixel intensity in the yellow dots detected [7, 17, 22].

Textural Features

GLCM-derived statistical texture features include Contrast, Energy, Homogeneity, and Correlation, which were computed alongside Entropy, which captures the randomness of dot distribution across the printed output [13, 15, 23].

Unsupervised Pattern Framework

Dimensionality reduction was applied in the first step, using Principal Component Analysis (PCA) to retain the dominant variance and disclose how the yellow dot patterns are organized within this feature space [24, 25]. It was then followed by t-distributed Stochastic Neighbor Embedding (t-SNE) to produce a two-dimensional visualization that actually preserves local neighborhood relationships [24]. The three clustering methods were compared, including K-Means, DBSCAN, and Gaussian Mixture Models (GMM) [13, 15]. The Cluster quality was judged internally via the Silhouette Score and Davies–Bouldin Index, and externally via the Adjusted Rand Index (ARI) [26].

Supervised Classification Framework

A supervised learning framework was then assembled on the extracted features to identify the source color-laser printers [10,27,28]. The dataset was split into 80/10/10 for training, validation, and independent test sets, preserving class proportions across the samples. The five classifiers were compared, including Decision Tree (DT), k-Nearest Neighbors (k-NN), Random Forest (RF), Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP). Furthermore, a 5-fold cross-validation framework was used during training for stable performance estimates and to limit overfitting for better decipherment [29,30].

Data Quality Checks and Outlier Handling

The same enhancement settings and feature-extraction pipeline were applied to every scanned image, and no samples were removed based on their feature values. The distribution of each descriptor was screened for extreme values using boxplots (whiskers at 1.5 × the interquartile range). Values beyond the whiskers were seen in several brands and were deliberately retained. This retention was necessary because multiple printer models were sampled per brand; given the wide gap within-brand spread, this variability is expected and does not itself indicate measurement error. Also, removing the samples would disproportionately thin the minority brands and alter the class distribution considered during the study. Furthermore, the documents in question will carry the same variability in the real world, and a model tuned on trimmed data would overstate the effect.

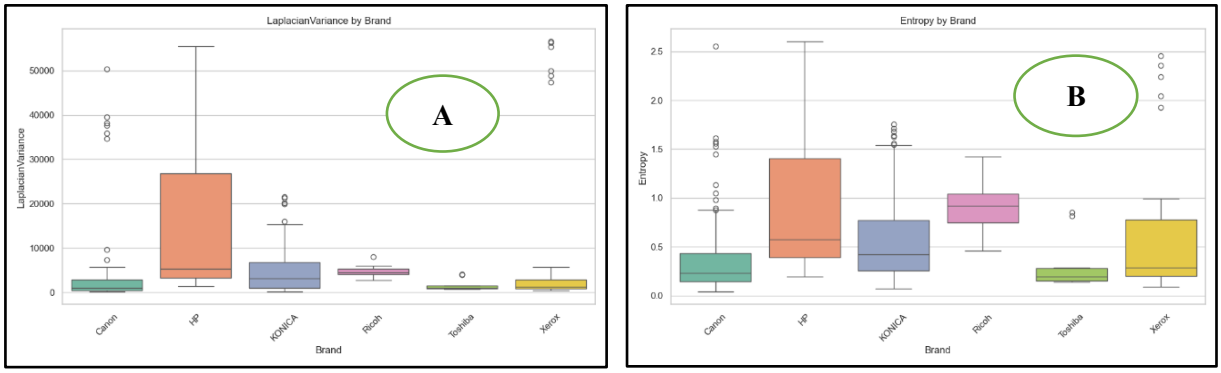
Results

Feature-Based Differentiation of Dot Patterns

The inter-brand variability in yellow dot patterns across printed documents was examined, and a comprehensive set of features was studied from scanned and collected printed outputs. The analysis of structural and textural descriptors shows brand-specific characteristics that support printer identification.

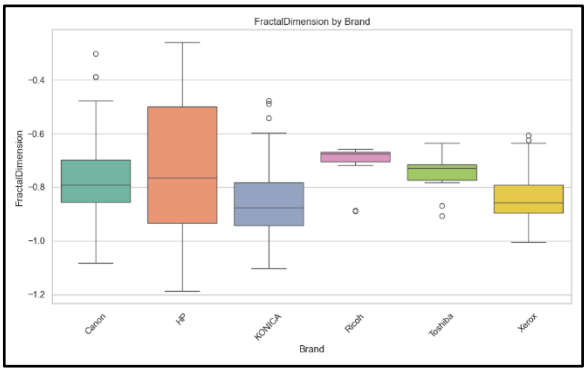
Laplacian variance emerged as a key indicator of edge sharpness across all brands. HP printers displayed the highest average variance (≈ 13,960), signifying noticeably sharper dot edges that are much more well-defined as compared to the other brands. In contrast, Toshiba showed the lowest variance (≈1,630), indicating smoother edges and a more diffuse dot-deposition process. Moreover, entropy analysis (Figure 1) reinforced these observations. Ricoh (≈0.91) and HP (≈0.87) recorded the highest entropy values, with greater complexity and randomness in the pixel intensity distributions across the printed document. Canon (≈0.38) and Toshiba (≈0.32) showed lower entropy, consistent with constant, repetitive dot structures.

The fractal dimension results (Figure 2) highlight the differences in structural complexity among the printer brands.



International Journal of Medical Toxicology & Forensic Medicine

Figure 1. Brand-wise distribution: (A) Laplacian Variance, (B) Entropy.



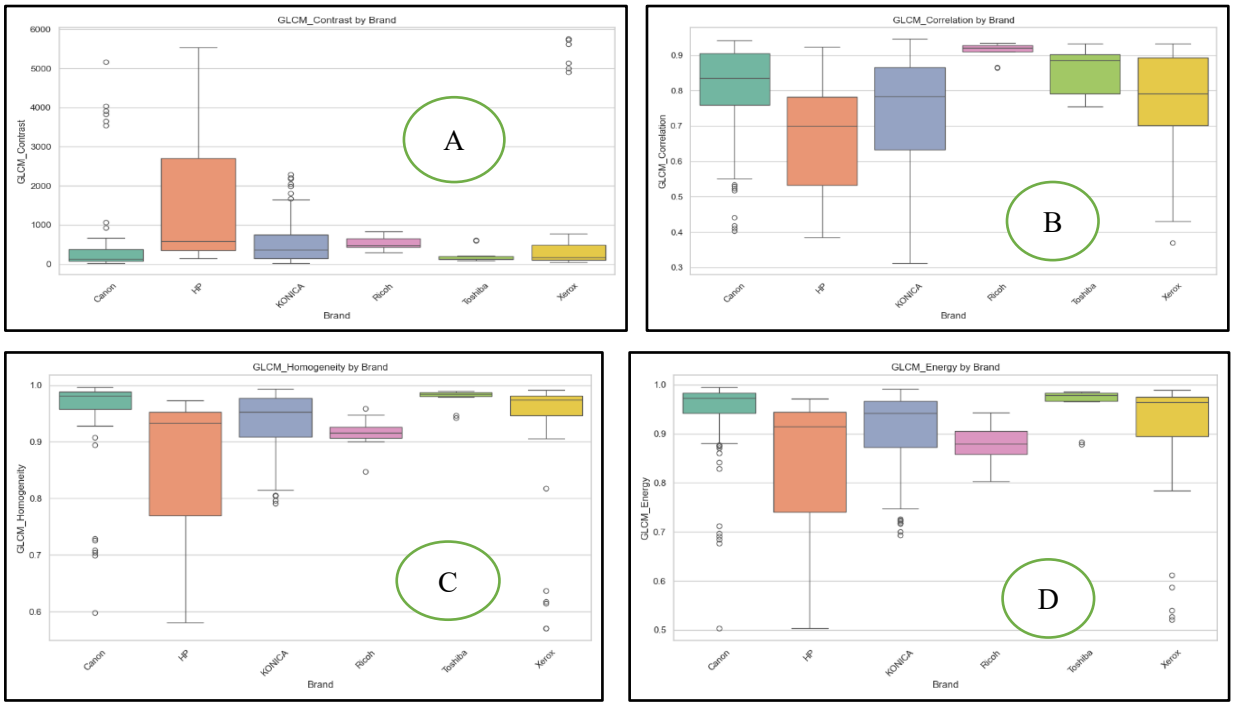
International Journal of Medical Toxicology & Forensic Medicine

Figure 2. Boxplot for Fractal Dimension values.

HP demonstrated a broad distribution ranging from ≈−1.2 to −0.3, indicating substantial variability in the self-similarity of the dot patterns. Ricoh showed tighter clustering around -0.85, suggesting more stable and uniform edge geometry across these dot patterns.

GLCM texture features added further discriminatory power. Contrast (Figure 3a) was highest for HP, with mean values above 1,400, and lowest for Toshiba (≈230), matching its smoother dot appearance. Correlation (Figure 3b) peaked for Ricoh (≈0.91), reflecting a strong linear dependency between neighboring pixels—a pattern far weaker in HP, which showed lower correlation (≈0.67). Homogeneity and Energy (Figure 3c–d) were highest for Toshiba and Canon (>0.95), indicating a highly regular, predictable dot-placement strategy compared with the more disordered texture seen in HP.

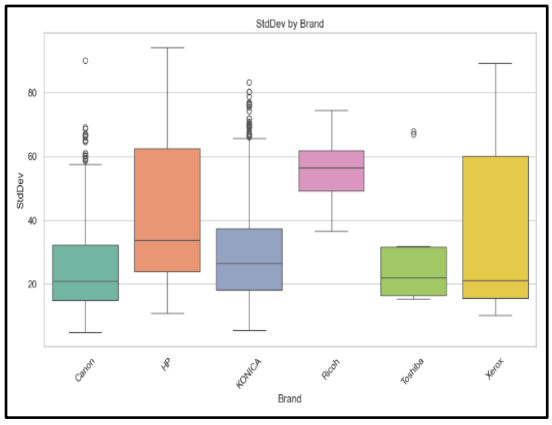
Also, pixel intensity distribution, measured by Standard Deviation (Figure 4), indicated that Ricoh printed outputs have the broadest dynamic range



International Journal of Medical Toxicology & Forensic Medicine

Figure 3. GLCM features across brands: (A) Contrast, (B) Correlation, (C) Homogeneity, (D) Energy.

(mean≈56), significantly higher than the narrow intensity spread observed in the brands Canon and Toshiba (≈28–30).



International Journal of
Medical Toxicology & Forensic Medicine

Figure 4. Standard deviation across brands.

Unsupervised Clustering and Separability Analysis

The t-SNE visualization (Figure 5) reduces the high-dimensional feature space to two dimensions. The plot reveals a semi-structured latent space, with distinct clusters also visible for brands with unique textures (HP and Ricoh). There was a significant overlap among the geometrically similar brands (Canon, Konica, and Xerox).



International Journal of
Medical Toxicology & Forensic Medicine

Figure 5. t-SNE visualization.

Table 1 confirms this overlap quantitatively: K-Means (k=6) produced a low ARI of 0.09 and a Silhouette Score of 0.33, showing the brands are not linearly separable in raw feature space, and DBSCAN could not resolve distinct clusters at all (Silhouette=0) given the uniform density of the overlapping brands. Together, these results indicate that the extracted features carry discriminative information, but simple proximity-based grouping cannot exploit it — motivating the move to non-linear supervised classifiers.

Table 1. Performance of unsupervised clustering algorithms.

Method	Clusters (k)	Silhouette	Davies-Bouldin Index	ARI
K-Means	6	0.33	0.94	0.09
GMM	6	0.20	1.51	-----
DBSCAN	3	-----	-----	-----

International Journal of
Medical Toxicology & Forensic Medicine

Supervised Classification Performance

Given these clustering limitations, we trained five supervised models to map the feature space to printer manufacturer labels; their performance on the independent test set is summarised in Table 2.

Table 2. Comparative performance of supervised classifiers.

Model	Accuracy (%)	Precision	Recall	F1-Score
Random Forest	73.9	0.72	0.74	0.71
SVM	71.7	0.74	0.72	0.67
MLP (Neural Net)	71.7	0.67	0.72	0.69
Decision Tree	67.4	0.68	0.67	0.67
k-NN	60.9	0.59	0.61	0.59

International Journal of
Medical Toxicology & Forensic Medicine

Random Forest was the most robust of the five, reaching the highest accuracy (73.9%) and F1 score (0.71); its ensemble structure captures the non-linear decision boundaries needed to separate geometrically similar brands such as Konica and Xerox. In contrast, SVM achieved the highest precision (0.74), indicating a more conservative approach with fewer false positives. However, this came at the expense of a lower F1 score (0.67), reflecting decreased recall performance for the minority classes.

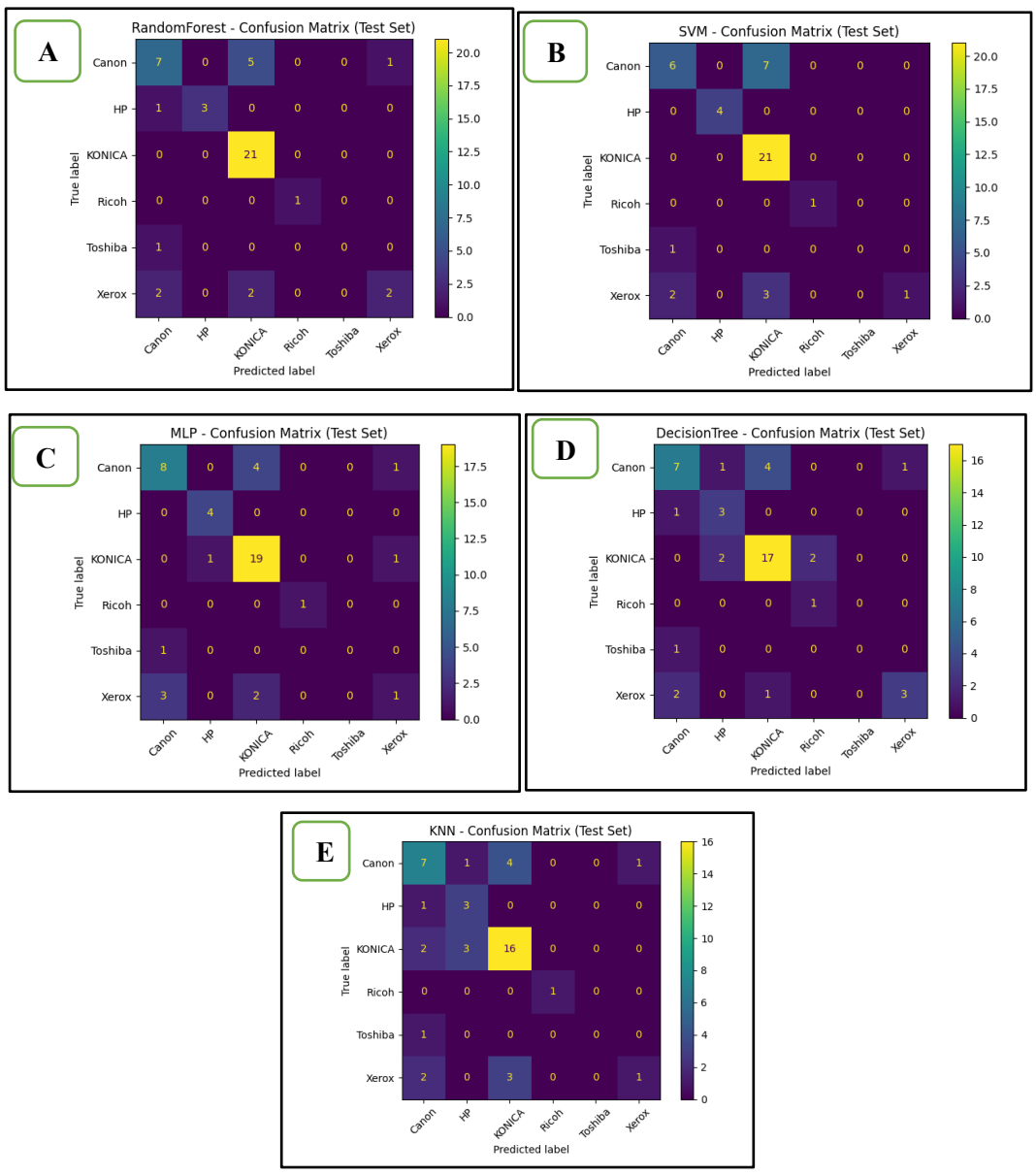
MLP, on the other hand, attained an equivalent frequency of 71.7% but demonstrated slightly lower precision (0.67) than both RF and SVM. This implied that this model was more sensitive to noise and overlapping features. K-NN performed the worst of all models for both accuracy (60.9%) and F1 score (0.59), underscoring the limitations of basic distance-based classifiers in separating such complex problems, as suggested by the t-SNE visualizations.

A closer examination of per-class performance (Figure 6) provides further insight into brand separation. Ricoh achieved near-perfect classification (precision=1.0, recall=1.0 across most models). HP also demonstrated almost perfect classification (precision=1.0, recall=1.0 in SVM). These brands exhibit distinctive high-entropy and very high-variance feature signatures, allowing them to be distinguished from the remaining classes. Konica, while achieving perfect recall (1.0) in both RF and SVM, exhibited reduced precision (≈0.68-0.75), indicating that samples from other brands were misclassified as Konica. This behavior can be attributed to class imbalance, in which Konica accounts for about 45% of the test samples, leading the models to favor it when faced with ambiguous feature patterns.

Classification performance deteriorated sharply for printer brands, especially those with limited representation in the training data. Toshiba was also misclassified across all models. Similarly, Xerox exhibited poor recall values, ranging from 17% to 33%, as predictions were often biased toward the dominant printer brands.

Discussion

This study demonstrated that the microscopic yellow dot patterns generated by color laser printers contain manufacturer-specific unique signatures for identification. The feature analysis further confirmed that physical toner characteristics, especially edge sharpness and pixel intensity distribution, differ significantly among manufacturers. The Laplacian



International Journal of Medical Toxicology & Forensic Medicine

Figure 6. Confusion matrix (A) Random forest (B) SVM (C) MLP (D) DT (E) KNN.

Variance was identified as a key discriminator: HP displayed high variance, indicating sharp, crisp yellow dots, whereas Toshiba and Canon demonstrated low variance. This observation aligns with research by Hamzehyan R et al. (2021) [15], which found that modern high-resolution printers often use frequency-modulated (FM) screening, which blurs the edges of toner particles. The results suggest that HP likely uses an amplitude-modulated (AM) screening process for its yellow dots, leading to sharper, noisier edges detected by the algorithm, while Toshiba's smoothness resembles the stochastic screening effects described in the literature [14, 31]. The findings of this study also support and validate the studies conducted by Salim and Abdalla (2019) [9]. They further reported that Ricoh employs a highly strict, grid-like dot placement that remains consistent despite scanning distortions. The statistical analysis further confirms that Ricoh samples demonstrated the highest GLCM Correlation (0.91), providing quantitative evidence of the structural regularity that they described. Moreover, the model's ability to identify Ricoh printers with near-perfect precision and recall of approximately 1.0 further reinforces the findings. The unsupervised clustering results were also evaluated to assess the natural separation of the brands without labels. K-means (k=6) failed to separate the brands, resulting in a low ARI, and DBSCAN failed to identify stable clusters. Isolated signatures also exist for brands like Ricoh and HP, while the feature spaces of Canon, Konica, and Toshiba overlap. The complexity of the feature space necessitated supervised learning models to significantly improve performance. The Random Forest classifier achieved the highest accuracy (73.9%) and F1 score (0.71), effectively capturing the non-linear relationships among the texture features considered. RF's performance over k-NN (Accuracy: 60.9%) is consistent with the study by Navarro et al. (2018) [30]. K-NN depends on Euclidean distance, which breaks down when classes overlap significantly, as discussed in the unsupervised analysis. In contrast, Random Forest uses hierarchical splitting, allowing it to separate geometrically similar brands like Konica and Xerox that perplexed the simpler models. Random Forest's strong performance over distance-based metrics, including k-NN (Accuracy: 60.9%), suggests that printer artifacts are not defined by simple global clusters but by complex hierarchical associations between features considered in this study.

A comprehensive review by Zhao et al. (2025) [2] records that most forensic studies utilize perfectly balanced datasets to report high accuracy metrics. In contrast, this study preserved the natural imbalance of a real forensic collection, where Konica was common, and Toshiba was rare. This imbalance is the primary driver of the performance gap between this study (74%) and the 90% or above accuracies reported by Phan-Ho et al. (2022) [32] and Darwish and Elgohary (2021) [33]. This finding serves as a significant stress test, showing that high accuracies reported in controlled literature may not fully translate to the "long-tail" distribution of real-world evidence without synthetic data amplification or cost-sensitive learning.

The overall accuracy of 73.9% should not be read as a single measure of fitness for casework. On the independent test set (n = 46), it corresponds to 34 correct attributions, with a 95% Wilson confidence interval of approximately 60–84%, so true performance may lie well away from the point estimate. More importantly, the aggregate conceals an uneven distribution of errors. Ricoh and HP were attributed almost perfectly, while Konica had perfect recall but precision of only about 0.68–0.75; Toshiba was never correctly identified, and Xerox recall was 17–33%. An examiner relying on the output would thus be most misled exactly where a brand is rare. The present model is not suitable as a stand-alone basis for opinion evidence. It can, however, play a limited investigative role as a screening or triage aid. Reaching a standard acceptable for casework would require larger, more balanced per-brand and per-model collections; an open-set reject option; calibrated confidence outputs with error rates reported for each brand; and blind validation on independent samples produced under varied paper, scanner, and print-setting conditions.

Conclusion

Color laser printers generate microscopic yellow dot patterns that retain textural and geometrical signatures, which also serve as a reliable steganographic feature in color laser printer examination. RF was the most effective classifier, achieving 73.9% accuracy and an F1 score of 0.71 on a challenging dataset. The feature analysis further establishes that Laplacian Variance and GLCM Entropy are the primary physical discriminators.

This study highlights the challenges of deploying automated forensic systems, including the use of realistic sample distributions, which reveals a significant vulnerability in identifying rare printer brands and a bias toward majority classes in forensic setups. The current model operates on a closed-set assumption, meaning it forces every questioned document to belong to one of the six known brands. The algorithm cannot directly reject an unknown printer, a major limitation of the study.

The findings support using MIC dot texture as a screening cue for manufacturer-level attribution, not for individualizing a specific device. In real casework,

the questioned document may differ in paper, toner state, print settings, age, or scanning and photocopying history, or may originate from a brand outside the six studied, none of which was tested here. At present, the approach should be regarded as a proof of concept for investigative triage, to be used alongside MIC decoding and conventional examinations rather than as stand-alone evidence. Practical deployment would need larger multi-site datasets, model-level labels, open-set rejection, robustness testing against paper, scanner, and copying variation, and blind-trial validation. Therefore, future work should focus on integrating morphological analysis to resolve ambiguity between similar brands. Furthermore, to mitigate class imbalance, algorithms designed to learn robust signatures for rare printers should be used with the limited sample size, thereby achieving much better accuracy for source identification.

Ethical Approval

Not applicable.

Acknowledgment

None.

Funding

No external funding was received for this research.

Conflicts of Interest

The authors report there are no competing interests to declare.

References

[1] Choi JH, Lee HY, Lee HK. Color laser printer forensics based on noisy features and a support vector machine classifier. Multimed Tools Appl. 2013;67(2):363-82. [DOI: 10.1007/s11042-011-0835-9]

[2] Zhao Y, Yang X, Chen X. A review of laser printer classification and identification. Forensic Sci Int. 2025;371:112480. [DOI: 10.1016/j.forsciint.2025.112480]

[3] Escher S, Strafe T. Robustness analysis of a passive printer identification scheme for halftone images. In: Proceedings of the IEEE International Conference on Image Processing (ICIP); 2017 Sep 17-20; Beijing, China. IEEE; 2017. p. 4357-61. [DOI: 10.1109/ICIP.2017.8297059]

[4] Mikkilineni AK, Chiang PJ, Ali GN, Chiu GTC, Allebach JP, Delp EJ III. Printer identification based on gray-level co-occurrence features for security and forensic applications. Proc SPIE. 2005;5681:430-40. [DOI: 10.1117/12.593796]

[5] Orhan B, Mao J, Sharma G. 2009 IEEE International Conference on Acoustics, Speech and Signal Processing: proceedings; 2009 Apr 19-24; Taipei, Taiwan. IEEE; 2009. [DOI: 10.1109/ICASSP.2009.5099042]

[6] Khanna N, Mikkilineni AK, Chiu GTC, Allebach JP, Delp EJ. Survey of scanner and printer forensics at Purdue University. Lect Notes Comput Sci. 2008;5158:22-34. [DOI: 10.1007/978-3-540-85303-9_3]

[7] Tsai MJ, Yuadi I, Tao YH. Decision-theoretic model to identify printed sources. Multimed Tools Appl. 2018;77(21):27543-87. [DOI: 10.1007/s11042-018-5938-0]

[8] Salim AS, Abdalla AA. Application of Adobe Photoshop CC 2018 for identifying the source of HP color laser printouts. Arab J Forensic Sci Forensic Med. 2019;1(2):1158-64. [DOI: 10.26735/16586794.2019.002]

[9] Salim AS, Abdalla AA. Determination of the identity and uniqueness of Ricoh-brand color laser printouts using Adobe Creative Cloud Photoshop 2018. Egypt J Forensic Sci. 2019;9(1):16. [DOI: 10.1186/s41935-019-0140-8]

[10] Salim AS, Abdalla AA. The state-of-the-art technique for determining the identity and uniqueness of common color laser printouts by coded dot matrix patterns. Arab J Forensic Sci Forensic Med. 2023;5(1):58-73. [DOI: 10.26735/RUGH9293]

[11] Ferreira A, Bondi L, Baroffio L, Bestagini P, Huang J, Dos Santos JA, et al. Data-driven feature characterization techniques for laser printer attribution. IEEE Trans Inf Forensics Secur. 2017;12(8):1860-73. [DOI: 10.1109/TIFS.2017.2692722]

[12] Choi JH, Lee HK, Lee HY, Suh YH. Color laser printer forensics with noise texture analysis. In: Proceedings of the 2010 ACM SIGMM Multimedia Security Workshop; 2010 Oct 26; Firenze, Italy. ACM; 2010. p. 19-24. [DOI:

10.1145/1854229.1854235]

[13] Shang S, Kong X. Printer and scanner forensics. In: Handbook of digital forensics of multimedia data and devices. Chichester: John Wiley & Sons; 2015. p. 375-410. [DOI: 10.1002/9781118705773.ch10]

[14] Chiang PJ, Ali GN, Mikkilineni AK, Delp EJ, Allebach JP, Chiu GTC. Extrinsic signatures embedding and detection for information hiding and secure printing in electrophotography. In: Proceedings of the American Control Conference; 2006 June 14-16; Minneapolis, MN, USA. IEEE; 2006. p. 2539-44. [DOI: 10.1109/ACC.2006.1656557]

[15] Hamzehyan R, Razzazi F, Behrad A. Printer source identification by feature modeling in the total variable printer space. J Forensic Sci. 2021;66(6):2261-73. [DOI: 10.1111/1556-4029.14822]

[16] Devi MU. Statistical measures for differentiation of photocopy from print technology: forensic perspective. Int J Comput Appl. 2014;105(1):1-7. [DOI: 10.5120/18328-9366]

[17] Tsai MJ, Tao YH, Yuadi I. Deep learning for printed document source identification. Signal Process Image Commun. 2019;70:184-98. [DOI: 10.1016/j.image.2018.09.006]

[18] Liu N, Zhang L, Chen C. A preliminary study on sources of banding artifacts for the identification of monochromatic laser printers. J Forensic Sci. 2017;62(6):1542-53. [DOI: 10.1111/1556-4029.13487]

[19] Arora S, Sarin RK, Puri P. Machine identification codes of color laser printers: revisiting privacy and security. Ind J Electrical Eng Comput Sci. 2025;40(1):137-45. [DOI: 10.11591/ijeecs.v40.i1.pp137-145]

[20] Salim AS, Abdalla AAM. Application of Adobe Photoshop CC 2018 for identifying color laser printer source of Xerox brand. Egypt J Forensic Sci. 2018;8(1):52. [DOI: 10.1186/s41935-018-0076-4]

[21] Verhoeven G, Karel W, Štuhec S, Doneus M, Trinks I, Pfeifer N. Mind your grey tones: examining the influence of decolorization methods on interest point extraction and matching for architectural image-based modeling. Int Arch Photogramm Remote Sens Spat Inf Sci. 2015;XL-5-W4:307-14. [DOI: 10.5194/isprsarchives-XL-5-W4-307-2015]

[22] Yang P, Baracchi D, Ni R, Zhao Y, Argenti F, Piva A. A survey of deep learning-based source image forensics. J Imaging. 2020;6(9):94. [DOI: 10.3390/jimaging6090094]

[23] Lee HY, Choi JH. Printer identification using intrinsic noise. In: Proceedings of the 5th International Conference on Ubiquitous Information Technologies and Applications; 2010 Dec 16-18; Sanya, China. IEEE; 2010. [DOI: 10.1109/ICUT.2010.5677893]

[24] Jiang ZF, Zhang QH, Wang YC, Liu YL, Zhao YW, Hao YY, et al. Prediction of laser printers and cartridges based on three-dimensional profiles via discrimination analysis. Forensic Sci Int. 2024;363:112186. [DOI: 10.1016/j.forsciint.2024.112186]

[25] Gál L, Oravec M, Gemeiner P, Čeppan M. Principal component analysis for the forensic discrimination of black inkjet inks based on the Vis-NIR fiber optics reflection spectra. Forensic Sci Int. 2015;257:285-92. [DOI: 10.1016/j.forsciint.2015.09.011]

[26] Karna A, Gibert K. Automatic identification of the number of clusters in hierarchical clustering. Neural Comput Appl. 2022;34(1):119-34. [DOI: 10.1007/s00521-021-05873-3]

[27] Kim DG, Hou JU, Lee HK. Learning deep features for source color laser printer identification based on cascaded learning. Neurocomputing. 2019;365:219-28. [DOI: 10.1016/j.neucom.2019.07.084]

[28] Elkasrawi S, Shafait F. Printer identification using supervised learning for document forgery detection. In: Proceedings of the 11th IAPR International Workshop on Document Analysis Systems (DAS); 2014 Apr 7-10; Tours, France. IEEE; 2014. p. 146-50. [DOI: 10.1109/DAS.2014.23]

[29] El Abady NF, Zayed HH, Taha M. An efficient source printer identification model using convolution neural network (SPI-CNN). Int J Adv Comput Sci Appl. 2023;14(3):745-53. [DOI: 10.14569/IJACSA.2023.0140386]

[30] Navarro LC, Navarro AKW, Rocha A, Dahab R. Connecting the dots: toward accountable

machine-learning printer attribution methods. J Vis Commun Image Represent. 2018;53:257-72. [DOI: 10.1016/j.jvcir.2018.04.002]

[31] Jamieson A, Moenssens AA, editors. Wiley encyclopedia of forensic science. Chichester: John Wiley & Sons; 2009. [DOI: 10.1002/9780470061589]

[32] Phan-Ho AT, Nguyen QT, Patrix J, Verny J. Source printer identification with microscopic printing using deep learning. IFAC-PapersOnLine. 2022;55(13):1177-82. [DOI: 10.1016/j.ifacol.2022.10.395]

[33] Darwish SM, Elgohary HM. Building an expert system for printer forensics: a new printer identification model based on niching genetic algorithm. Expert Syst. 2021;38(4):e12624. [DOI: 10.1111/exsy.12624]