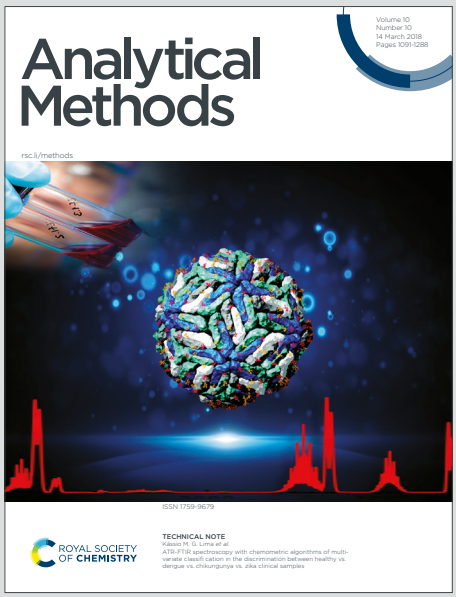
Analytical Methods

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: X. Lu, J. Zhang, J. Wu, X. Zhang, H. Ren, H. Chen, K. Ma and F. Li, *Anal. Methods*, 2025, DOI: 10.1039/D5AY00526D.



ROYAL SOCIETY OF CHEMISTRY

rsc.li/methods

View Article Online
DOI: 10.1039/D5AY00526D

ARTICLE

Nondestructive and Rapid Identification of Stamp Pad Ink Based on Hyperspectral Imaging and Extreme Learning Machine

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Lu Xiaoquan,[a] Zhang Jianqiang,[b] Wu Jiaquana,[a] Zhang Xinyu,[a] Ren Huihui,[a] Chen Hang,[a] Ma Kun*[a] and LI Fan*[b]

The examination of stamp pad ink in questioned documents serves as a crucial scientific basis for forensic authentication. This study presents a novel rapid classification framework integrating Hyperspectral Imaging (HSI) and Extreme Learning Machine (ELM) to address the challenges of timeliness and accuracy in nondestructive ink detection. A total of 24 photosensitive ink samples from 21 brands were collected, generating 72 standardized stamped impressions. Spectral-spatial data were acquired using an HSI system (400-1000 nm, 5 nm spectral resolution). preprocessed by Multiplicative Scatter Correction (MSC) to mitigate substrate interference. Experimental results demonstrate that the HSI-MSC-ELM framework achieved an accuracy of 98.38% on the test set without feature dimensionality reduction (full 121 spectral bands), outperforming Random Forest (RF) by 4.63% and Backpropagation Neural Network (BPNN) by 6.34%. Crucially, the detection time was only 1.59 seconds – 28× faster than RF (45.90 s) and 285× faster than BPNN (453.36 s). This approach provides a simple, nondestructive, and efficient solution for forensic document examination, with potential to replace traditional destructive techniques.

Introduction

In the field of forensic science, stamp impressions serve as critical physical evidence in document authentication, and their examination directly affects the legitimacy and reliability of the chain of evidence. Advances in forgery techniques and materials science have necessitated a shift in stamp impression analysis from traditional morphological comparison to a comprehensive technological framework that integrates physical and chemical analysis to improve identification accuracy. In forensic practice, beyond verifying stamp consistency based on morphological features, forensic examiners must also determine whether the impressions in different documents originate from the same type of stamp pad ink.

Given the short lifespan of stamp pad ink paste, which fades easily and hardens quickly, liquid-based stamp pad ink has become the most widely used alternative. Identifying the composition and type of stamp pad ink plays a crucial role in forensic investigations, aiding both criminal investigations and legal proceedings. Commercially available stamp pad inks are categorized into three main types based on their manufacturing process: ordinary ink, atomic ink, and photosensitive ink.[1] The primary chemical components of stamp pad ink include pigments, resins, and solvents.[2] Due to differences in manufacturers and production processes, commercially available inks contain varying compositions and proportions of these components.

To date, numerous studies on stamp pad ink identification have been conducted, and the primary technical approaches can be categorized into spectroscopic and chromatographic analysis. Spectroscopic techniques, including Ultraviolet-Visible Spectroscopy (UV-Vis),[1] Raman spectroscopy,[2,6] fluorescence spectroscopy,[3] and infrared spectroscopy,[4,5] are characterized by rapid detection, minimal sample requirements, and nondestructive testing. However, these methods still face technical limitations in acquiring comprehensive spectral information, particularly when attempting to distinguish ink samples with similar chemical compositions. In contrast, chromatographic techniques, such as High-Performance Liquid Chromatography (HPLC),[7-9] Thin-Layer Chromatography (TLC),[10] and other chromatographic methods,[11] offer high detection accuracy but are inherently destructive to the sample.

Hyperspectral Imaging (HSI) is an effective nondestructive ink detection technique that integrates spatial imaging and spectral analysis. Compared to traditional methods, HSI offers higher spectral resolution, broader spectral range, and richer data representation. The core principle of HSI involves the simultaneous acquisition of spectral information, full-band imaging, and radiation intensity data within a specified wavelength range, thereby providing multi-dimensional data support for ink identification. This technology has been applied in multiple fields, including ecological monitoring,[12,13] agricultural and food safety analysis,[14,15] medical diagnostics,[16] cultural heritage preservation,[17] and forensic investigations.[18] In forensic document examination, Reed et al.[19] demonstrated that HSI achieved different levels of discrimination accuracy for

[a] College of Science, Kunming University of Science and Technology, Kunming, 650500,China. E-mail: 453937936@qq.com
[b] College of Criminal Investigation, Yunnan Police College, Kunming, 650106, China. E-mail: lifan1010@sina.com

red (1.00), blue (0.90), and black gel ink (0.40), with an overall discrimination accuracy of 0.76. Devassy et al.[20] conducted a systematic comparison of Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE) in hyperspectral ink data analysis, evaluating their effectiveness in dimensionality reduction and visualization, providing key methodological insights for the application of HSI in forensic science. Wang et al.[21] employed HSI for ink data collection and utilized One-Dimensional Convolutional Neural Networks (1D-CNN) and Backpropagation Neural Networks (BPNN) to classify 20 types of stamp pad ink. Their findings indicate that 1D-CNN outperformed BPNN in terms of classification stability and efficiency, achieving training and validation loss values of 0.068 and 0.075, and final classification accuracies of 98.30% and 97.94%, respectively. Despite these advancements, existing hyperspectral ink analysis methods still face several key technical challenges: overall classification accuracy remains suboptimal and requires further improvement; existing methods (e.g., CNN-based models) achieve high accuracy but suffer from excessive computational costs (single training sessions > 75 min and testing times > 25 min), thereby limiting practical forensic applications.

To address these challenges, this study proposed a novel rapid and nondestructive ink identification framework by integrating HSI, Multiplicative Scatter Correction (MSC), and Extreme Learning Machine (ELM). The goal is to enhance classification accuracy while significantly reducing computational complexity. Experimental results demonstrate that compared to BPNN and Random Forest (RF) models, the ELM model achieves superior performance, allowing for fast, nondestructive, and highly accurate identification of stamp pad ink types.

Experiment and Methods

Experimental Materials

This study collected 24 photosensitive stamp pad ink samples from 21 different brands, covering 24 models (Table 1). A total of 24 new, uninked stamps (diameter: 38 mm) were used, each uniformly engraved with the phrase “Experimental Sample Seal”. The printing medium was M&G A4 copy paper (manufactured by Shanghai M&G Stationery Co., Ltd.; Model: APYVQ958; dimensions = 210 mm × 297 mm; weight = 80 g/m²). Before acquiring HSI data, each stamp pad ink sample was individually injected into its corresponding pre-numbered stamp (No.1-24). Each stamp was then pressed three times consecutively, generating a total of 72 standardized stamp impressions. To ensure the stability and reliability of subsequent analyses, all stamp samples were left at room temperature for two weeks to allow the ink to fully cure and settle.

Table 1 Photosensitive ink brands and models

ID	Brand and model	ID	Brand and model
1	Deyin	16	Deli NO.9040
2	Zeyuan	17	Jingyi
3	Noch	18	WZ-3630
4	Caiyou	19	Aosheng
5	Hachidai	20	Mingchuang
6	Guancheng	21	Qixin B3722
7	Shengbei	22	M&G AYZ97508
8	Qixin	23	DuPont
9	Baiyiyuan	24	Print Up
10	Shachihata		
11	Kingdee		
12	Yaxin		
13	Arxin		
14	Deli NO.9879		
15	M&G 97509		

Image Acquisition and Processing

This study utilized the SHIS-N220 hyperspectral imaging system (Shenzhen Zhongda Ruihe Technology Co., Ltd.) for data acquisition, with system configuration and parameter settings strictly adhering to spectral analysis standards. The imaging unit was equipped with a 2048 × 2046 pixel area-array Charge-Coupled Device (CCD) detector, covering a spectral response range of 400-1000 nm in the Visible to Near-Infrared (VNIR) region, with a spectral resolution of 1 nm and a sampling interval of 5 nm, capturing 121 continuous spectral bands. The system employed a four-light-source symmetrical setup (50 W halogen lamps at a 45° incident angle) with a vertical imaging optical path to ensure uniform illumination across the sample surface (Fig. 1).

The data acquisition process consisted of three key steps: environmental control, equipment calibration, and standardized scanning. Firstly, the HSI system was preheated for 10 min in a darkroom environment to ensure stable radiation intensity; Secondly, full-band radiometric calibration was performed using a Spectralon® 99% reflectance standard panel to eliminate environmental light scattering and detector dark current noise; Thirdly, a high-resolution hyperspectral cube dataset was acquired at a constant scanning speed of 0.5 mm/s, achieving a spatial resolution of 30 μm/pixel.

For dataset development, the ENVI 5.3 software platform was used to execute standardized data processing. In the densely pigmented area of each stamp impression (excluding edge diffusion regions), 50 independent Regions of Interest (ROI) were manually selected (20 × 20 pixels each, with spatial distribution and selection dimensions in Fig. 2a-b). The spectral reflectance mean values of all pixels within the ROI for the 400-1000 nm wavelength range were extracted in batch mode, forming a structured dataset of 3,600 feature vectors (2,600 for training, 1,300 for testing). 30 regions of interest from the hyperspectral images of the unused unprinted paper were selected as baseline references. A strict dataset partitioning strategy was implemented to ensure that training and testing samples remained physically isolated.

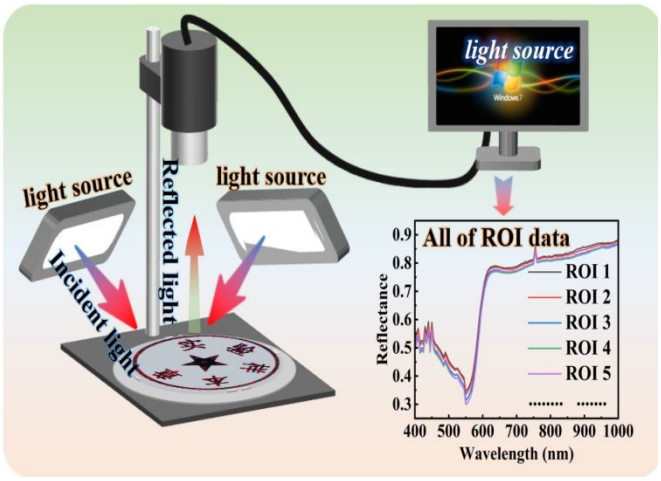


Fig. 1 HSI system configuration schematic.

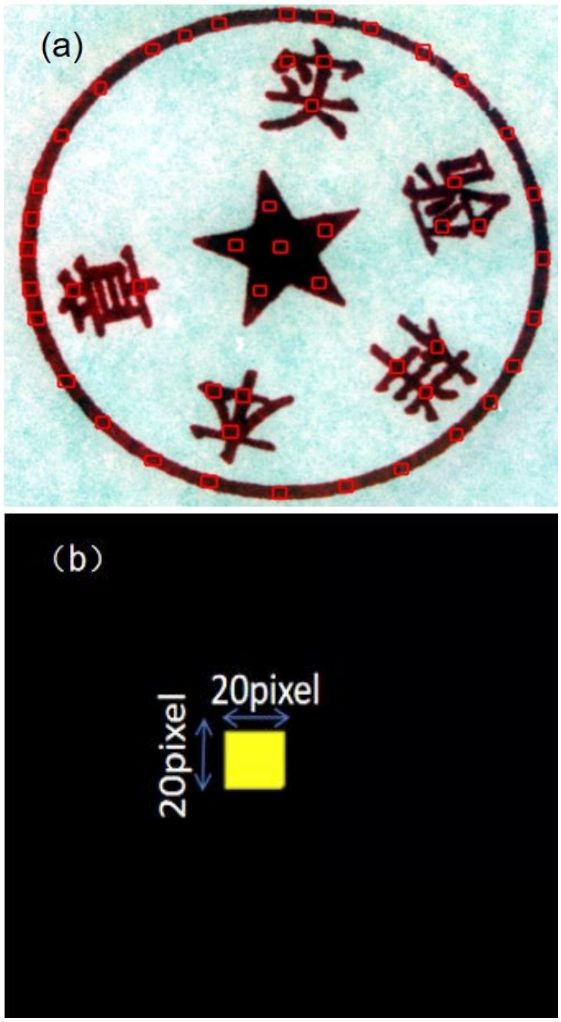


Fig. 2 Schematic of samples and their spectral selection: (a) Schematic of ROI for print selection; (b) Sample of (a).

Model Development

BPNN. BPNN is a multi-layer feedforward neural network designed for training nonlinear differentiable function weights.[22] It consists of two primary processes: forward propagation of information and backpropagation of errors. A typical BPNN comprises several interconnected layers, including an input layer, multiple hidden layers, and an output layer. The dataset is fed into the input layer, where the hidden layers compute the inner product of the weight matrix and bias coefficients. The Rectified Linear Unit (ReLU) function is employed as the activation function, and a dropout layer is incorporated to prevent overfitting. Finally, the SoftMax function is applied at the output layer for classification of stamp pad ink types.

(1) Output of hidden layer H_i

$$H_j = g\left(\sum_{i=1}^{n} w_{ij}x_i + a_j\right) \quad (1)$$

where a_j is the bias vector from the input layer to the hidden layer; w_{ij} is the weight matrix from the input layer to the hidden layer; and g is the activation function.

(2) Output of output layer

$$O_k = \sum_{j=1}^{l} H_j w_{jk} + b_k \quad (2)$$

where w_{jk} is the weight matrix connecting the hidden layer to the output layer; b_k is the bias vector of the output layer; and l is the hidden layer output.

(3) Error calculation

$$E = \frac{1}{2}\sum_{k=1}^{m} (Y_K - O_K)^2 \quad (3)$$

where m is the number of neurons in the input layer; Y_K is the expected output; and k = 1, ..., m.

The BPNN model architecture is designed by selecting key hyperparameters, including the number of fully connected layers, learning rate, and the number of hidden layer neurons. To ensure comparability of training results and enhance experimental efficiency, the number of training iterations for BPNN was fixed at 1,000 epochs in this study.

RF. Since its introduction by Breiman in 2001,[23] the RF algorithm has become one of the most representative ensemble learning techniques in the field of machine learning. RF constructs models by integrating multiple weak classifiers,[24] typically Classification and Regression Trees (CART), which significantly enhance prediction accuracy and generalization performance. When the dependent variable in a dataset is categorical, RF applies CART decision trees as classifiers, effectively handling classification tasks.

The RF algorithm employs two key criteria for decision-making: information gain and Gini index.

(1) Information gain: The information gain $g(X, A)$ of a feature A in a training dataset X is defined as the difference between the empirical entropy $H(X)$ of the dataset X and the empirical conditional entropy $H(X \mid A)$ of the dataset given a specific feature A. $g(X, A)$ can be expressed as follows:

$$g(X, A) = H(X) - H(X \mid A) \quad (4)$$

(2) Gini index: The RF algorithm utilizes the Gini Index to determine the optimal splitting attribute during decision tree development. The Gini index measures the impurity of a dataset, where a lower Gini Index indicates higher dataset purity after a split.

$$Gini(P) = 1 - \sum_{i=1}^{i} P_i^{\ 2} \quad (5)$$

where i is the number of sample classes and P_i is the probability of a sample belonging to class i.

The RF algorithm architecture is primarily optimized by tuning two key hyperparameters: the number of decision trees in the ensemble learning framework and the minimum number of samples per leaf node, which controls tree depth and model complexity. To establish a reproducible model evaluation framework, the experiment fixed the number of decision trees at 50 as a baseline configuration, with a Bootstrap sampling rate of 0.8. Additionally, a grid search strategy was employed to explore the effect of varying the minimum sample size per leaf node (1-20) on model performance.

ELM. The ELM algorithm was proposed by Huang et al.[25] One of its key characteristics is that the weights of the hidden layer nodes are either randomly assigned or manually set and do not require iterative updates. Theoretically, this approach enables extremely fast learning while maintaining strong generalization performance. Compared to conventional neural networks, ELM can perform classification and regression tasks more efficiently and stably.[26]

Let X represents the input layer dataset, containing elements x1, …, xd, H represents the hidden layer dataset, containing elements h1, h2, …, hL, and Y represents the output layer dataset, containing elements y1, …, yk, the relationship among these datasets can be expressed as follows:

$$H = \omega X + b \quad (6)$$

$$Y = \beta H \quad (7)$$

where ω is the weight vector connecting the input layer to the hidden layer nodes; b is the bias vector for the hidden layer nodes; and β is the weight vector connecting the hidden layer to the output layer.

The ELM primarily selects the training set randomly. It does not require knowledge of the intermediate hidden layer dataset; it only needs to establish the corresponding relationships to obtain the output dataset.[27,28] This approach eliminates the need for iterative weight updates, significantly reducing computational complexity. Therefore, ELM is faster and more accurate than traditional neural network methods.

$$Y = \beta g(\Omega X + b) \quad (8)$$

where $g(x)$ is the activation function in the hidden layer; $g(\Omega X + b)$ is the output of the hidden layer, where xis the input layer dataset; Y is the output layer dataset; Ω is the weight vector from the input layer to the hidden layer; b is the bias vector for the hidden layer; and β is the weight vector from the hidden layer to the output layer (Fig. 3).

In the hyperparameter optimization of the ELM algorithm, this study focused on tuning the key structural parameter—the number of hidden layer nodes. To establish a standardized evaluation framework, the experiment adopted a fixed architecture using the Sigmoid activation function, where weights between the input and output layers were orthogonally initialized. A five-fold cross-validation strategy was employed to investigate the effect of varying the number of hidden nodes (10-600 with a step size of 20) on the generalization performance of the model.

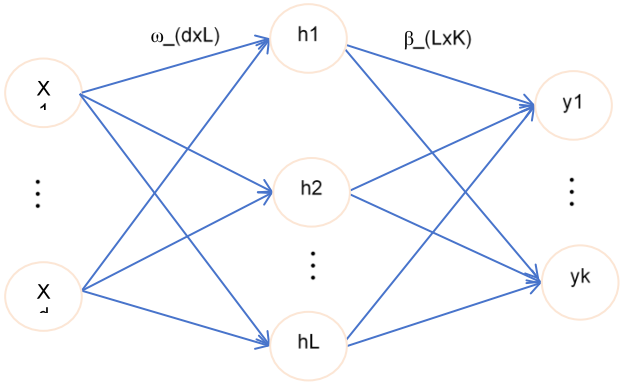


Fig. 3 ELM network structure diagram.

Measures of classification performance

As a visualization tool, the confusion matrix is not only used to evaluate the accuracy of supervised learning but also serves as an assessment metric for unsupervised learning. Additionally, the confusion matrix provides a detailed breakdown of classification performance, highlighting the accuracy of predictions for each category. Fig. 4 illustrates the basic structure of a confusion matrix. This study employed key evaluation metrics, including Accuracy (Acc), Precision (PC), Sensitivity (SN), Specificity (SP), and F1-score, with the corresponding calculations defined as follows:

$$Acc = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \quad (9)$$

$$PC = \frac{TP}{TP + FP} \quad (10)$$

$$SN = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad (11)$$

$$SP = \frac{TN}{TN + FP} \quad (12)$$

$$F1-score=\frac{2\times \text{Precision}\times \text{Sensitivity}}{\text{Precision}+\text{Sensitivity}} \quad (13)$$

where True Positive (TP) is the number of correctly classified positive samples; True Negative (TN) is the number of correctly classified negative samples; False Positive (FP) is the number of negative samples incorrectly classified as positive; and False Negative (FN) is the number of positive samples incorrectly classified as negative.[29]

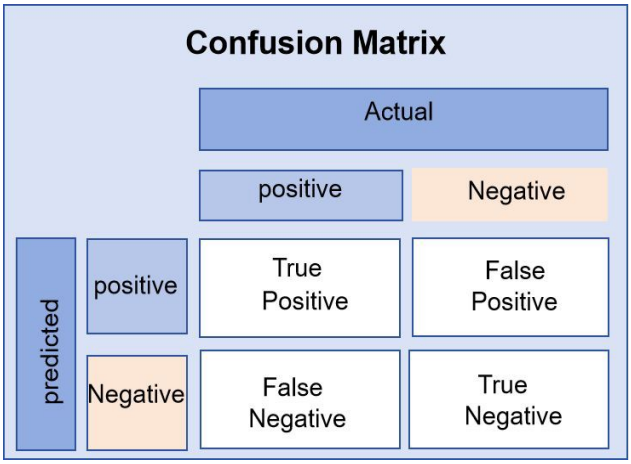


Fig. 4 Basic form of confusion matrix.

Results and Discussion

Hyperspectral Spectral Analysis

The raw reflectance spectrum of each pixel in the hyperspectral stamp pad ink image is shown in Figure 5. As evidenced by the spectral graph, the uppermost curve characterizes the high-reflectance properties of the paper substrate. Within the 400–650 nm waveband, a pronounced spectral distinction emerges between the paper and forensic stamp ink. At approximately 550 nm, the paper's reflectance exhibits a sharp surge to near-unity levels (≈1.0), this reflectance persisting at near-saturation stability throughout longer wavelengths. In contrast, The reflectance spectra of the 24 stamp pad inks showed a rapid downward trend in the wavelength range of 450-550 nm, with a small absorption peak at 450 nm. In the range of 550-670 nm, the absorption peaks increase dramatically, while after 670 nm, the spectral curve remains relatively flat with minimal variation. These results show that the chemical composition of different brands of stamp pad inks is highly similar, resulting in a highly consistent reflectance spectral pattern. Given the paper's forensically-significant reflectance in the 400-650 nm spectrum – establishing a non-bridgeable divergence in optical responses from stamp ink – this study's methodology applies spectral substrate filtration to enable exclusive focus on forensic ink spectral modeling. This similarity presents a significant challenge for direct visual classification, hence the need for hyperspectral data analysis using machine learning.

Preprocessing

View Article Online
DOI: 10.1039/D5AY00526D

The raw spectral data contain both sample-specific information and noise components. Preprocessing operations are essential to reduce noise impacts and enhance model accuracy. This study employed multiple spectral preprocessing techniques to construct the stamp pad ink classification model, using raw spectral data (NO processing) as the control group. The study integrated four preprocessing techniques: Savitzky-Golay first derivative (SG+1D), Savitzky-Golay second derivative (SG+2D), Standard Normal Variate (SNV), and MSC. Among these, MSC effectively corrects the scattering effects in individual sample spectra, eliminating baseline shifts and intensity distortions while significantly improving the resolution of absorption features. SNV applies mean centering and standard deviation normalization ($z=(x-\mu)/\sigma$), which suppresses systematic errors caused by sample-to-sample scattering differences. Derivative-based methods employ a 15 nm window width and a second-order polynomial fit, where first- and second-order derivatives help in baseline drift suppression and overlapping peak resolution enhancement. These preprocessing techniques remove non-chemical spectral variations induced by light scattering, laying the foundation for building high-precision classification models.

To establish the optimal preprocessing method, various preprocessing techniques were tested before developing the quantitative classification models. The BPNN model validation results for different preprocessing methods are summarized in Table 2, where the BPNN algorithm parameters were maintained as follows: training iterations = 1000, hidden neurons = 30, learning rate = 0.001, and a three-layer fully connected architecture with 121 neurons in the first layer, 30 neurons in the second layer, and 24 neurons in the final output layer. In Table 2, when using MSC preprocessing, the training and test set accuracies reached 97.58% and 92.04%, respectively (runtime = 453.36s), achieving the highest test set accuracy among all preprocessing techniques. Consequently, MSC preprocessing was selected as the standard preprocessing operation for all subsequent model training procedures. Fig. 6 illustrates the MSC-preprocessed spectral results.

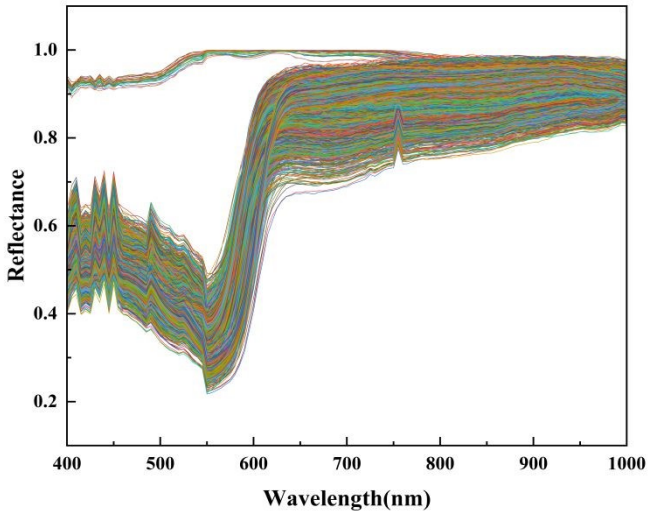


Fig. 5 Original data Spectral Curve Graph.

View Article Online
DOI: 10.1039/D5AY00526D

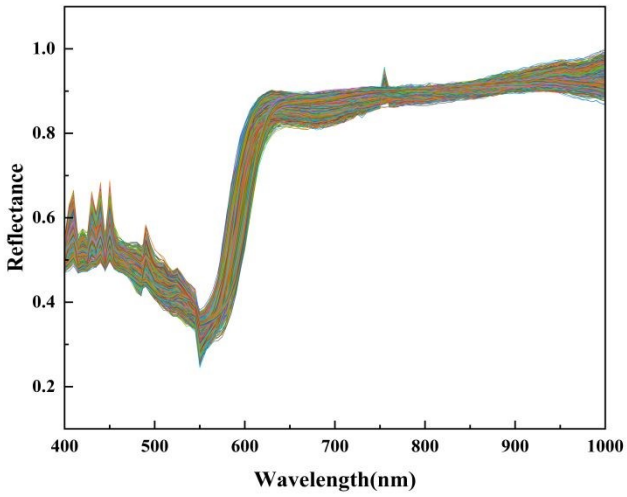


Fig. 6 MSC-preprocessed data Spectral Curve Graph.

Table 2 Accuracy of different preprocessing methods based on BP algorithm

Algorithm	Preprocessing method	Training set accuracy/%	Testing set accuracy /%
BP	NO	92.94%	79.07%
	SG+1D	72.42%	58.80%
	SG+2D	78.61%	65.56%
	SNV	91.39%	85.09%
	MSC	97.58%	92.04%

RF Training Model Results

Experimental results indicate that the number of leaf nodes significantly impacts model complexity and generalization ability (Fig. 7). As the number of leaf nodes increases from 1 to 7, the testing set accuracy slightly decreases from 96.04% to 95.71% (Δ = 0.33 percentage points), while the training accuracy drops from 100% to 98.92%, demonstrating that moderately limiting node splitting can effectively suppress overfitting. However, when the leaf node count further increases to 20, the testing set accuracy declines to 93.60% (a reduction of 2.11%), and the training accuracy simultaneously drops to 97.35%, indicating that an excessive number of leaf nodes leads to model performance degradation. These results confirm that the number of leaf nodes plays a crucial role in controlling decision boundary complexity. Specifically, setting the leaf node count to 7 achieves an optimal balance between model sensitivity and robustness.

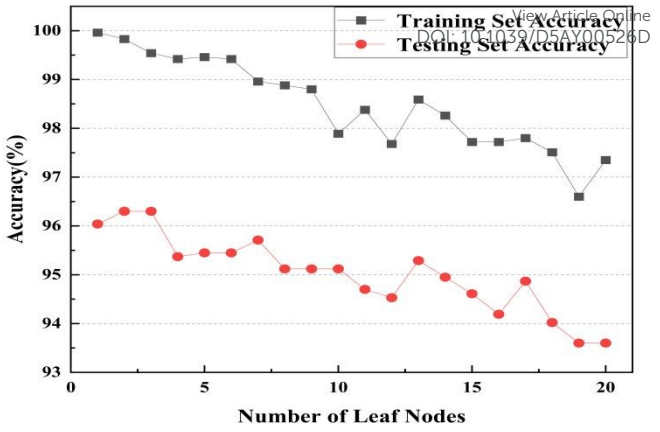


Fig. 7 Effect of number of leaf nodes on model accuracy.

Fig. 8 illustrates the effect of decision tree count on Out-of-Bag (OOB) error with leaf node size fixed at 7 samples (minleaf=7). The out-of-bag error is the average error for each training sample calculated using predictions from the trees that do not contain it in their bootstrap sample. This constrained leaf node configuration demonstrates that the number of decision trees has a significant nonlinear effect on model convergence. As the number of trees increases from 1 to 15, the OOB error decreases exponentially from 0.6310 to 0.05223, representing a 91.72% reduction, which highlights the improvement in prediction consistency achieved through ensemble learning under fixed leaf node conditions. However, once the number of trees exceeds 15, the error fluctuation range narrows to ±0.005, with the OOB error decreasing slightly from 0.05223 to 0.04353 between 15 and 30 trees, confirming the marginal benefit saturation effect of RF with minleaf=7. These results indicate that setting the number of decision trees to 15 combined with leaf node control (minleaf=7) ensures an optimal trade-off between computational efficiency (training time = 45.90s) and generalization performance (OOB error = 0.0435), achieving Pareto optimality.

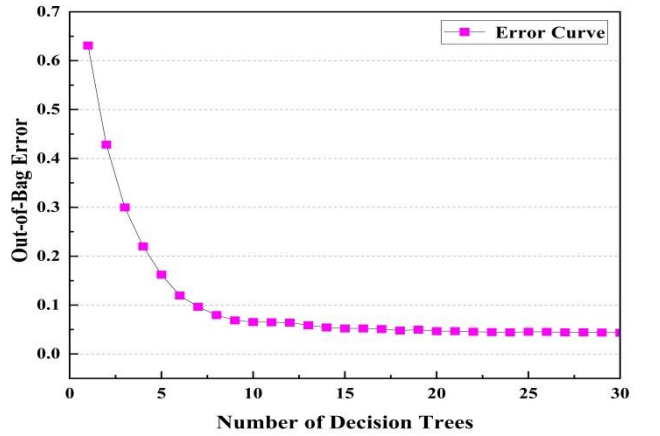


Fig. 8 Randomized forest out-of-bag error versus number of trees curve.

Feature importance scores were calculated based on the mean decrease in Gini impurity.30 For each decision tree in the Random Forest, the Gini impurity reduction induced by splitting at a specific feature (wavelength) was recorded. The final importance score for feature i is the averaged reduction across all trees:

$$\mathrm{Im}\,portance_i = \frac{1}{N_{trees}} \sum_{t=1}^{N_{trees}} \Delta Gini(i,t) \tag{14}$$

where ΔGini(i,t) is the Gini decrease by feature i in tree t, and Ntrees = 15. Scores were normalized to a -1–4 scale for visualization in Fig. 9. Each feature index maps to a discrete spectral sampling point across the 400-1000 nm spectral range, with a uniform spectral resolution of 5 nm, generating 121 wavelength-specific channels. Fig. 9 depicts the feature importance scores, revealing that the 400-600 nm wavelength range (feature indices 1, 3, 36, etc.) significantly contributes to classification. Among them, index 1 (3.62), index 3 (1.76), and index 36 (1.74) correspond to absorption peaks near 400 nm, 500 nm, and 575 nm, respectively, aligning with the electronic transition characteristics of benzotriazole based photosensitie-rs. Additionally, the 800-1000 nm wavelength range (indices 90 and 107) also exhibits relatively high scores (0.70, 0.78), correlating with the synergistic absorption of phthalocyanine-based dyes. In contrast, the 640-765 nm range has an average importance score of only 0.018, with over 70% of its feature scores being zero, and index 80 even showing a negative score (-0.27), indicating that this spectral range contributes little to classification and may introduce noise. By applying a feature importance threshold of >0.40, the top 36 most informative features were selected, reducing the data dimensionality by 70.2% (121 → 36). Despite this reduction, the testing set accuracy decreased by 2.02 percentage points (94.87% → 92.85%), while model inference speed improved by approximately 1.5 times (45.90 s → 30.21 s). These results underscore the dual value of focusing on key spectral regions, facilitating both efficient dimensionality reduction and performance optimization in hyperspectral ink classification.

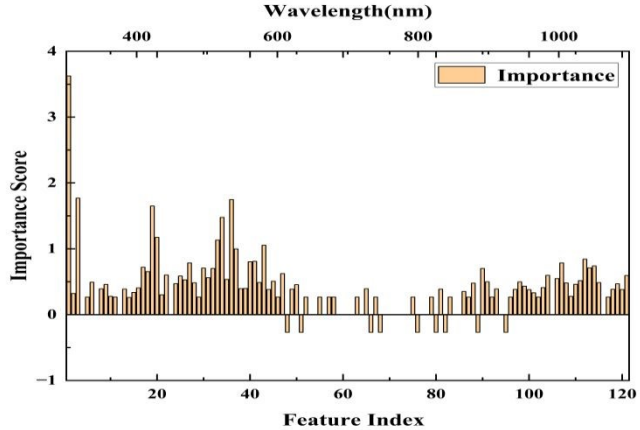


Fig. 9 Feature importance distribution over 400-1000 nm spectral range.

After parameter optimization, the RF model configuration was set to 7 leaf nodes, 15 decision trees, and a feature selection threshold of 36. On the 3,600-sample dataset under identical 36-feature subspace constraints, RF achieved training accuracy of 96.27% and test accuracy of 92.85%, significantly outperforming BPNN (testing set accuracy = 88.61%).The performance advantage is attributed to the dual randomness mechanism (Bootstrap sampling and feature subspace selection), which effectively suppresses overfitting, and the high-discriminability feature focus, which enhances spectral difference capture capability. Compared to the BPNN model, the RF model demonstrated superior training efficiency (68.02%reduction in computation time) and test classification accuracy (4.24% improvement).

ELM Training Model Results

Following the performance analysis of the RF model, this section further explores the classification effectiveness of ELM to compare the strengths and weaknesses of different algorithms. Fig. 10 illustrates the effect of hidden neuron number in the neural network on accuracy. As the number of neurons increases, the training accuracy consistently improves, indicating that increasing model complexity enhances the ability to capture training data patterns. However, the test accuracy peaks within the 200-400 neuron range, with the highest recorded accuracy at 360 neurons ((testing set accuracy = 98.38%). Beyond this point, as the number of neurons continues to increase, the test accuracy gradually declines, reaching approximately 96.93%, suggesting that overfitting becomes more pronounced. The gap between training and test accuracy significantly expands beyond 310 neurons, reaching 95.47% at 600 neurons, highlighting that excessively complex network structures impair generalization ability. Therefore, limiting the number of hidden neurons to 360 achieves an optimal balance between performance and model complexity.

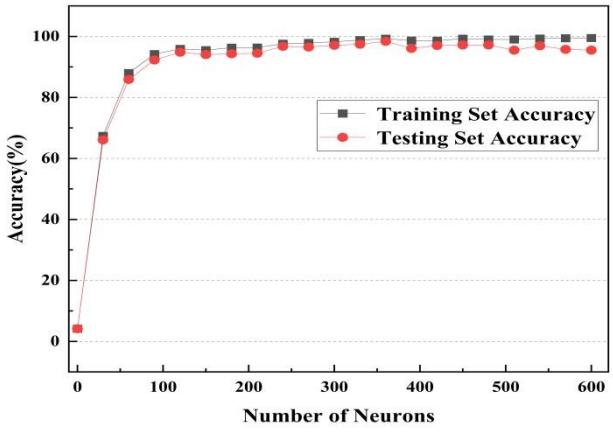


Fig. 10 Effect of hidden neuron number on model performance and accuracy trends in training and testing sets.

Model Evaluation

To ensure equitable comparison, as shown in table 3, both ELM and BPNN models were retrained on the identical 36-feature

View Article Online
DOI: 10.1039/D5AY00526D

subset selected by RF. The test accuracy of ELM decreased marginally to 94.72% (from 98.38%), while BPNN suffered a substantial drop to 88.61% (from 92.04%). Notwithstanding the feature reduction, ELM consistently outperformed both RF and BPNN, achieving inference in merely 0.97 seconds demonstrating its superior robustness under dimensionality constrained conditions.

Table 3 Performance comparison with 36 optimal features

Algorithms	Training set accuracy/%	Testing set accuracy/%	ΔAcc* (%)	Computation time (s)
BP	95.32%	88.61%	3.34%	94.47
RF	96.27%	92.85%	2.02%	30.21
ELM	97.77%	94.72%	3.66%	0.97

*ΔAcc: The performance gap between full-feature testing accuracy and optimal-36-feature testing accuracy.

Table 4 outlines the modeling accuracy and computational efficiency of different algorithms. Under a unified MSC preprocessing condition, the study evaluated the modeling performance of BPNN, RF, and ELM. Experimental results indicate that the ELM algorithm demonstrated significant overall advantages, achieving a training accuracy of 99.25% and a test accuracy of 98.38%, surpassing RF and BPNN by 2.64 and 1.67 percentage points in training accuracy, and by 4.63 and 6.34 percentage points in test accuracy, respectively. Additionally, ELM exhibited exceptional computational efficiency, requiring only 1.59 s for processing, which is faster than RF (45.90 s) and BPNN (453.36 s), respectively, demonstrating that ELM not only ensures high predictive accuracy but also possesses superior real-time processing capabilities. The RF algorithm, benefiting from its ensemble learning mechanism, achieved a test accuracy of 93.75%, outperforming BPNN by 1.17 percentage points, but its computational cost was 285 times higher than ELM, highlighting a significant efficiency bottleneck. The BPNN algorithm, constrained by the local convergence characteristics of gradient descent, achieved a test accuracy of only 92.04%, while incurring an extremely high computational cost of 453.36 s, revealing its practical limitations in real-world applications.

Table 4 Accuracy comparison of three algorithms

Algorithms	Pre-processingmethod	Training set accuracy/%	Testing set accuracy/%	Computation time (s)
BP	MSC	97.58%	92.04%	453.36
RF		96.61%	93.75%	45.90
ELM		99.25%	98.38%	1.59

Critically, the ELM model achieved superior performance without relying on feature dimensionality reduction. As evidenced in Table 4, ELM attained a test accuracy of 98.38% using all 121 spectral bands, surpassing RF (93.75%) by 4.63% and BPNN (92.04%) by 6.34%. This result underscores ELM's inherent capability to harness high-dimensional hyperspectral data efficiently. Unlike RF, which required meticulous feature selection (top 36 bands) to optimize accuracy and speed (Table 3), ELM delivered higher accuracy (98.38% vs. 92.85%) and faster inference (1.59 s vs. 30.21 s) even without such preprocessing. The elimination of feature engineering not only simplifies the analytical workflow but also enhances practical utility in forensic settings where rapid, high-confidence decisions are paramount.

To evaluate the model performance of ELM, BPNN, and RF, this study employed four evaluation metrics: PC, SN, SP, and F1-score for systematic analysis. In Tables 5and 6, he BPNN algorithm achieved an average F1-score of 0.976 on the training set (Table 5) but significantly dropped to 0.923 on the test set (Table 6), indicating weak generalization ability. Performance fluctuations were particularly evident in complex samples, such as test sample 11, where the F1-score plummeted to 0.615 (SN = 0.444, PC = 1.000), while samples 8, 9, and10 recorded F1 scores of 0.789, 0.636, and 0.818, respectively. These results suggest that the BPNN model is highly sensitive to noise and prone to local optima, making it challenging to handle high-dimensional data classification tasks reliably.

The RF model outperformed BPNN overall, achieving an average F1-score of 0.947 on the test set. However, its stability remained insufficient, as evidenced by a 4.1% performance gap (0.988 → 0.947). Particularly in class-imbalanced scenarios, RF struggled to recognize minority class samples effectively, such as test sample 21 (SN = 0.877, F1-score = 0.851) and test sample 22 (SN = 0.795, F1-score = 0.823). These results suggest that RF requires feature weighting or sampling strategies to enhance classification robustness.

ELM exhibited significant advantages in both the training and test sets, demonstrating superior performance across all metrics. The PC values reached 1.000 in 16 samples (e.g., samples 1-7 and 14-16), while the SP values consistently exceeded 0.992. The average F1-score of 0.983 on the test set represented a 6.00% and 3.60% improvement over BPNN and RF, respectively. Its single-hidden-layer feedforward structure, combined with randomly generated hidden layer parameters and a generalized inverse matrix solution, eliminated iterative parameter tuning, achieving a training efficiency over 200 times faster than BPNN, while maintaining high accuracy at a significantly lower computational cost.

The comparative analysis in this study highlights ELM as the most effective method for hyperspectral stamp pad ink classification, demonstrating both stability (training-testing F1-score difference of only 0.017) and computational efficiency, far exceeding traditional models. The generalization limitations of BPNN and the class imbalance sensitivity of RF underscore the shortcomings of conventional approaches, whereas ELM, with its unique network structure and analytical learning mechanism, offers a highly reliable solution for forensic document authentication. Although ELM demonstrates remarkable performance in controlled laboratory settings, its adaptability to deteriorated samples (e.g., faded/contaminated impressions) requires further validation. Nevertheless, the model's exceptional efficiency enables rapid recalibration for complex scenarios.

View Article Online
DOI: 10.1039/D5AY00526D

Table 5 Performance comparison of BPNN, RF, and ELM training models

brochure	BPNN (PC/SN/SP/ F1-score)	RF (PC/SN/SP/ F1-score)	ELM (PC/SN/SP/ F1-score)
1	1.000 /1.000/1.000/1.000	1.000/1.000/1.000/1.000	1.000/1.000/1.000/1.000
2	1.000/1.000/1.000/1.000	1.000/1.000/1.000/1.000	1.000/1.000/1.000/1.000
3	1.000 /1.000/1.000/1.000	1.000/1.000/1.000/1.000	1.000/1.000/1.000/1.000
4	1.000 /0.990/1.000/0.995	1.000/1.000/1.000/1.000	1.000/1.000/1.000/1.000
5	1.000/0.990/1.000/0.995	0.990/0.990/0.999/0.990	1.000/1.000/1.000/1.000
6	1.000/1.000/1.000/1.000	0.990/0.990/0.999/0.990	1.000/1.000/1.000/1.000
7	1.000/1.000/1.000/1.000	1.000/1.000/1.000/1.000	1.000/1.000/1.000/1.000
8	0.895/0.981/0.995/0.936	0.977/0.977/1.000/0.988	0.960/0.950/0.998/0.955
9	0.978/0.885/0.999/0.930	1.000/1.000/0.999/0.990	0.951/0.960/0.997/0.955
10	0.913/1.000/0.995/0.954	1.000/1.000/0.999/0.995	1.000/0.982/1.000/0.990
11	1.000/0.666/1.000/0.800	0.951/0.95610.998/.956	1.000/0.990/1.000/0.995
12	0.859/0.995/0.993/0.920	0.961/0961/0.997/0.956	0.990/0.990/0.999/0.990
13	0.990/1.000/0.999/0.995	0.976/0.979/1.000/0.987	0.980/1.000/0.999/0.990
14	0.918/0.971/0.996/0.944	1.000/1.000/0.999/0.990	0.990/0.990/0.999/0.990
15	1.000/1.000/1.000/1.000	1.000/1.000/1.000/1.000	1.000/1.000/1.000/1.000
16	0.990/1.000/0.999/0.995	1.000/1.000/1.000/1.000	1.000/1.000/1.000/1.000
17	0.972/1.000/0.998/0.985	0.959/0.959/0.999/0.974	0.990/1.000/0.999/0.995
18	1.000/0.961/1.000/0.980	0.980/0.980/0.998/0.976	1.000/0.990/1.000/0.995
19	0.963/1.000/0.998/0.981	0.989/0.989/0.999/0.984	0.990/1.000/0.999/0.995
20	1.000/1.000/1.000/1.000	1.000/1.000/1.000/1.000	1.000/0.990/1.000/0.995
21	0.970/0.923/0.998/0.946	0.950/0.950/0.999/0.964	0.990/0.990/0.999/0.990
22	0.927/0.971/0.996/0.948	0.981/0.981/0.997/0.967	0.980/0.990/0.999/0.985
23	1.000 /1.000/1.000/1.000	1.000/1.000/1.000/1.000	1.000/1.000/1.000/1.000
24	1.000/1.000/1.000/1.000	1.000/1.000/1.000/1.000	1.000/1.000/1.000/1.000
Average	0.975/0.953/0.996/0.976	0.988/0.988/0.995/0.988	0.992/0.992/0.999/0.992

Table 6 Performance comparison of BPNN, RF, and ELM testing models

brochure	BPNN (PC/SN/SP/ F1-score)	RF (PC/SN/SP/ F1-score)	ELM (PC/SN/SP/ F1-score)
1	1.000 /1.000/1.000/1.000	1.000 /1.000/1.000/1.000	1.000/0.979/1.000/0.989
2	1.000/1.000/1.000/1.000	1.000/1.000/1.000/1.000	1.000/1.000/1.000/1.000
3	1.000 /1.000/1.000/1.000	1.000 /1.000/1.000/1.000	1.000/1.000/1.000/1.000
4	1.000 /0.977/1.000/0.988	1.000 /1.000/1.000/1.000	1.000/1.000/1.000/1.000
5	0.978 /1.000/0.999/0.989	0.975/0.975/0.999/0.975	1.000/1.000/1.000/1.000
6	1.000 /1.000/1.000/1.000	0.977/0.977/0.999/0.977	1.000/1.000/1.000/1.000
7	1.000 /1.000/1.000/1.000	1.000/1.000/0.999/0.989	1.000/1.000/1.000/1.000
8	0.652/1.000/0.976/0.789	0.919/0.919/1.000/0.958	0.976/0.836/0.999/0.911
9	1.000/0.466/1.000/0.636	0.912/1.000/0.995/0.954	0.859/1.000/0.992/0.924
10	0.873/0.800/0.993/0.818	1.000/1.000/0.996/0.961	1.000/0.959/1.000/0.979
11	1.000/0.444/1.000/0.615	0.934/0.934/0.993/0.895	0.979/0.959/0.999/0.969
12	0.882/1.000/0.994/0.937	0.913/0.913/0.998/0.933	0.960/0.979/0.998/0.969
13	0.771/0.977/0.987/0.862	0.865/0.865/0.999/0.918	0.960/1.000/0.998/0.980
14	0.756/0.622/0.991/0.682	0.978/0.978/0.992/0.900	1.000/0.979/1.000/0.989
15	1.000/1.000/1.000/1.000	1.000/1.000/0.996/0.966	1.000/1.000/1.000/1.000
16	1.000/1.000/1.000/1.000	1.000/1.000/0.996/0.952	1.000/1.000/1.000/1.000
17	0.725/1.000/0.983/0.841	0.903/0.903/1.000/0.949	0.980/1.000/0.999/0.989
18	1.000/0.933/1.000/0.965	0.804/0.804/0.999/0.881	1.000/1.000/1.000/1.000
19	0.937/1.000/0.997/0.967	0.961/0.961/0.991/0.892	1.000/1.000/1.000/1.000
20	1.000/0.977/1.000/0.988	0.932/0.932/1.000/0.964	1.000/1.000/1.000/1.000
21	0.907/0.866/0.996/0.886	0.877/0.877/0.992/0.851	0.959/0.959/0.998/0.959
22	0.854/0.911/0.993/0.881	0.795/0.795/0.994/0.823	0.959/0.959/0.998/0.959
23	1.000 /1.000/1.000/1.000	1.000 /1.000/1.000/1.000	1.000 /1.000/1.000/1.000
24	1.000/1.000/1.000/1.000	1.000/1.000/1.000/1.000	1.000/1.000/1.000/1.000
Average	0.920/0.953/0.990/0.923	0.948/0.9483/0.997/0.947	0.984/0.983/0.999/0.983

Conclusion

To address the challenges of timeliness and accuracy in nondestructive stamp pad ink detection for forensic identification, this study proposed a rapid classification framework integrating HSI and ELM. Experimental results demonstrated that the proposed framework achieved a classification accuracy of 98.38% across 24 photosensitive ink samples, with a detection time of only 1.59 s. Critically, even without feature dimensionality reduction (i.e., using full 121 spectral bands), the ELM model outperformed both RF and BPNN by significant margins. ELM achieved a test accuracy of 98.38%, surpassing RF (93.75%) by 4.63% and BPNN (92.04%) by 6.34%. More importantly, ELM accomplished this with a computational time of merely 1.59 seconds—28 times faster than RF (45.90 s) and 285 times faster than BPNN (453.36 s). This dual advantage of superior accuracy and extreme efficiency underscores ELM's unique suitability for real-time forensic applications where rapid, high-confidence analysis is paramount.

However, the potential application for analyzing ink aging—while acknowledged—was not experimentally explored. Further assessment is needed to determine the reliability of the ELM model when analyzing deteriorated samples or operating under heterogeneous environmental conditions. Future work should also validate the model under realistic forensic scenarios involving diverse substrates and variable lighting conditions to strengthen ecological validity.

Data Availability Statement

Data are available from the corresponding author upon reasonable request.

Author contributions

Lu Xiaoquan wrote and revised the original manuscript; Zhang Xinyu, Wu Jiaquan and Ma Kun provided theoretical guidanceand overall planning; Zhang Jianqiang and LI Fan revised the original manuscript; Ren Huihui and Chen Hang proofread the manuscript.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

The authors acknowledge Kunming Public Security Bureau for technical support regarding experimental samples. This work was supported by the Physical Evidence Spectral Technology Innovation Team of Yunnan Police College (202105AE160007),and the Basic Project Research of Yunnan Police College (202101AU070011).

References

1 X. F. Guo, *Guangdong Gongan Keji*, 2017, **25**, 18-22.
2 W. Han, J. T. Huang and H. Wang, J. Peoples Public Security Univ. China (Sci Technol), 2016, **22**, 28-32.
3 Z. Y. Zhang, Q. F. Song, Y. Zhu, L. M. Xing and Y. J. Wang, *J. Criminal Invest Police Univ. China*, 1998, **02**, 37-38.
4 G. Ouyang, B. Li, P. Zhao, X. Guo and C. Wang, *J. Forensic Sci.*, 2019, **64**, 1203-1212.
5 S. Sharma, D. Garg, R. Chophi and R. Singh, *Forensic Chem.*, 2021, **26**, 100377.
6 A. Raza and B. Saha, *Sci. Justice*, 2013, **53**, 332-338.
7 J. X. Zou, Y. Y. Wang and G. J. Shi, *Chinese Journal of Forensic Sciences*, 2013, **05**, 28-31.
8 J. Zhang, J. X. Zou, G. J. Shi, A. D. Yang, L. X. Lin and Y. X. Dong, *Forensic Science and Technology*,2010, **02**, 37-39.
9 C. Neumann and P. Margot, *Forensic Sci. Int.*, 2009, **185**, 29-37.
10 Y. Zhu and B. Li, *Guangdong Gongan Keji*, 2018, **26**, 30-32.
11 B. Li, *J. Forensic Sci.*, 2014, **59**, 1403-1409.
12 C. Li, H. Li, Y. Zhou and X. Wang, *Sustainability*, 2024, **16**, 3582.
13 G. S. Kim, Y. Gwon, E. J. Oh, D. Kim, J. H. Kwon and Y. D. Kim, *Appl. Sci.*, 2023, **13**, 4631.
14 J. Qin, M. S. Kim, K. Chao, D. E. Chan, S. R. Delwiche and B.-K. Cho, *Appl. Sci.*, 2017, **7**, 125.
15 J. Ma, D. W. Sun, H. Pu, J. H. Cheng and Q. Wei, *Annu. Rev. Food Sci. Technol.*, 2019, **10**, 197-220.
16 A. Mohammed Ridha, N. A. M. Isa and A. Tawfik, *J. Imaging*, 2024, **10**, 174.
17 Z. Pei, Y. M. Huang and T. Zhou, *Photonics*, 2023, **10**, 1104.
18 W. Sun, R. L. Chen and J. X. Luo, *Laser & Optoelectronics Progress*, 2021, **58**, 0600007.
19 G. Reed, K. Savage, D. Edwards and N. Nic Daeid, *Sci. Justice*, 2014, **54**, 71-80.
20 B. Melit Devassy and S. George, *Forensic Sci. Int.*, 2020, **311**, 110194.
21 S. Wang, H. He, R. Lv, W. He, C. Li and N. Cai, *J. Forensic Sci.*, 2022, **67**, 550-561.
22 G. Qian and L. Zhang, *Appl. Soft Comput.*, 2018, **70**, 1034-1041.
23 L. Breiman, *Mach. learn.*, 2001, **45**, 5-32.
24 S. Asadi, S. Roshan and M. W. Kattan, *J. Biomed. Inf.*, 2021, **115**, 103690.
25 G. B. Huang, Q. Y. Zhu and C. K. Siew, *Neurocomputing*, 2006, **70**, 489-501.
26 J. Wang, S. Y. Lu, S. H. Wang and Y. D. Zhang, *Multimed. Tools Appl.*, 2022, **81**, 41611-41660.
27 M. Adegoke, A. C. S. Leung and J. Sum, in *Neural Information Processing: 25th International Conference,* Springer-Verlag, 2018, pp. 680-689.
28 S. Kuila, N. Dhanda and S. Joardar, *Multimed. Tools Appl.*, 2023, **82**, 29857–29881.
29 X. Y. Wang, C. H. Zhu, Z. T. Fu, L. X. Zhang and X. X. Li, *Spectrosc. Spect. Anal.*, 2019, **39**, 1864-1869.
30 Y. Zhang, B. Nie, J. Du, J. Chen, Y. Du, H. Jin, X. Zheng, X. Chen and Z. Miao, *PeerJ Comput. Sci.*, 2023, *9*, e1711.

Data are available from the corresponding author upon reasonable request.

View Article Online
DOI: 10.1039/D5AY00526D