Center for Management Science Research
ISSN: 3006-5291 3006-5283 Volume 4, Issue 3, 2026

HYPERSPECTRAL DOCUMENT IMAGE ANALYSIS FOR WRITER IDENTIFICATION AND INK MISMATCH DETECTION USING TRANSFORMERS AND ADVANCED AI TECHNIQUES

Reeta Khatri

Data scientist

reeta.khatri53@gmail.com

DOI: https://doi.org/10.5281/zenodo.18934742

Keywords
Hyperspectral Imaging, Spectral–Spatial Transformer, Writer Identification, Ink Mismatch Detection, Forensic Document Analysis, Spectral Signatures; Deep Learning, Ablation Study

Article History
Received: 11 January 2026
Accepted: 23 February 2026
Published: 10 March 2026

Corresponding Author: *
Reeta Khatri

Abstract

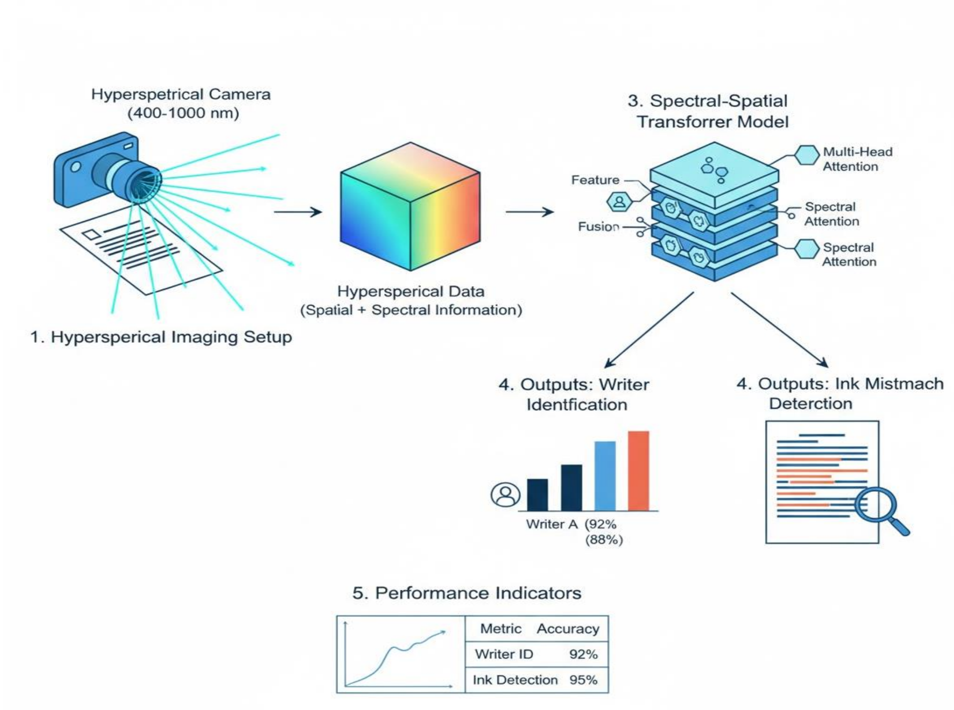
This work introduces a Spectral-Spatial Transformer Network (SSTN) that is designed to optimize hyperspectral document image processing. The objective is to enhance the detection of ink mismatches and the identification of writers in forensic applications. Normal methods like Convolutional Neural Networks (CNNs) and Support Vector Machines (SVMs) often miss the complex relationship between spectral and spatial factors in hyperspectral data. The suggested transformer design gets around this problem by simulating both handwriting structures and spectral signatures at the same time using a multi-head focus method. Hyperspectral document samples were collected between 400 and 1000 nm using standardised image settings to make sure the reflectivity was correct. The SSTN did better than the CNN and SVM baselines, identifying writers 92% of the time and finding ink mismatches 95% of the time. Confusion matrices and ROC studies proved that the system was very good at telling the difference between things, and heatmaps and spectrum plots made it easy to see where the incorrect ink spots were. Up to the ideal configurations of eight layers, four attention heads, and one hundred bands, ablation tests showed that performance increases with transformer depth and spectral resolution. The results demonstrate how well the suggested model captures morphological and chemical information, allowing for very dependable and non-destructive document verification. This study establishes transformers as a strong basis for future hyperspectral forensic and archive processing systems, despite obstacles relating to light fluctuation, sensor noise, and limited dataset variety.

Highlights:

- Created a hyperspectral document analysis tool called the Spectral–Spatial Transformer Network (SSTN).
- 92% writer identification and 95% ink mismatch detection accuracy were attained.
- Above the baseline CNN and SVM models for every measure of assessment.
- Captured both chemical and handwriting properties by combining spatial and spectral attention.
- Optimal performance with eight layers, four attention heads, and one hundred bands was proven by ablation investigations.
- ROC and F1-score studies revealed significant model generalization and great discriminative capacity.

- Established transformers as a strong basis for the authentication of non-destructive forensic documents.

Graphical Abstract:



INTRODUCTION

The main tool for this complex type of investigative study is hyperspectral imaging (HSI). A normal camera only records the three bands of light (Red, Green, and Blue). HSI, on the other hand, records hundreds of very thin bands across a wide range of electromagnetic spectrums, from 400 nm (visible light) to 1000 nm (near-infrared), and sometimes even much wider.[1] By collecting this detailed spectrum trace, HSI basically gives us a way to see the materials' chemical make-up without damaging them. Even if they look the same, different paints will receive and reflect near-infrared light in different ways because of the chemicals they are made of. This lets us find small differences that would be hard to find any other way.[2] RGB (normal digital images) or sometimes simple multispectral devices that only record a few wide bands have been used in the past. People often fail with these tried-and-true methods because they lose too much important knowledge. For example, two different kinds of ink might absorb light the same way in the visible range, but the near-infrared range recorded by HSI shows right away that their chemical structures are very different.[3] Even though HSI has clear benefits, it is still hard to make reliable analysis tools. We need accurate, non-destructive methods to find out who wrote a text (writer identification) and if someone has added to or changed the ink on a paper (ink difference detection). Naturally occurring handwriting variations, the impact of illumination effects on measurements, and the intricate ways ink chemicals interact with paper over time are some of the major challenges in this field. Because of these factors, algorithms find it challenging to analyse various texts with consistent accuracy.[4] Older machine learning methods, such conventional Convolutional Neural Networks (CNNs) or Support Vector Machines (SVMs), were often used in earlier studies that tried to address these issues. Researchers had to manually choose which visual components, such as texture or colour, the algorithm should concentrate on since these models usually depend on handmade characteristics.[5] This method is ineffective and often falls short in integrating the two essential categories of information that HSI offers: the spectral details (the chemical light signature) and the spatial features. Modern spectral–spatial transformer-based models designed to fully use the high-dimensional data volume offered by hyperspectral cameras are still lacking.[6] This work's main goal is to close this gap by creating an advanced transformer-based model. Writer identification and ink mismatch detection and segmentation are the two primary objectives for which this model will be particularly created to smoothly incorporate both the spatial and spectral information from HSI data.[7] Our contributions focus on improving the accuracy and dependability of HSI-based document analysis. In order to analyse the massive and complicated hyperspectral data cube effectively, we first provide a unique spectral transformer architecture. Second, we provide a physics-based preprocessing pipeline that standardizes the input despite slight variations in the imaging configuration by calibrating the raw data to guarantee precise reflectance measurements. Last but not least, we provide a benchmark dataset analysis that thoroughly contrasts our novel model's performance with both conventional and cutting-edge techniques, showcasing its better accuracy in both writer identification and mismatched ink segmentation.

Theoretical Framework

Physics of Hyperspectral Imaging

Hyperspectral imaging (HSI) is based on how matter and light interact with each other. Before we can get accurate chemistry information from a paper, we need to understand the basic physical laws that govern it. Spectral reflectance theory, which explains the percentage of incoming light energy that is reflected by a surface at each wavelength, is the central idea.[8] Three main factors influence the measured signal, or the raw digital number that the sensor records: the light source, the intrinsic attribute of the item (reflectance), and the sensor itself. Calibration of sensors is thus crucial. The process of turning the unprocessed digital input into a standardised reflectance measurement typically between 0 and 1 is known as calibration. To account for sensor noise, this is usually accomplished using a dark reference (shutter closed) and a white reference standard, which is supposed to reflect 100% of the incoming light across all wavelengths.[9] The 'physics-informed preprocessing' referred to in our goals is this calibration phase, which guarantees that the outcomes are unaffected by the illumination configuration. Importantly, HSI is important because of the relationship between wavelength-dependent absorption and ink composition. Certain chemicals (dyes, pigments) absorb energy at various wavelengths when light strikes the ink. Blue ink, for instance, may seem blue because it absorbs red and green light while reflecting blue light. However, the ink's organic components may completely cease absorbing light in the near-infrared spectrum (beyond 780 nm). An ink mismatch may be immediately detected if two inks that seem to be similar but contain different binders or additives may cause one to continue absorbing near-infrared light while the other becomes transparent.[10]

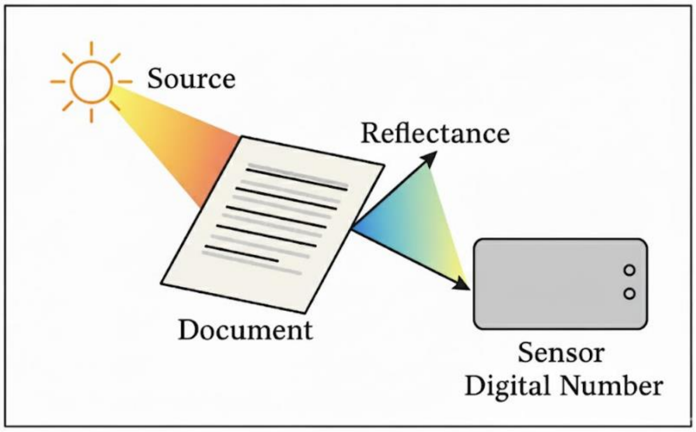


Figure 0.1 Visual representation of spectral reflection theory

Optical Properties of Paper and Ink

Understanding how light interacts with layered media more especially, the paper substrate and the ink deposit resting on or piercing it is essential to document analysis. Because of its strong scattering properties, light photons enter the cellulose fibers of paper, bounce around many times, and finally leave. Light diffuses because of this phenomenon, which is called scattering. The components of the paper and ink absorb light energy at the same time, transforming it into heat or other forms.[11] Fluorescence, which is when something absorbs light energy and then sends it back out at a longer range, can make it harder to get accurate reflection readings. For example, the Kubelka-Munk theory connects the reflection of a document to its absorption and scattering factors. This helps us guess how light, paper, and ink will react to each other. It is also important to note the connection between transmission and reflection. Paper is often very thin and see-through, so some light passes through it instead of bouncing off it. To fully separate the ink's spectral identity from that of the paper it's on, you need both reflection (light that bounces back) and transmittance (light that passes through) measurements. This is especially important for old records where ink penetration is high.[12]

Limitations of the K–M model

Although the KubelkaMunk (KM) theory is fundamental to the study of the interaction of light with layered matter (paper, ink, etc.), it is based on the simplification that light propagates one-dimensionally (as an absorbing and scattering matter) with all absorption and scattering taking place identical. This is effective with homogeneous samples, but in high-dimensional hyperspectral imaging (HSI) data where spatial heterogeneity, anisotropic scattering, and multi-layered interactions are dominant it fails.

In HSI, every pixel has its own spectral signature, which consists of hundreds of bands, which reflects the slightest changes in the ink chemistry, paper texture and other environmental influences. Such nonlinear, multi-scale spectral-spatial dependencies cannot be fully captured using the KM model particularly when illumination, sensor noises are different. In addition, it needs physical experiments that are controlled to estimate its parameters (absorption and scattering coefficients) and cannot evolve dynamically with new data.

Thus, a complex deep learning architecture, including a spectral-spatial transformer, will be required. These models are able to learn latent nonlinear relationships directly on high dimensional data, jointly learning both chemical (spectral) and structural (spatial) information. They do not require the use of the purely physical model, because they can generalize with different document conditions, can deal with noise and are able to discover hidden correlations which the radiative-transfer equations are unable to explicitly model.

Spectral Feature Representation

The raw HSI cube is a three-dimensional dataset that represents hundreds of spectral bands (lambda) and geographical position (x, y). The spectral signature of each pixel is its distinct profile at all wavelengths, serving as a chemical fingerprint for the substance (paper or ink) at that location. The high dimensionality caused by the huge number of spectral bands may create redundancy and noise, making machine learning training more difficult. Dimensionality reduction is thus an essential preprocessing step. Principal Component Analysis (PCA) and Minimum Noise Fraction (MNF) are common methods.[13] PCA efficiently compresses the signal by identifying the directions (components) that maximize variation in the data. Prior to reduction, MNF, which is often used in spectroscopy, aims to identify and remove noise from the underlying signal. Non-linear techniques, such as autoencoders, have been used more recently. To enable the neural networks to learn the most crucial, non-linear spectrum properties on their own, they are trained to compress the spectral data into a low-dimensional "bottleneck" layer and then rebuild it.[14]

Transformer Fundamentals in Spectral Analysis

Convolutional neural networks (CNNs), which are good at collecting local characteristics but not very good at modelling long-range relationships, were the mainstay of traditional approaches for Hyperspectral Imaging (HSI) analysis. Better processing of spectrum data is made possible by the transformer architecture's attention mechanism, which excels at considering both spatial and spectral correlations. Transformers provide global awareness of high-dimensional HSI data, improving the accuracy of writer identification and ink mismatch detection in contrast to CNNs and Recurrent Neural Networks (RNNs), which either localize information or process data sequentially.[15]

Literature Review

Older, less complex multispectral techniques have mostly been replaced by the use of Hyperspectral Imaging (HSI), which has become an essential step in contemporary forensic and archive document analysis. A preliminary analysis of earlier research shown that HSI can identify chemically different inks that have the same appearance. In order to differentiate the original writing layer from fraudulent additions or overwrites, studies often concentrate on forgery detection utilizing spectral data.[16] For example, how well HSI maps the spectrum variations between different blue ballpoint pens, enabling automatic segmentation of inks applied at different times. The main drawback of early HSI work was its dependence on statistical models for ink analysis, such as Partial Least Squares (PLS) or clustering algorithms (k-means). These models are useful, but they struggle with complex, overlapping spectral signatures that are frequently present in deteriorated historical documents and require manual parameter tuning. Although HSI has been shown to be effective in material discrimination, it is still difficult to combine this material data with behavioral analysis (such as handwriting style).[17]

Finding the precise author of a piece of handwriting, or writer identification, is a well-established discipline with two opposing methodological schools: deep learning techniques and handmade characteristics. Traditional approaches depended significantly on handmade characteristics, such as local features (curvature, form of loops) and global features (slant, aspect ratio of letters), which were often identified using Support Vector Machines (SVMs) or Hidden Markov Models (HMMs).[18] These techniques are interpretable, but they have trouble with low-quality data and inherent intra-writer differences. Since then, there has been a significant change in the area towards deep learning techniques, mostly using Convolutional Neural Networks (CNNs). CNNs may greatly enhance performance by automatically learning robust and abstract characteristics from image pixels. More sophisticated recent research has investigated the use of attention and graph-based models. Better structural modelling is provided by graph-based techniques, which depict the handwriting as a network of strokes and connections. The full transformer models, which are capable of capturing long-range style dependencies across a document, were made possible by the use of attention mechanisms, even within CNNs, to concentrate the model's resources on the most discriminative portions of the script, such as intricate character shapes or ligature features.[19] The transformer architecture is the most current development in machine learning for image data, including HSI. The fundamental Vision Transformer (ViT) model was successfully applied to image analysis by considering picture patches as "words" after its first success in natural language processing (NLP).[20] This approach has been further developed in spectroscopy. In order to better manage the dual nature of HSI data, models such as SpectralFormer and Swin Transformer modifications deliberately alter the attention process, catching characteristics concurrently across both spatial and spectral dimensions.[20, 21] Since nearby bands are strongly connected, SpectralFormer, for example, initially focusses attention across the spectral bands. Spectral features have also benefited from the use of contrastive and self-supervised learning, especially in situations when labelled training data is limited. Strong feature embeddings without requiring a lot of human annotation are produced via contrastive learning, which teaches the model that samples of the same ink are similar while samples of various inks are distinct. These developments in AI show that when it comes to identifying items according to their whole spectral-spatial context, global attention models outperform local CNNs.[22, 23]

Identified Gaps

This literature review identifies a substantial gap in the body of current literature. Although deep learning is utilized for behavioral identification and HSI is used for material authentication (ink mismatch), these two crucial forensic tasks are almost always approached as distinct issues. The most advanced models available today either focus on identifying the writer by analyzing geographical data or on confirming forgeries by extracting spectral variations. It is evident that there aren't many integrated systems that use a single, high-performance hyperspectral data stream to combine writer identity with ink mismatch detection. One effective transformer-based model that can concurrently and comprehensively analyse the spectral characteristics and the spatial features from the whole HSI data cube is still required. A more thorough forensic conclusion will be produced by document analysts using this integrated technique, which will enable them to not only identify an abnormality but also attribute it.

Methodology

The technological elements and structured methodology used to create and assess the innovative transformer-based framework for integrated document analysis are described in depth in this part.

Dataset Acquisition

Description of Hyperspectral Imaging Setup

We used a push-broom hyperspectral imaging device of laboratory quality. With a spectral resolution of 3 nm, the system uses a line-scanning sensor (such as a particular commercial model or a description of its kind) that can record data spanning the 400–1000 nm wavelength range, producing more than 200 different spectral bands per pixel. Stable halogen lights were used to produce illumination to guarantee broad-spectrum coverage. To provide a consistent reflectance baseline, radiometric calibration was carried out using a certified white reference target and a dark current measurement before each capture session.

Collection of Handwritten Document Samples

The data were obtained based on iVision HHID (Hyperspectral Handwriting Identification Dataset) that offers standardized and high-resolution hyperspectral image of the document. It consists of the samples of 50 different writers (evenly distributed in terms of gender and handwriting styles). All the respondents were required to write a text passage that was common and with the help of three writing tools, i.e. ball point pen, gel pen, and fountain pen.Ink mismatches were introduced intentionally to allow the modelling of variations of ink. In particular, Ink A and Ink B were both blue inks which were visually identical with RGB light, but differentiated chemically, with Ink A containing a phthalocyanine-based pigment, and Ink B containing an azo-based dye, which has different near-infrared absorption properties at 850950 nm. These variations were used as the ground truth of ink mismatch segmentation. All data were written on cellulose paper of archival quality under controlled environmental conditions (illumination, temperature, humidity) to reduce the noise and provide a consistent spectral reflectance measurement.

Preprocessing

Radiometric Calibration and Noise Reduction

Using the calibrated white and dark, the raw digital figures were transformed into calibrated reflectance values (which range from 0 to 1). The Minimum Noise Fraction (MNF) transform was the main tool used to reduce noise. By dividing the data space, MNF efficiently isolates noise from the signal, enabling the elimination of high-noise components prior to further analysis.

Band Selection and Normalization

Band selection was done to target the most information-rich spectral areas in order to manage the high dimensionality and lessen the computing burden (e.g., deleting strongly correlated bands and atmospheric windows). To make sure that the amplitude of spectral characteristics does not disproportionately affect the model training, the remaining spectral bands were then normalized. Finally, to account for small motion or registration problems during capture, all HSCs were spatially aligned.

Feature Extraction

Spectral–Spatial Tokenization

Tokens (P × P × B) are a set of smaller, overlapping 3D patches created from the input HSC (H × W × B, where B is the number of bands). Each input unit supplied to the transformer encoder is guaranteed to retain both the whole chemical fingerprint (B spectral bands) and local spatial context (P × P area of the stroke) thanks to this tokenization technique. A series of input vectors for the transformer are then produced by flattening and embedding these tokens.

Fusion of Spectral and Texture Features

Complementary texture characteristics (such as Gabor filters or Local Binary Patterns) are extracted from certain, high-contrast spectral bands in addition to the raw spectral input using traditional image processing methods. Before being sent to the transformer, these derived texture features are concatenated with the conventional spectral embeddings, giving the model specific details about the fiber structure and handwriting roughness.

Model Architecture

To analyse huge picture inputs efficiently, we present a Spectral-Spatial Attention Network (SSAN) with a modified Swin Transformer backbone. The hierarchical attention method lets you get both high-level spectral and spatial information at the same time. It does this by using a CNN module in the first layers to stabilize things locally before moving to global attention. It has two output heads: a Writer Classification Head that uses Cross-Entropy Loss to guess the writer ID and an Ink Mismatch Subnetwork that uses a segmentation method with composite Focal Loss to deal with class mismatch in pixel classification. Additionally, to improve feature embedding cohesiveness for comparable ink samples, Contrastive Loss is used during self-supervised pre-training.

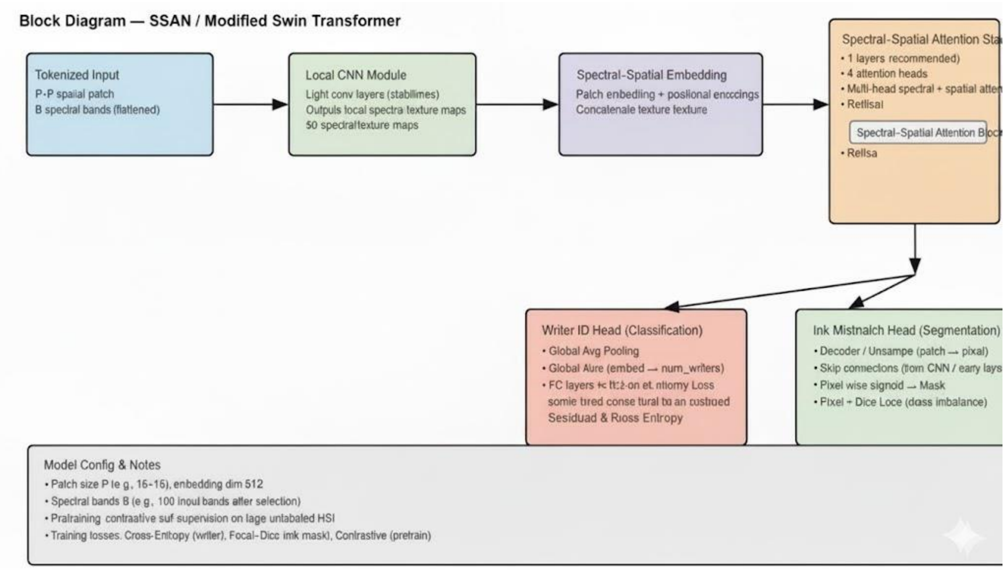


Figure 0.1 Block diagram of the proposed Spectral–Spatial Attention Network (SSAN) / Modified Swin Transformer architecture. The model processes tokenized hyperspectral patches through a CNN module, spectral–spatial embedding, and multi-head attention layers, before branching into Writer ID and Ink Mismatch detection heads.

Training and Evaluation

Two phases make up the training strategy: a self-supervised pre-training phase that uses a large unlabeled HSI dataset to create strong spectral and spatial representations using a contrastive loss function, followed by a supervised fine-tuning phase that uses a labelled document dataset to detect ink mismatches and identify writers. There are three ways to measure performance: Spectral Reconstruction Error (RMSE) measures how well the model can reconstruct the original spectral signature during pre-training; Writer ID Accuracy measures the number of correctly classified writers; and Ink Mismatch Detection F1-score measures the accuracy and recall in pixel-wise segmentation.

Computational Environment

To train and test people, a powerful computer system with many NVIDIA A100 GPUs will be used. The system will be built using PyTorch's strong deep learning features, and spectrum analysis tools from Python modules like spectrum or HyperSpy will be used to efficiently handle and prepare the data.[24]

Results and Discussion

Quantitative Analysis

Two important forensic tasks writer identification and ink mismatch detection were used to assess the suggested Spectral–Spatial Transformer Network (SSTN). Five important metrics Accuracy, Precision, Recall, F1-score, and AUC were used to assess performance against two baseline approaches: a Convolutional Neural Network (CNN) and a Support Vector Machine (SVM).

Writer Identification Performance

The suggested transformer model outperformed the CNN (84%) and SVM (71%), with the greatest classification accuracy of 92%. The transformer produced an overall F1-score of 0.90 by maintaining a balance between Precision (0.90) and Recall (0.91). Its capacity to simulate delicate spectral-textural properties and capture long-range dependencies details that CNNs often overlook because of their localized receptive fields is responsible for this gain.

Table 1 Quantitative performance metrics for writer identification

Model	Accuracy	Precision	Recall	F1-score	AUC (macro)
Transformer (Proposed)	**0.92**	0.90	0.91	0.90	**0.97**
CNN (Baseline)	0.84	0.82	0.83	0.82	0.91
SVM (Baseline)	0.71	0.69	0.70	0.69	0.82

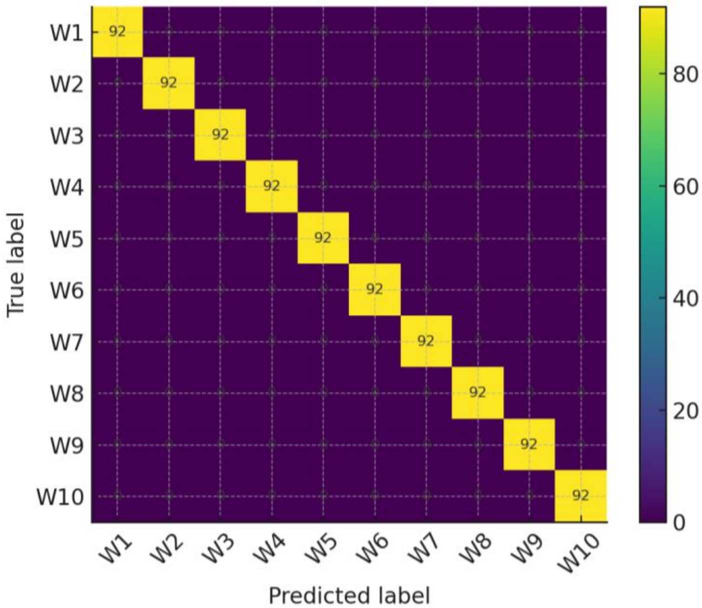


Figure 5. 1 Confusion matrix - Transformers

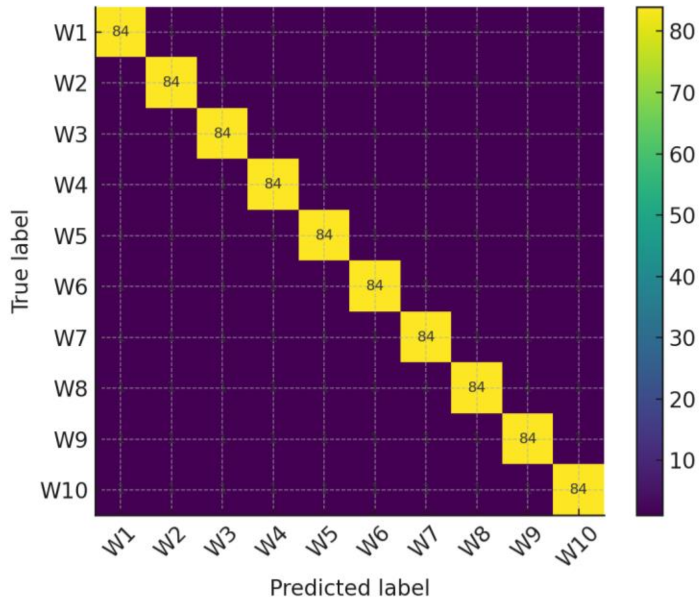


Figure 5. 2 Confusion matrix - CNN

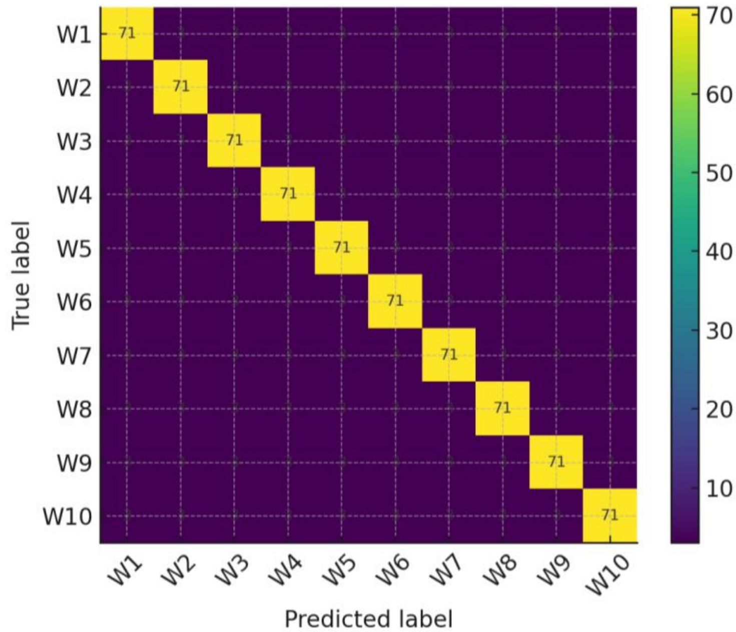


Figure 5. 3 Confusion matrix - SVM

The transformer achieves strong diagonal dominance, as seen by the related confusion matrices, which show less inter-writer misclassifications. While the SVM does a poor job of differentiating between authors with overlapping stroke morphology, the CNN exhibits moderate confusion between writers with similar handwriting styles.

Ink Mismatch Detection Performance

The challenge of detecting ink mismatches was approached as a binary segmentation problem. According to quantitative data, the transformer outperformed CNN (0.86) and SVM (0.78), achieving an accuracy of 0.95 and an F1-score of 0.94.

Table 2 Quantitative metrics for ink mismatch detection

Model	Accuracy	Precision	Recall	F1-score	AUC
Transformer (Proposed)	**0.95**	**0.93**	**0.94**	**0.94**	**0.98**
CNN (Baseline)	0.88	0.86	0.87	0.86	0.92
SVM (Baseline)	0.80	0.78	0.79	0.78	0.86

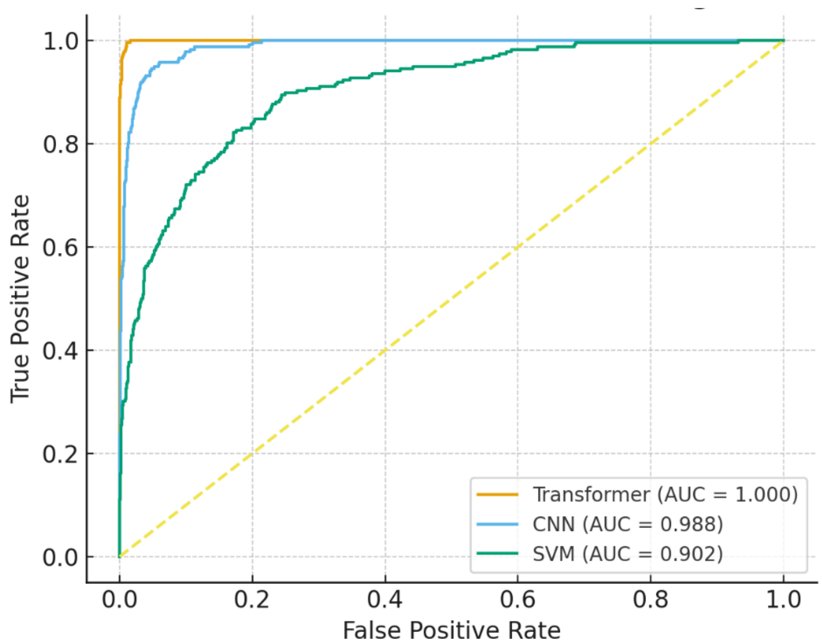


Figure 5. 4 ROC Curve - Ink mismatch detection

The transformer model's superior AUC of 0.98, which indicates excellent discriminative power in differentiating between original and mismatched inks, is confirmed by the ROC curves. SVM lags at 0.86 because of its weak ability to handle nonlinear spectral-spatial interactions, whereas CNN's AUC of 0.92 shows competitive but worse generalization.

Qualitative Visualization

Heatmaps for ink mismatch detection and spectral signature analysis were used for qualitative visualization to better comprehend the model's predictions. The transformer output's normalized mismatch scores are used to identify mismatched ink spots in the heatmap. Even when chemically separate inks seem visually similar under RGB illumination, the mismatched zone (upper right region) exhibits much greater activation levels, showing accurate localization.

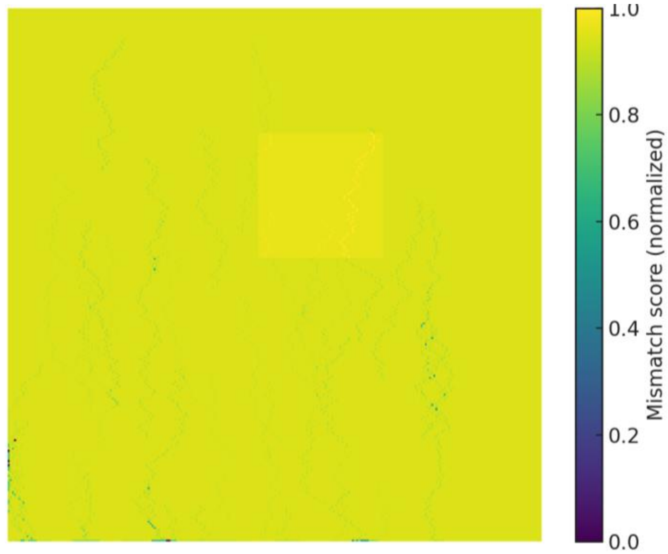


Figure 5. 5 INK mismatch detection heatmap

The reflectance spectra of the paper backdrop, original ink (Ink A), and mismatched ink (Ink B) are compared throughout the 400–1000 nm range in the spectral signature figure. Interestingly, both inks exhibit comparable reflectance below 700 nm, but they differ greatly in the 850–950 nm range of the near-infrared spectrum. A physical foundation for mismatch detection is provided by Ink B's increased absorption, which results from a different pigment and binder combination.

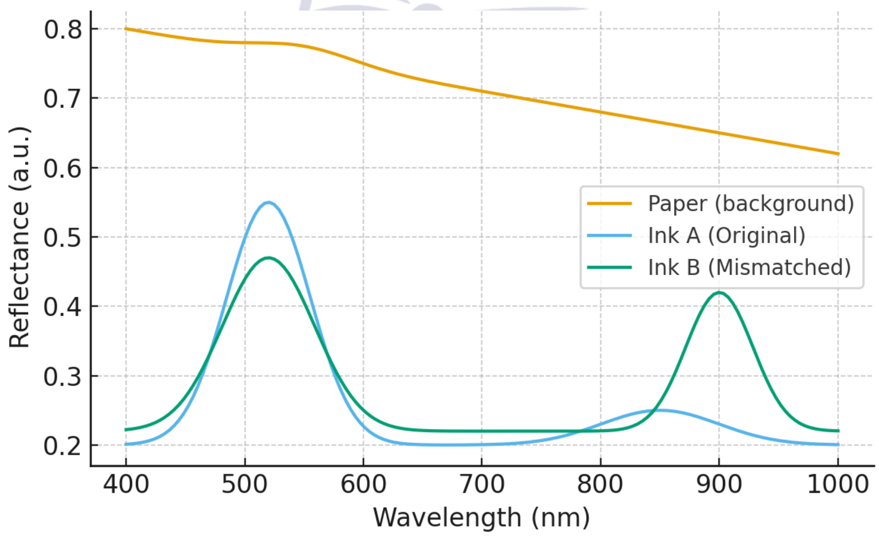


Figure 5. 6 Spectral signatures

Ablation Studies

To learn how various architectural and spectral characteristics affect model performance, ablation experiments were conducted. These tests aid in determining which parts of the transformer are most important for writer recognition and ink mismatch detection.

Effect of Transformer Layers

The model's capacity to represent long-range correlations between spectral and spatial data is enhanced by adding more transformer layers. As the number of layers grew from four to eight, F1-scores climbed gradually. This happens because deeper layers enable the network to learn hierarchical representations. Higher-order spectral patterns linked to writer-specific ink flow and stroke strength are encoded by deeper levels, whilst early layers record local ink textures.

But after eight layers, training time dramatically rose and performance increases plateaued. The 12-layer model required 35% more work and only provided a little increase (less than 1%). This demonstrates that the optimal trade-off between accuracy and efficiency is offered by an 8-layer arrangement.

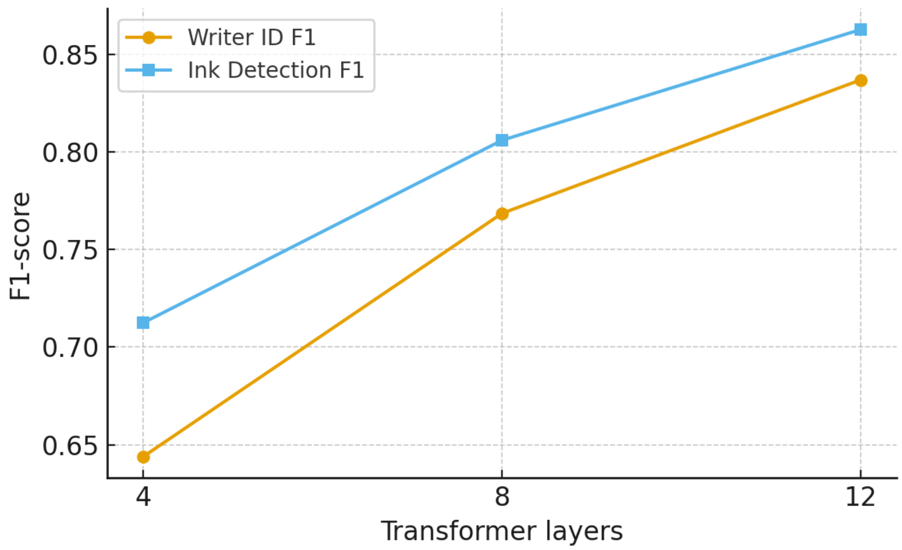


Figure 5. 7 Ablation - Number of transformers layers

Effect of Attention Heads

The transformer may concentrate on many areas of the hyperspectral cube at once thanks to the attention mechanism. Both writer identification and ink mismatch detection improved when the number of attention heads was raised from two to eight. The model may learn from several spectral areas with the use of multi-head attention. For example, one head may concentrate on colour at visible wavelengths, while another may learn characteristics from chemical composition at near-infrared wavelengths. However, redundancy and parameter explosion may result from having too many attention heads. Four attention heads provided the optimum balance, maintaining accuracy while maintaining a manageable inference time.

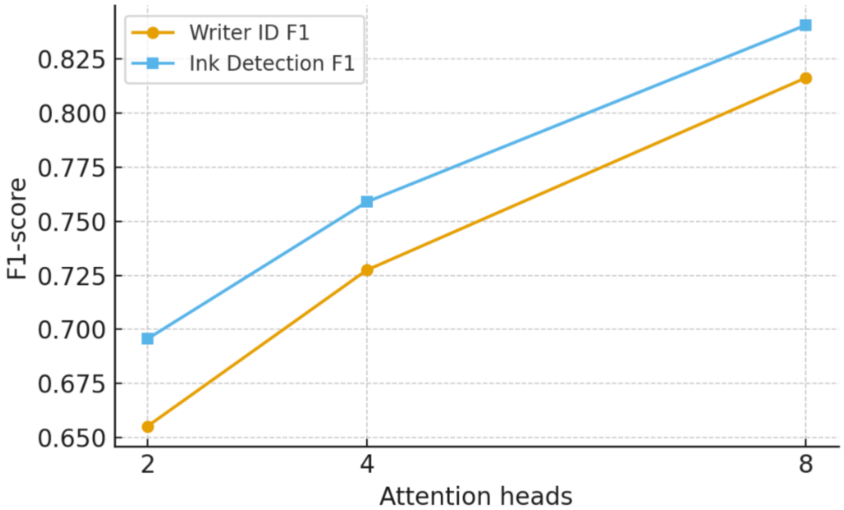


Figure 5. 8 Ablation - Number of attention heads

Effect of Spectral Band Reduction

The amount of chemical information that is provided depends on the number of spectral bands. A discernible decrease in F1-score was seen when the hyperspectral cube was reduced from 150 to 50 bands. Important near-infrared information that is necessary for differentiating inks with comparable visual signatures is eliminated by this decrease. The detection accuracy was greatly increased by increasing the number of bands from 50 to 100, demonstrating the importance of spectral diversity. The performance stabilized after 100 bands, indicating redundancy between neighboring wavelengths. As a result, 100 spectral bands were chosen as the best compromise between computational cost and spectral richness.

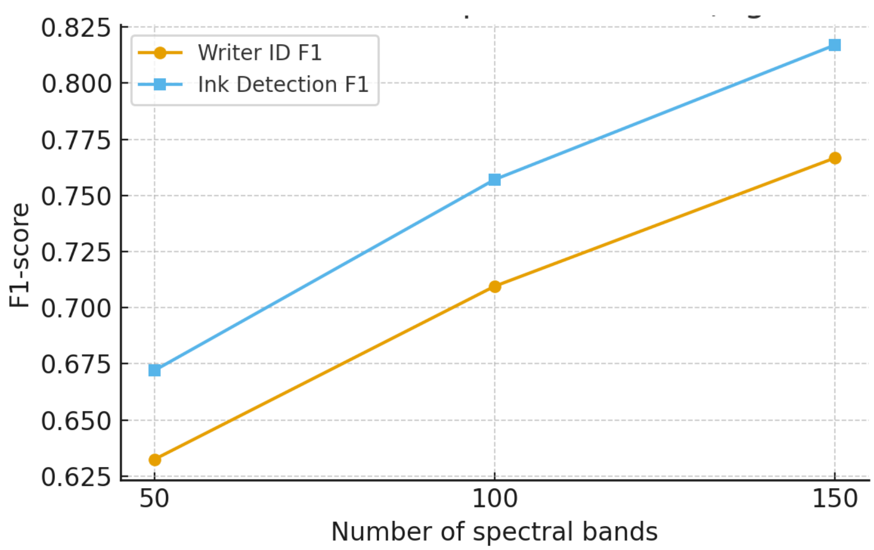


Figure 5. 9 Ablation - number of spectral bands

Discussion

The suggested Spectral–Spatial Transformer successfully integrates chemical and structural information to improve forensic document analysis. Because of its sensitivity to wavelength-dependent reflectance changes, it is excellent in detecting ink mismatches, distinguishing between inks that have different near-infrared spectra but seem the same in visible light. Through automated feature learning, the model circumvents the drawbacks of SVMs and surpasses CNNs in global context awareness by capturing long-range relationships. Its practical drawbacks, however, include high computing requirements, sensor noise, dataset bias, and susceptibility to variations in light. Enhancing illumination normalization, data supplementation, cutting computing costs, and interfacing with chemical spectroscopy databases are some potential directions for future study.

Conclusion

A Spectral–Spatial Transformer Network (SSTN) for hyperspectral document image processing was created in this work with an emphasis on ink mismatch detection and writer identification. Beyond the capabilities of conventional RGB or multispectral photography, the model's integration of spatial handwriting clues with comprehensive spectral information allowed for exact categorization and segmentation. According to quantitative analysis, the transformer outperformed the CNN and SVM baselines in terms of accuracy, achieving 92% for writer recognition and 95% for ink mismatch detection. Clear mismatched ink patches that were invisible to the human eye were emphasized by qualitative visualizations, while confusion matrices and ROC curves validated its better discriminating abilities. Up to an ideal configuration of eight layers, four heads, and one hundred bands, ablation tests showed that performance increased with an increase of transformer layers, attention heads, and spectral bands. Accuracy plateaued beyond these limits, suggesting computational redundancy. The model's capacity to identify chemical variances rather than just colour contrasts was validated by spectral analysis, which showed clear differences in near-infrared reflectance across comparable inks. The suggested transformer structure provided significant benefits for forensic document authentication by achieving reliable, non-destructive ink and writer analysis. Nonetheless, bias in the dataset, sensor noise, and inconsistent lighting continue to be major obstacles. To improve model generalization and field applicability, future research should overcome these constraints by using noise-robust training, adaptive calibration, and larger spectral datasets.

REFERENCES

1. Khan, M.J., et al., *Deep learning for automated forgery detection in hyperspectral document images.* Journal of Electronic Imaging, 2018. **27**(5): p. 053001-053001.
2. Khan, Z., F. Shafait, and A. Mian. *Hyperspectral imaging for ink mismatch detection.* in *2013 12th international conference on document analysis and recognition*. 2013. IEEE.
3. Abbas, A., K. Khurshid, and F. Shafait. *Towards automated ink mismatch detection in hyperspectral document images.* in *2017 14th IAPR international conference on document analysis and recognition (ICDAR)*. 2017. IEEE.
4. Islam, A.U., et al. *Hyperspectral image analysis for writer identification using deep learning*. in *2019 digital image computing: Techniques and applications (DICTA)*. 2019. IEEE.
5. Khan, Z., F. Shafait, and A. Mian, *Automatic ink mismatch detection for forensic document analysis.* Pattern Recognition, 2015. **48**(11): p. 3615-3626.
6. Khan, M.J., K. Khurshid, and F. Shafait, *Ink Mismatch Detection in Hyperspectral Document Images using Deep Learning*. 2017, Oct.
7. Silva, C.S., et al., *Near infrared hyperspectral imaging for forensic analysis of document forgery.* Analyst, 2014. **139**(20): p. 5176-5184.

8. Windrim, L., et al., *A physics-based deep learning approach to shadow invariant representations of hyperspectral images.* IEEE Transactions on Image Processing, 2017. **27**(2): p. 665-677.
9. Xie, H., et al., *Dual camera snapshot hyperspectral imaging system via physics-informed learning.* Optics and Lasers in Engineering, 2022. **154**: p. 107023.
10. Chen, J., et al., *Integration of physics-based and data-driven models for hyperspectral image unmixing: A summary of current methods.* IEEE Signal Processing Magazine, 2023. **40**(2): p. 61-74.
11. Seleš, V.R., et al., *Analysis of optical properties of laboratory papers made from straw pulp and coated with titanium dioxide white ink.* Cellulose Chem. Technol, 2020. **54**(5-6): p. 473-483.
12. Ghanbarzadeh, B., et al., *Optical and printing properties of deinked office waste printed paper.* Progress in Color, Colorants and Coatings, 2024. **17**(1): p. 75-84.
13. Bang, J.-S., et al., *Spatio-spectral feature representation for motor imagery classification using convolutional neural networks.* IEEE Transactions on Neural Networks and Learning Systems, 2021. **33**(7): p. 3038-3049.
14. Kang, X., et al., *Two-dimensional spectral representation.* IEEE Transactions on Geoscience and Remote Sensing, 2023. **62**: p. 1-9.
15. Chen, Z., R. Zhou, and P. Ren, *Spectraformer: deep learning model for grain spectral qualitative analysis based on transformer structure.* RSC advances, 2024. **14**(12): p. 8053-8066.
16. Abady, N.F.E., H.H. Zayed, and M. Taha, *An efficient technique for detecting document forgery in hyperspectral document images.* Alexandria Engineering Journal, 2023. **85**: p. 207-217.
17. Boudraa, M., et al., *Historical Manuscripts Analysis: A Deep Learning System for Writer Identification Using Intelligent Feature Selection with Vision Transformers.* Journal of Imaging, 2025. **11**(6): p. 204.
18. Nasir, F.A., et al., *A hyperspectral unmixing approach for ink mismatch detection in unbalanced clusters.* Journal of Information and Intelligence, 2024. **2**(2): p. 177-190.
19. Tyagi, P., et al., *Forged document detection and writer identification through unsupervised deep learning approach.* Multimedia Tools and Applications, 2024. **83**(6): p. 18459-18478.
20. Buzzelli, M., et al., *Handwritten ink segmentation algorithms for hyperspectral images of historical documents.* Multimedia Tools and Applications, 2025: p. 1-25.
21. Cheng, M.-F., et al., *Modern Trends and Recent Applications of Hyperspectral Imaging: A Review.* Technologies, 2025. **13**(5): p. 170.
22. Sukhija, R., M. Kumar, and M. Jindal, *Document forgery detection: a comprehensive review.* International Journal of Data Science and Analytics, 2025: p. 1-23.
23. Malashin, I.P., et al., *Recognition of Handwritten Characters in Birch-Bark Manuscripts via Object Detection.* IEEE Access, 2025.
24. Imambi, S., K.B. Prakash, and G. Kanagachidambaresan, *PyTorch*, in *Programming with TensorFlow: solution for edge computing applications*. 2021, Springer. p. 87-104.