applied sciences

MDPI

Article

Few-Shot Hyperspectral Representation Learning for Multi-Task Forensic Document Analysis Using Episodic Prototypical Networks

Rida Nasir [1,2], Khurram Khurshid [3] and Mansoor Hayat [4,*]

[1] Department of Electrical Engineering, Institute of Space Technology (IST), Islamabad 44000, Pakistan; ridanasir15@hotmail.com
[2] iVision Laboratory, Institute of Space Technology (IST), Islamabad 44000, Pakistan
[3] Department of Computing, Institute of Space Technology (IST), Islamabad 44000, Pakistan; khurram.khurshid@ist.edu.pk
[4] Section of Neurosurgery, Department of Surgery, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, MB R3T 1R9, Canada
* Correspondence: mansoor.hayat@umanitoba.ca

Abstract

Hyperspectral imaging has emerged as a promising modality for forensic document analysis due to its ability to capture spectral and material information beyond conventional RGB representations. However, existing hyperspectral forensic studies typically focus on individual tasks and provide limited insight into the transferability of learned representations across different forensic applications. This study investigates episodic hyperspectral representation learning for forensic document analysis using a shared spectral–spatial embedding space learned through prototypical optimization. The proposed framework is evaluated on the complete UWA WIHSI and iVision HHID hyperspectral datasets across multiple forensic applications, namely writer identification, writer verification, ink mismatch detection, forgery detection, gender prediction, and age prediction.The proposed method achieved high performance on the controlled UWA WIHSI dataset across multiple forensic tasks. On the more challenging iVision HHID dataset, it attained a writer identification accuracy of 90.74%, together with writer verification and ink mismatch detection ROC–AUC scores of 0.9936 and 0.9915, respectively. Cross-dataset experiments further highlighted both the potential and the challenges of transferring learned hyperspectral representations across different acquisition conditions and writer populations, while extensive ablation studies confirmed the importance of spectral information, embedding design, and optimization strategy in learning effective hyperspectral representations. Overall, the results show that shared hyperspectral representations can support diverse forensic document analysis applications within a common embedding space while providing insight into the opportunities and challenges of developing transferable hyperspectral representations for forensic settings.

Keywords: hyperspectral imaging; forensic document analysis; episodic prototypical learning; few-shot learning; writer identification; writer verification; ink mismatch detection; forgery detection; gender prediction; age prediction; spectral–spatial representation; metric learning; deep learning

Check for updates

Academic Editor: Andrea Prati

Received: 2 June 2026
Revised: 27 June 2026
Accepted: 2 July 2026
Published: 9 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Forensic document analysis plays an important role in criminal investigations, legal proceedings, and document authentication. Common tasks include writer identification, writer verification, forgery detection, and ink analysis. Traditional techniques such as microscopic inspection and chemical analysis can provide reliable evidence but are often time-consuming, require expert interpretation, and may risk damaging the original document [1]. These limitations have motivated the development of automated and non-destructive computational approaches.

Most automated document analysis systems rely on conventional RGB images. Handcrafted feature extraction and deep learning methods have demonstrated promising performance for writer identification and related tasks [2,3]. However, RGB representations are limited to visible-spectrum information and cannot capture the material characteristics of handwriting and ink. Consequently, documents produced using different inks or by different writers may appear visually similar, making reliable discrimination difficult.

Hyperspectral imaging (HSI) addresses this limitation by capturing spatial and spectral information across multiple wavelengths, providing a spectral signature for every pixel [4,5]. The additional spectral information enables the analysis of subtle differences in ink composition and handwriting characteristics that are not observable in RGB images. Consequently, HSI has been successfully applied to writer identification, ink mismatch detection, and forgery analysis using both classical machine learning and deep learning approaches [6–8].

Despite these advances, several challenges remain. Most hyperspectral forensic studies focus on individual tasks, such as writer identification or ink analysis, and are often evaluated on limited dataset subsets, restricting their ability to capture realistic variability [9]. Furthermore, the transferability of learned hyperspectral representations across different forensic tasks and datasets remains insufficiently explored. Consequently, it is unclear whether a shared hyperspectral representation can effectively support multiple forensic applications within a unified framework.

To address this gap, this work investigates episodic hyperspectral representation learning for forensic document analysis under limited-data conditions. The objective is to evaluate whether a shared spectral–spatial embedding learned through episodic prototypical optimization can support multiple forensic tasks. Unlike conventional multi-task learning frameworks that jointly optimize task-specific prediction heads using multiple objective functions, the proposed approach learns a single shared embedding that is subsequently evaluated across multiple downstream forensic document analysis tasks. By learning transferable similarity relationships between samples, episodic learning is particularly well suited to forensic scenarios where annotated data are limited [10].

The proposed framework is evaluated across six forensic applications, namely writer identification, writer verification, ink mismatch detection, forgery detection, gender prediction, and age prediction. Experiments are conducted using the complete UWA-WIHSI and iVision HHID datasets under both within-dataset and cross-dataset evaluation settings to assess the transferability of the learned representations [11]. The proposed method adopts a shared representation learning approach rather than a conventional multi-task learning architecture. Its primary contribution is to demonstrate that a single episodically learned hyperspectral representation can generalize across multiple forensic document analysis tasks without introducing a new backbone architecture or metric-learning algorithm.In this work, the term “multi-task” refers to the capability of a single learned hyperspectral representation to support multiple downstream forensic document analysis tasks, rather than conventional multi-task learning based on joint task-specific optimization.

The main contributions of this work are summarized as follows:

- We investigate episodic hyperspectral representation learning for forensic document analysis and evaluate its effectiveness under few-shot learning conditions.
- We propose a shared hyperspectral representation learning framework that learns a common spectral–spatial embedding capable of supporting multiple downstream forensic document analysis tasks without requiring task-specific prediction heads or joint multi-task optimization.
- We present a comprehensive evaluation across six forensic applications, including writer identification, writer verification, ink mismatch detection, forgery detection, gender prediction, and age prediction.
- We evaluate the framework using the complete UWA-WIHSI and iVision HHID datasets and investigate cross-dataset transferability under different acquisition conditions.
- We provide detailed experimental analysis highlighting both the strengths and limitations of episodic hyperspectral representations for forensic document analysis.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 presents the proposed methodology. Section 4 describes the experimental setup and evaluation protocol. Section 5 presents the experimental results and discussion.Section 6 presents the ablation studies. Section 7 presents the conclusions of the study. Finally, Section 8 discusses the limitations of the present study and outlines future research directions.

2. Related Work

Forensic document analysis encompasses a wide range of tasks, including writer identification, ink mismatch detection, forgery analysis, and document authentication. Early approaches in this domain relied primarily on handcrafted feature extraction from grayscale or RGB images. These methods focused on capturing structural and statistical characteristics of handwriting, such as stroke orientation, curvature, and texture, followed by classification using traditional machine learning techniques [12,13]. While effective in controlled environments, these approaches are inherently limited by their dependence on manually designed features and their inability to capture deeper representations of writing characteristics.

The introduction of deep learning has significantly advanced the field of document analysis. Convolutional neural networks (CNNs) have been widely adopted for writer identification and related tasks, enabling automatic feature learning directly from raw image data. Techniques such as patch-based learning, transfer learning using pre-trained networks, and feature aggregation strategies have demonstrated strong performance across multiple datasets [14,15]. Despite these improvements, RGB-based methods remain constrained by the limited information content of visible-spectrum images. In many forensic scenarios, documents written with different inks or by different writers may exhibit similar visual appearances, reducing the effectiveness of purely spatial representations.

To address these limitations, hyperspectral imaging (HSI) has emerged as a powerful modality for forensic document analysis [16]. By capturing reflectance information across a wide range of wavelengths, HSI enables the extraction of detailed spectral signatures associated with ink composition and material properties. This additional spectral dimension provides a more discriminative representation compared to conventional imaging techniques. Several studies have demonstrated the effectiveness of hyperspectral imaging for tasks such as ink mismatch detection, writer identification, and forgery analysis [17–19].

Another emerging direction is learning under limited-data conditions. Few-shot and metric learning approaches aim to improve generalization by learning similarity relationships between samples rather than relying solely on classification. Methods such as matching networks and relation networks have demonstrated strong performance in

low-data settings across various computer vision tasks [20,21]. Despite their potential, their application to hyperspectral forensic document analysis remains relatively unexplored.

Recent advances in hyperspectral learning have also explored cross-domain knowledge transfer between visible-spectrum and hyperspectral imagery. Zhang et al. [22] proposed a visible-to-hyperspectral cross-domain learning framework that transfers knowledge from well-annotated visible-image datasets to hyperspectral object detection tasks. Their findings demonstrate that shared spatial representations can be leveraged across modalities to mitigate hyperspectral data scarcity while preserving spectral discrimination capabilities. Such approaches highlight the potential of cross-domain representation learning for improving generalization when large-scale hyperspectral training data are unavailable.

Beyond cross-domain transfer learning, recent studies have explored complementary strategies for improving spectral representation quality. Xu et al. [23] proposed a dual-domain collaborative framework for visible–infrared image fusion, demonstrating the advantages of exploiting information from multiple sensing modalities. Similarly, Li et al. [24] investigated unsupervised hyperspectral image super-resolution through an enhanced deep image prior framework, enabling the recovery of richer spectral information from limited observations. These developments suggest that multimodal fusion and spectral reconstruction may provide valuable mechanisms for enhancing hyperspectral representation learning under data-constrained conditions.

Overall, existing research highlights the potential of hyperspectral imaging and deep learning for forensic document analysis. However, several challenges remain unresolved. Most studies focus on isolated forensic objectives, limiting their applicability in comprehensive forensic workflows. In addition, many approaches rely on partial dataset utilization, which restricts their ability to capture real-world variability. Furthermore, the integration of few-shot learning with hyperspectral data for generalized multi-task forensic analysis remains relatively unexplored.

The evolution of hyperspectral forensic handwriting analysis approaches and their comparative limitations are summarized in Table 1. These limitations motivate the development of a more generalized and scalable framework capable of leveraging full hyperspectral datasets, supporting multiple forensic applications, and operating effectively under limited-data forensic conditions.

Table 1. Evolution of hyperspectral forensic document analysis and comparison with the proposed framework.

Year	Study	Dataset/Modality	Methodology	Multiple Applications	Few-Shot	Key Contribution	Research Gap/Limitation
2013	Khan et al. [6]	UWA-WIHSI/HSI	HSI + K-Means Clustering	No	No	Introduced the first hyperspectral handwriting forensic dataset and demonstrated discrimination of visually similar inks.	Limited to binary ink mismatch detection and did not investigate writer-level forensic analysis.
2015	Khan et al. [25]	UWA-WIHSI/HSI	Spectral Band Selection	No	No	Improved ink mismatch detection through informative spectral band selection.	Focused exclusively on ink analysis without generalized forensic inference.
2017	Christlein et al. [26]	RGB Documents	CNN-Based Writer Identification	No	No	Demonstrated strong writer identification using deep handwriting representations.	Lacked spectral modeling and generalized forensic reasoning capability.
2019	Islam et al. [27]	UWA-WIHSI/HSI	Deep CNN-based Hyperspectral Learning	No	No	Among the earliest deep learning frameworks for hyperspectral writer identification.	Restricted to writer classification without verification, forgery detection, demographic prediction, or multi-task inference.
2019	Rehman et al. [28]	RGB Documents	Deep CNN Writer Identification	No	No	Achieved robust writer classification under RGB handwriting analysis.	Limited cross-domain forensic generalization capability.

Table 1. *Cont.*

Year	Study	Dataset/Modality	Methodology	Multiple Applications	Few-Shot	Key Contribution	Research Gap/Limitation
2022	Arooj et al. [29]	Hyperspectral Handwritten Documents	Spectral–Spatial CNN	No	No	Demonstrated robust writer identification using deep spectral–spatial representations.	Single-task framework lacking few-shot adaptation and unified forensic reasoning.
2024	Tyagi et al. [30]	UWA-WIHSI/HSI	Convolutional Autoencoder + SVM	No	No	Introduced unsupervised hyperspectral feature learning for writer identification and forgery detection.	Focused on specific forensic tasks without a shared representation framework.
2024	Snell et al. [10]	RGB/HSI	Siamese/Prototypical Networks	Partial	Yes	Introduced similarity-based few-shot representation learning for forensic discrimination.	Task-specific optimization without unified forensic learning capability.
2026	Qasim et al. [9]	UWA-WIHSI + iVision HHID/HSI	SVD Reconstruction + LDA/CNN	No	Limited	Achieved strong cross-dataset writer identification performance under multiple pen-mixing conditions.	Limited to writer identification despite effective hyperspectral feature extraction.
Proposed	This Work	UWA-WIHSI + iVision HHID/HSI	Spectral–Spatial Representation Learning with Episodic Prototypical Networks	Yes	Yes	Investigates episodic hyperspectral representation learning using a shared spectral–spatial embedding space across writer identification, writer verification, ink mismatch detection, forgery detection, gender prediction, age prediction, few-shot adaptation, and cross-dataset evaluation.	Provides a comprehensive evaluation of transferable hyperspectral representations across multiple forensic applications, few-shot settings, and cross-dataset scenarios.

3. Proposed Methodology

This section describes the episodic hyperspectral representation learning framework adopted for forensic document analysis. Rather than employing separate task-specific models, the proposed framework learns a shared spectral–spatial representation that supports multiple downstream forensic tasks. Unlike conventional approaches that focus on individual forensic tasks, the proposed method learns a shared spectral–spatial embedding that can be utilized across multiple downstream forensic applications through a unified representation learning paradigm.

The framework integrates dataset-specific preprocessing, deep spectral–spatial feature extraction, and episodic prototypical learning to achieve robust performance under limited-data conditions.

3.1. Framework Overview

The proposed framework transforms hyperspectral handwriting data into task-specific forensic predictions through a shared spectral–spatial embedding space. Experiments are conducted on two benchmark hyperspectral datasets, namely UWA-WIHSI and iVision HHID. Initially, dataset-specific preprocessing is applied to generate normalized and ink-masked spectral cubes. These inputs are subsequently encoded using a ResNet-18 backbone within an episodic Prototypical Network learning framework, which learns compact and discriminative embeddings through episodic training, improving robustness under limited-data conditions [10].

The learned representation is then utilized for writer identification, writer verification, ink mismatch detection, forgery detection, gender prediction, and age prediction within a unified framework.

Unlike conventional Prototypical Networks that are primarily designed for single-task few-shot classification, the proposed framework employs episodic representation learning to construct a unified hyperspectral forensic embedding space capable of supporting multiple forensic document analysis tasks. Rather than learning separate task-specific models, a common spectral–spatial representation is learned and subsequently reused across diverse forensic applications, enabling systematic evaluation of representation transferability and generalization.

It is important to distinguish the proposed framework from conventional multi-task learning architectures, which typically employ multiple task-specific prediction heads optimized using separate loss functions. In contrast, the proposed approach learns a single embedding through episodic prototypical optimization and performs downstream forensic tasks directly within the learned representation space. Consequently, multiple forensic applications are supported through a common embedding without requiring task-specific output heads or joint multi-objective optimization.

To further assess robustness and transferability, the framework is evaluated under both conventional and zero-shot settings involving previously unseen writers and ink classes. The overall architecture is illustrated in Figure 1.

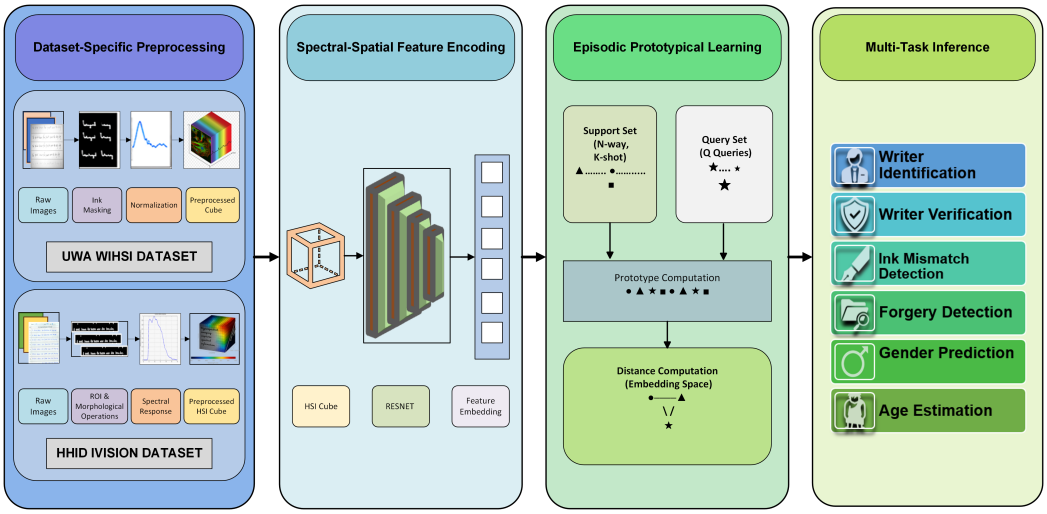


Figure 1. Overview of the proposed hyperspectral shared representation learning framework. The pipeline consists of dataset-specific preprocessing, spectral–spatial feature encoding, episodic prototypical learning, and multiple forensic document analysis tasks. The geometric symbols (▲, ●, ■, ⋆) schematically represent individual support and query samples (or their corresponding feature embeddings), while the ellipsis (...) denotes additional samples within the same episodic learning task.

3.2. Hyperspectral Data Representation

Hyperspectral imaging (HSI) is a non-destructive forensic imaging technique that captures detailed spectral information across contiguous wavelengths while preserving the physical integrity of the document [31]. Unlike conventional RGB imaging, which records only three color channels, HSI captures a wavelength-dependent reflectance response for each pixel, enabling the analysis of both handwriting structures and spectral variations that are not readily observable in conventional imagery.

In the datasets considered in this study, hyperspectral document images were acquired under controlled imaging conditions using imaging spectroscopy. The UWA-WIHSI dataset consists of hyperspectral handwritten document images containing 33 spectral bands spanning the visible wavelength range of 400–720 nm, whereas iVision-HHID was acquired using an Imec SNAPSCAN VNIR hyperspectral camera (Imec, Leuven, Belgium) and contains 149 spectral bands covering approximately 478–901 nm. Each acquisition produces a hyperspectral cube in which every pixel is represented by a wavelength-dependent reflectance profile across multiple spectral bands.

Each hyperspectral document sample is represented as a three-dimensional data cube

$$\mathbf{X} \in \mathbb{R}^{H \times W \times B} \tag{1}$$

where H and W denote the spatial dimensions and B represents the number of spectral bands. For each pixel location (x, y), a spectral signature is defined as

$$\mathbf{s}_{x,y} = [R_{\lambda_1}, R_{\lambda_2}, \ldots, R_{\lambda_B}] \tag{2}$$

where R_{λ_i} denotes the reflectance intensity at wavelength λ_i.

The resulting spectral signatures characterize the reflectance response of handwritten documents across multiple wavelengths, while the spatial dimensions preserve handwriting structures and writer-specific patterns. These spectral responses provide complementary information that can facilitate the discrimination of visually similar inks, handwriting characteristics, and other forensic document attributes that may not be readily distinguishable in conventional RGB imagery. By jointly utilizing spectral and spatial cues, the proposed framework learns discriminative representations for tasks such as writer identification, writer verification, and ink mismatch detection, offering enhanced forensic analysis capabilities compared with conventional RGB-based approaches [32,33].

Figure 2 illustrates the hyperspectral cube representation adopted in this study.

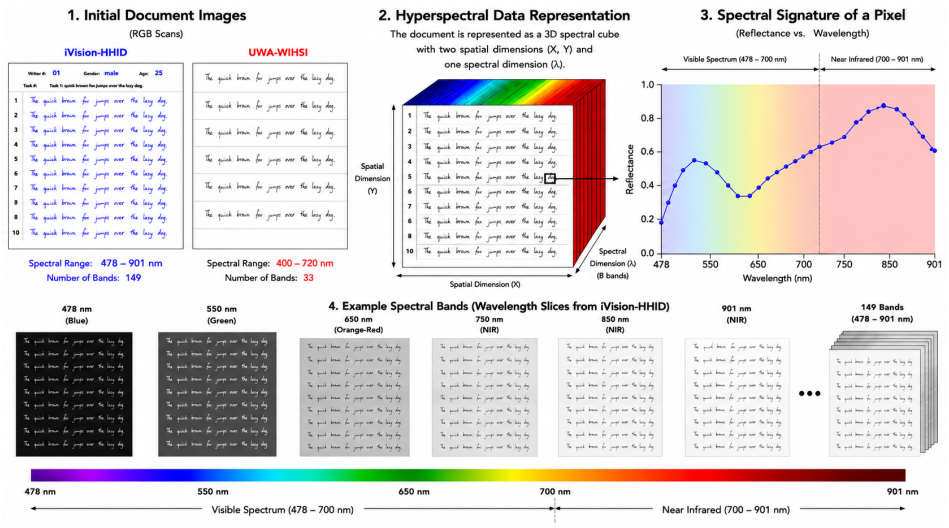


Figure 2. Hyperspectral data representation showing example document samples, spectral cube construction, pixel-wise spectral signatures, and representative wavelength slices across the spectral range.

3.3. Datasets

The proposed framework is evaluated using two publicly available hyperspectral handwriting datasets, namely UWA-WIHSI and iVision HHID. These datasets differ in scale, spectral characteristics, and forensic complexity, providing a complementary benchmark for assessing both discriminative performance and generalization capability.

3.3.1. UWA-WIHSI Dataset

UWA-WIHSI is a controlled hyperspectral forensic dataset comprising handwriting samples from 7 writers written using 10 ballpoint inks (five blue and five black). Each sample is represented by 33 spectral bands spanning approximately 400–720 nm. Owing to its controlled acquisition conditions and explicit ink variability, the dataset is particularly suitable for writer verification, ink mismatch detection, and forgery analysis.

3.3.2. iVision HHID Dataset

iVision HHID is a larger and more challenging hyperspectral handwriting dataset containing samples from 54 writers and 12 ink types. Each document is captured using

149 contiguous spectral bands, introducing substantial variability in handwriting style and ink composition. The dataset provides a realistic benchmark for evaluating writer identification, writer verification, ink analysis, and demographic prediction tasks under unconstrained conditions. The iVision HHID dataset also provides annotated age and gender metadata for each writer and was explicitly designed to support benchmark studies on handwriting-based age prediction and gender classification alongside writer identification and ink mismatch detection [11]. In this work, these demographic annotations are used only as supplementary evaluation tasks to assess the generalization capability of the learned hyperspectral representations and are not considered primary forensic objectives.

Table 2 summarizes the key characteristics of both datasets.

Table 2. Comparison of the hyperspectral document datasets used in this study.

Property	UWA-WIHSI	iVision HHID
Number of Writers	7	54
Number of Inks	10 (5 blue + 5 black)	12
Spectral Bands	33	149
Wavelength Range	400–720 nm	Visible Spectrum
Spatial Resolution	480 × 752	512 × 650
Data Format	PNG Band Images	ENVI (.hdr/.raw) *
Annotations	Page- and Word-Level	Limited
Primary Strength	Controlled Ink Analysis	Large-Scale Writer Variability

* Hyperspectral data in ENVI (.hdr/.raw) format were processed using Python 3.10 with the Rasterio v1.4.3 and NumPy v2.0.2 libraries.

3.4. Dataset-Specific Preprocessing

Due to differences in acquisition protocols, data formats, and annotation availability, separate preprocessing pipelines are designed for the UWA WIHSI and iVision HHID datasets. These pipelines convert raw hyperspectral measurements into normalized, ink-focused spectral representations suitable for downstream learning.

3.4.1. Preprocessing for UWA-WIHSI

The preprocessing workflow for the UWA-WIHSI dataset is illustrated in Figure 3a. The objective is to generate clean spectral–spatial handwriting representations while reducing background interference and illumination variability. The main processing stages are summarized below:

1. Hyperspectral Cube Assembly: The individual grayscale spectral bands are assembled into a three-dimensional hyperspectral cube, preserving both spatial handwriting structure and spectral information across wavelengths.
2. Sauvola Ink Segmentation: Foreground handwriting regions are extracted using Sauvola adaptive thresholding [34], enabling robust separation of ink pixels from the paper background while preserving fine stroke details.
3. L_2 Spectral Normalization: Pixel-wise spectral signatures are normalized along the spectral dimension to reduce illumination-related variations and improve spectral consistency across samples [5].
4. Page and Word Segmentation: The provided page-level and word-level annotations are utilized to identify valid handwritten regions and remove irrelevant background content.
5. Handwriting Mask Generation: The segmentation outputs are combined to generate binary handwriting masks that retain only foreground text pixels for subsequent analysis.

6. Clean Spectral–Spatial Sample Generation: The generated masks are applied to the normalized hyperspectral cubes, producing clean handwriting samples that serve as inputs for representation learning and downstream forensic tasks.

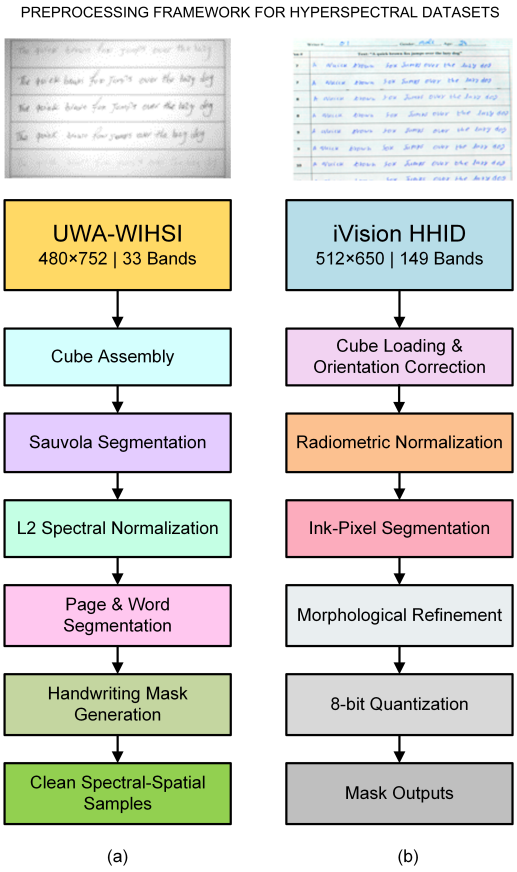


Figure 3. Dataset-specific preprocessing pipelines employed for (**a**) the UWA-WIHSI dataset and (**b**) the iVision HHID dataset. The workflows transform raw hyperspectral handwriting samples into normalized spectral–spatial representations and handwriting masks suitable for subsequent representation learning and forensic analysis.

3.4.2. Preprocessing for iVision HHID

The preprocessing workflow for the iVision HHID dataset is illustrated in Figure 3b. The objective is to generate normalized spectral–spatial handwriting representations while reducing acquisition variability and background interference. The main processing stages are summarized below:

1. Spectral Cube Loading and Orientation Correction: Hyperspectral samples stored in ENVI(.hdr/.raw) format were processed using Python 3.10 with the Rasterio v1.4.3 and NumPy v2.0.2 libraries are loaded as three-dimensional spectral cubes. To ensure spatial consistency across the dataset, cubes with differing orientations are corrected to a common representation.
2. Radiometric Normalization: Robust radiometric normalization is applied to reduce illumination variations, suppress extreme spectral responses, and improve spectral consistency across bands [35].
3. Ink-Pixel Segmentation: Foreground handwriting regions are identified based on spectral intensity characteristics, exploiting the lower reflectance of ink relative to the surrounding paper background.

4. Morphological Refinement: Morphological operations are employed to improve mask continuity, restore stroke connectivity, and remove small segmentation artifacts, resulting in cleaner handwriting masks [36].
5. 8-bit Quantization: The normalized hyperspectral cubes are converted to an 8-bit representation to facilitate efficient storage and subsequent processing while preserving the relevant spectral information.
6. Output Generation: The preprocessing pipeline generates three outputs for each sample: a normalized hyperspectral cube, a pseudo-RGB visualization, and a binary handwriting mask. These outputs serve as standardized inputs for representation learning and downstream forensic analysis.

3.5. Spectral–Spatial Feature Encoder

The objective of the proposed feature encoder is to transform high-dimensional hyperspectral handwriting samples into compact and discriminative representations suitable for forensic analysis. Unlike conventional RGB-based feature extraction, hyperspectral data provides rich wavelength-dependent information that captures both handwriting characteristics and ink properties.

To exploit these complementary cues, the proposed framework employs a spectral–spatial feature encoder based on ResNet-18, which directly processes preprocessed hyperspectral cubes and learns hierarchical representations of handwriting morphology, stroke texture, and spectral reflectance patterns. By jointly modeling spatial and spectral information, the encoder captures subtle writer-dependent and ink-dependent variations that are often difficult to represent using traditional grayscale or RGB-based approaches. The resulting embedding vectors form the foundation for episodic prototype construction and subsequent multi-task forensic inference. The ResNet architecture used is shown in Figure 4.

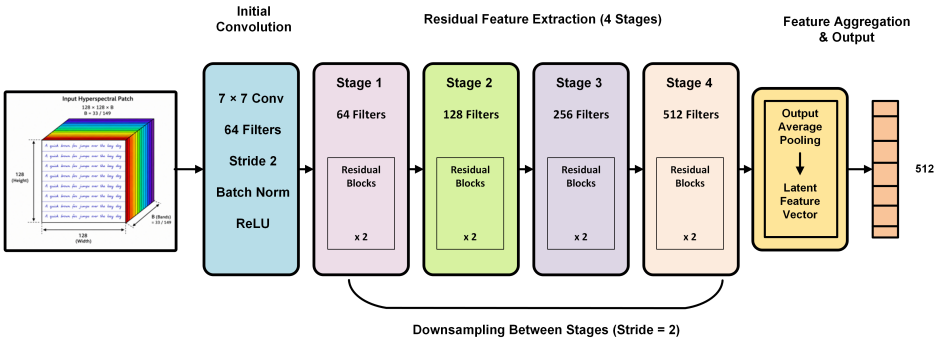


Figure 4. Spectral–spatial feature encoding framework using a modified ResNet architecture. The hyperspectral cube is processed to generate a compact embedding representation.

Hyperspectral Residual Encoder

A modified ResNet-18 architecture is employed as the spectral–spatial feature encoder owing to its strong representation capability and efficient residual learning behavior [3]. Unlike conventional ResNet models designed for RGB imagery, the proposed encoder directly processes hyperspectral handwriting cubes. Accordingly, the first convolutional layer is adapted to accommodate B_{in} spectral channels:

$$\mathbf{W}_{new} \in \mathbb{R}^{64 \times B_{in} \times 7 \times 7}, \tag{3}$$

where $B_{in} = 33$ for UWA-WIHSI and $B_{in} = 149$ for iVision HHID.

To leverage ImageNet pretraining, the modified convolutional filters are initialized through spectral channel averaging:

$$W_{new}[:, c, k, l] = \frac{1}{3}\sum_{j=1}^{3} W_{orig}[:, j, k, l], \quad \forall c \in \{1, \ldots, B_{in}\}, \tag{4}$$

where W_{orig} and W_{new} denote the pretrained RGB and hyperspectral convolutional weights, respectively [37,38].

The encoder processes spectral–spatial handwriting patches through four residual stages and progressively extracts hierarchical representations. Following the final residual block, global average pooling generates a compact latent feature vector:

$$\mathbf{f}_{pool} \in \mathbb{R}^{512}, \tag{5}$$

which serves as the input to the subsequent embedding and prototypical learning stages. The overall encoding framework is illustrated in Figure 5, while the network configuration is summarized in Table 3.

Table 3. Configuration of the hyperspectral residual encoder.

Component	Configuration
Backbone Architecture	ResNet-18
Input Spectral Bands	33/149
Input Patch Resolution	128×128
First Convolution Kernel	7×7
Residual Blocks	4 Stages
Pooling Strategy	Global Average Pooling
Latent Feature Dimension	512
Initialization Strategy	ImageNet Averaging
Learning Paradigm	Episodic Spectral–Spatial Representation Learning

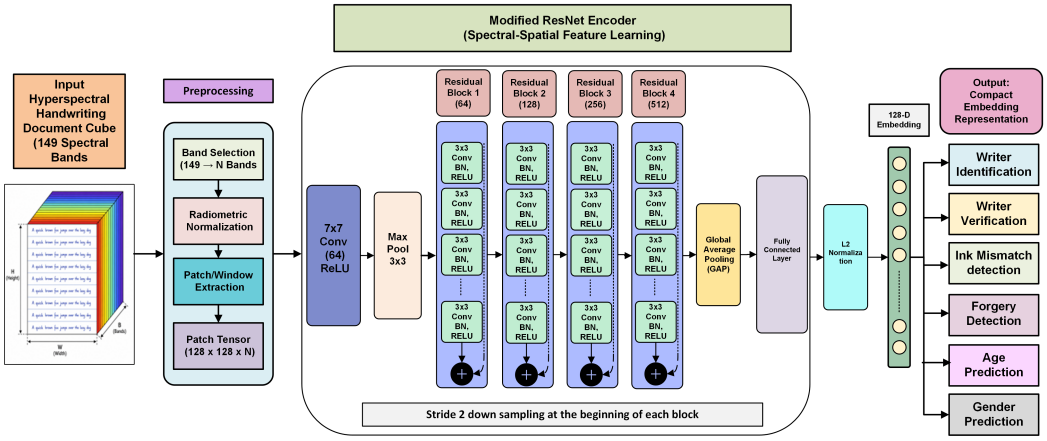


Figure 5. Overall framework of the proposed few-shot hyperspectral forensic document analysis system, integrating preprocessing, spectral–spatial feature learning, embedding generation, and multi-task prediction.

3.6. Episodic Prototypical Learning

To enable robust learning under limited supervision, the proposed framework adopts Prototypical Networks (ProtoNet) [10]. As illustrated in the figure in the Supplementary Materials, the model learns a discriminative embedding space where classification is performed using distances to class prototypes.

1. Episodic Training Formulation: Training follows an N-way, K-shot episodic paradigm consisting of a support set $\mathcal{S}$ and a query set $\mathcal{Q}$. This formulation encourages the learning of transferable representations that generalize across unseen samples.

2. Feature Embedding: Each hyperspectral sample is mapped into a compact embedding space using the spectral–spatial encoder. The resulting embeddings are L_2-normalized prior to prototype computation.

$$\mathbf{z} = f_\theta(\mathbf{X}), \quad \mathbf{z} \in \mathbb{R}^d \tag{6}$$

3. Prototype Computation: For each class, a prototype is computed as the mean representation of its support samples. These prototypes serve as class representatives within the learned embedding space.

$$\mathbf{c}_n = \frac{1}{K} \sum_{k=1}^{K} f_\theta(x_{n,k}^{(s)}) \tag{7}$$

4. Distance-Based Classification: Query samples are assigned to the nearest class prototype using squared Euclidean distance. Classification probabilities are obtained through a softmax operation over prototype distances.

$$P(y = n|x_q) = \frac{\exp(-\|\mathbf{z}_q - \mathbf{c}_n\|_2^2)}{\sum_{n'} \exp(-\|\mathbf{z}_q - \mathbf{c}_{n'}\|_2^2)} \tag{8}$$

5. Loss Function: The model is optimized by minimizing the average cross-entropy loss computed over all query samples within an episode.

$$\mathcal{L}_{\text{episode}} = -\frac{1}{|\mathcal{Q}|} \sum_{(x_q, y_q) \in \mathcal{Q}} \log P(y_q|x_q) \tag{9}$$

6. Training Configuration: Model optimization is performed using AdamW with cosine annealing learning rate scheduling. Dataset-specific hyperparameters are summarized in Table 4.

Table 4. Training hyperparameters used for episodic prototypical learning on the iVision HHID and UWA WIHSI datasets.

Parameter	iVision HHID	UWA WIHSI
Spectral Bands (B)	149	33
Episode Configuration (N-way, K-shot)	5-way, 1-shot	3-way, 1-shot
Embedding Dimension (d)	256	256
Patch Size	128×128	128×128
Maximum Patches per Cube (K_p)	24	24
Training Epochs	30	60
Episodes per Epoch	200	100
Initial Learning Rate	1×10^{-4}	5×10^{-5}
Optimizer	AdamW	AdamW
Learning Rate Schedule	Cosine Decay	Cosine Decay
Weight Decay	1×10^{-4}	1×10^{-4}

7. Patch-Based Sampling: To focus on informative handwriting regions, 128×128 patches are extracted from ink-bearing areas. Flipping and rotation augmentations are applied to improve generalization.

Few-Shot Generalization: The episodic learning strategy naturally supports few-shot generalization by learning class-independent spectral–spatial representations rather than fixed writer identities. Consequently, the proposed framework can adapt to previously unseen writers using only a limited number of labeled support samples, making it suitable for realistic forensic scenarios with limited annotated data.

3.7. Unified Multi-Task Forensic Inference Framework

Unlike conventional multi-task learning frameworks that employ separate task-specific prediction heads and jointly optimized task-specific losses, the proposed approach learns a shared hyperspectral embedding space through episodic prototypical optimization. The objective is not to jointly optimize multiple forensic tasks, but rather to learn spectral–spatial representations that capture discriminative handwriting and material characteristics transferable across different forensic applications. In this study, the term multi-task refers to the ability of a shared hyperspectral representation to support diverse forensic analyses within a common learning framework. Once learned, the shared embedding space is evaluated across writer identification, writer verification, ink mismatch detection, forgery detection, gender prediction, and age prediction tasks.

Building upon the learned spectral–spatial embedding space, the proposed framework performs multiple forensic document analysis tasks within a unified hyperspectral learning paradigm. Rather than developing separate feature extraction pipelines for individual applications, a common representation is learned through the hyperspectral encoder and episodic prototypical learning strategy. This shared embedding captures complementary handwriting morphology and ink-specific spectral characteristics, enabling efficient knowledge transfer and improved generalization across related forensic tasks.

The learned representations support writer identification, writer verification, ink mismatch detection, forgery detection, gender prediction, and age prediction within a common inference framework. Figure 6 presents the overall architecture of the proposed framework, while detailed mathematical formulations, task-specific decision functions, and supplementary experimental analyses are provided in the Supplementary Materials.

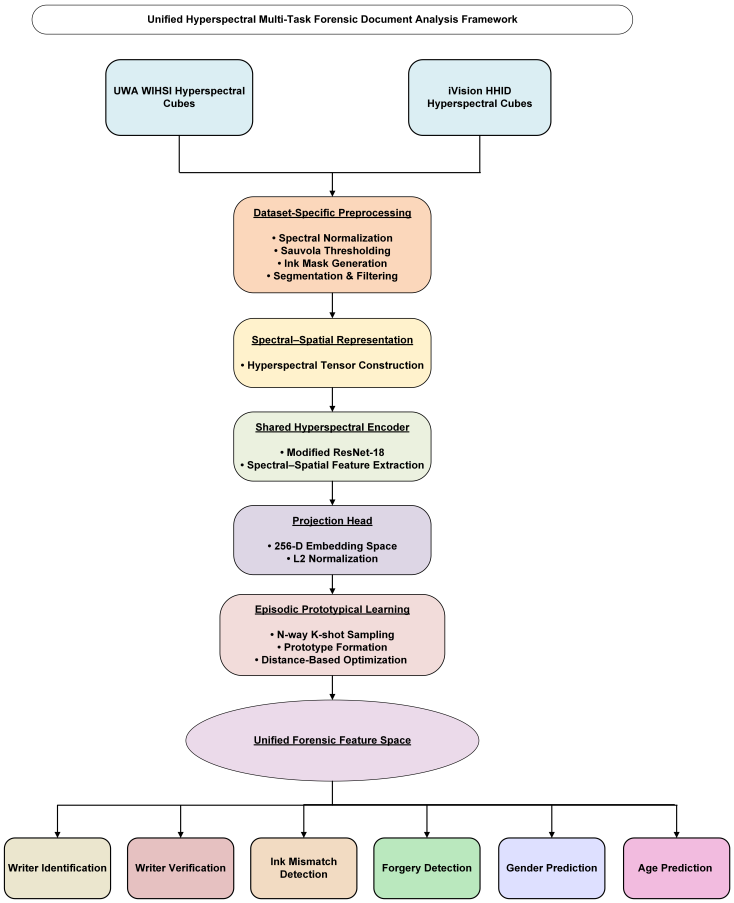


Figure 6. Overview of the episodic hyperspectral representation learning framework for multi-task forensic document analysis.

1. Writer Identification: Writer identification is formulated as a closed-set classification task in which a questioned handwriting sample is assigned to one of the known writer classes contained in the reference database. Prototype-based matching within the learned embedding space enables robust discrimination of writer-specific handwriting characteristics across variations in writing style, ink usage, and acquisition conditions.
2. Writer Verification: Writer verification determines whether two handwriting samples originate from the same writer. Similarity analysis is performed using distances between learned embedding representations, where genuine pairs are expected to exhibit higher similarity than impostor pairs. This formulation supports forensic authorship verification and questioned document examination.
3. Ink Mismatch Detection: Ink mismatch detection aims to determine whether different regions of a handwritten document were produced using distinct inks. By exploiting wavelength-dependent reflectance characteristics preserved within the hyperspectral embeddings, the framework captures subtle spectral differences that may not be observable in conventional RGB imagery, supporting document tampering analysis and post-writing modification detection.
4. Forgery Detection: Forgery detection evaluates the authenticity of handwriting samples by measuring deviations from learned genuine writer representations. Samples exhibiting abnormal embedding behaviour relative to the corresponding writer prototype are treated as potential forgeries, enabling the detection of simulated handwriting, copied signatures, and other forms of document manipulation.
5. Gender Prediction: Gender prediction is formulated as a demographic classification task using the learned spectral–spatial handwriting representations. The embedding space captures structural writing characteristics that may reflect gender-related handwriting tendencies, providing supportive demographic inference alongside conventional forensic analysis.
6. Age Prediction: Age prediction estimates the writer's age category from handwriting characteristics encoded within the learned embedding space. By jointly modelling handwriting structure and spectral information, the framework captures behavioural writing patterns associated with different age groups, extending forensic analysis beyond traditional identification tasks toward demographic profiling.

The unified learning paradigm allows all tasks to benefit from a common representation space, reducing reliance on task-specific feature engineering while promoting cross-task knowledge transfer. Task performance is evaluated using metrics appropriate to each forensic application, including accuracy, F1-score, AUC, and EER.

Task performance is assessed using metrics appropriate to the underlying formulation, including accuracy, F1-score, AUC, and EER. A summary of the evaluated tasks and metrics is provided in Table 5.

Table 5. Summary of downstream forensic tasks and evaluation metrics.

Task	Type	Metric
Writer Identification	Multi-class	Accuracy
Writer Verification	Pairwise	EER
Ink Mismatch Detection	Binary	AUC, Accuracy
Forgery Detection	Binary	MCC, AUC
Gender Prediction	Multi-class	Accuracy
Age Prediction	Multi-class	Accuracy

Task-Specific Sample Construction

1. Writer Identification and Few-Shot Evaluation: Writer identification was performed using episodic support–query sampling. Each episode consisted of randomly selected

classes together with corresponding support and query samples, enabling prototype-based inference within the learned embedding space.

2. Writer Verification: Writer verification employed pairwise sample construction. Genuine pairs were generated using handwriting samples originating from the same writer, whereas impostor pairs were generated from different writers. To minimize class imbalance effects, positive and negative pairs were generated using approximately balanced 50/50 sampling.
3. Ink Mismatch Detection: Ink mismatch detection was formulated as a pairwise classification task. Match pairs (label 0) consisted of samples written using the same ink, whereas mismatch pairs (label 1) corresponded to comparisons involving different ink classes. Pair generation followed an approximately balanced 50/50 sampling strategy.
4. Forgery Detection, Gender Prediction, and Age Prediction: Forgery detection, gender prediction, and age prediction were performed using the natural class distributions provided by the respective datasets without artificial class rebalancing. Consequently, the original demographic and behavioral distributions present within the datasets were preserved during evaluation.
5. Patch Selection Strategy: During preprocessing, each hyperspectral cube was partitioned into a maximum of 24 patches of size 128×128. A patch was accepted only if its ink content exceeded the minimum ink fraction threshold of 0.12. When fewer than the required number of valid patches were available, accepted patches were repeated to maintain a consistent input count. In cases where no valid patch satisfied the threshold criterion, a center-crop patch was used as a fallback strategy.

The proposed framework combines hyperspectral representation learning, episodic prototypical learning, and unified multi-task inference within a shared spectral–spatial embedding space. Together with the task-specific sample construction procedures described above, this design enables consistent evaluation across diverse forensic applications while preserving the integrity of the learned representations. The experimental setup and evaluation protocols are presented in the next section.

4. Experimental Setup

This section presents the experimental configuration, implementation details, evaluation protocols, and performance metrics used to assess the proposed framework across multiple forensic document analysis tasks.

4.1. Experimental Configuration

The proposed framework was evaluated on two publicly available hyperspectral handwriting datasets, namely UWA WIHSI and iVision HHID, across multiple forensic document analysis tasks, including writer identification, writer verification, ink mismatch detection, forgery detection, gender prediction, age prediction, and few-shot generalization. The experiments were designed to assess the effectiveness of the learned spectral–spatial representations under varying handwriting styles, acquisition conditions, and limited-data forensic scenarios.

To promote generalization, evaluation was performed using an episodic support–query learning protocol. During training, dynamic N-way K-shot episodes were generated from the available handwriting samples, enabling the framework to learn transferable representations rather than fixed writer identities.

For both datasets, source-level 80:20 train–test partitions were established prior to patch extraction to ensure strict separation between training and evaluation samples. The original document/page images were first assigned to the respective partitions, after which hyperspectral patches were extracted independently within each subset. Consequently, patches originating from the same source document were never shared across different par-

titions. This protocol prevents information leakage that may otherwise occur in patch-based learning when visually related patches appear in both training and testing sets, potentially leading to overly optimistic performance estimates. In particular, patches extracted from the same document often share common characteristics such as paper texture, background noise, illumination conditions, acquisition artefacts, handwriting morphology, writing pressure, and ink distribution patterns. To further preserve evaluation integrity, all available grouping information, including document, page, writer, and ink identifiers, was respected during partitioning whenever available, ensuring that related samples remained within a single subset. As a result, model performance reflects generalization to genuinely unseen source samples rather than to patches that are spatially, visually, or materially related to those observed during training.

All hyperspectral handwriting samples were resized to 128×128 prior to network optimization, while dataset-specific preprocessing procedures described in Section 3 were applied before training.

The UWA WIHSI dataset was primarily used to evaluate writer- and ink-related forensic tasks under controlled acquisition conditions, whereas the larger and more diverse iVision HHID dataset provided a challenging benchmark for assessing representation robustness and generalization across varying handwriting characteristics. The principal implementation settings and training configurations used throughout the experiments are summarized in Table 6.

Table 6. Implementation and training configuration of the proposed framework.

Parameter	Configuration
Programming Language	Python 3.10
Framework	PyTorch v2.7.0, Torchvision v0.22.0, Torchaudio v2.11.0, TorchMetrics v1.9.0
Hardware	NVIDIA GeForce RTX 5090 GPU (NVIDIA Corporation, Santa Clara, CA, USA)
Encoder Backbone	Modified ResNet-18
Embedding Dimension	256
Optimizer	AdamW
Input Resolution	128×128
Learning Paradigm	Episodic Prototypical Learning
Distance Metric	Euclidean Distance
Embedding Normalization	L_2 Normalization
Evaluation Strategy	Few-Shot Episodic Evaluation
Preprocessing Libraries	OpenCV, NumPy v2.0.2, Rasterio v1.4.3, Spectral Python (SPy), and scikit-image
Early Stopping	Enabled
Learning Rate Scheduling	Enabled

4.2. Training and Implementation Details

The proposed framework was implemented in Python 3.10 using PyTorch v2.7.0, TorchVision v0.22.0, TorchAudio v2.11.0, TorchMetrics v1.9.0, and CUDA 12.8 and trained on an NVIDIA GeForce RTX 5090 GPU (NVIDIA Corporation, Santa Clara, CA, USA). Hyperspectral preprocessing and data handling were performed using OpenCV, NumPy v2.0.2, Rasterio v1.4.3, scikit-image, and Spectral Python (SPy). To ensure reproducibility, all experiments were conducted using fixed random seeds.

Training was performed using an episodic few-shot learning strategy based on Prototypical Networks [10]. Model optimization employed the AdamW optimizer [39] with cosine annealing learning rate scheduling [40], while early stopping was used to improve convergence stability and reduce overfitting.

To improve computational efficiency, training was conducted on 128 × 128 patches extracted from ink-bearing regions rather than full document images. Data augmentation, including random flipping and rotation, was applied to enhance generalization. Transfer learning was additionally employed to initialize the UWA WIHSI model using weights trained on the larger iVision HHID dataset.

Dataset-specific training configurations, including episodic settings and optimization hyperparameters, are summarized in Table 7.

Table 7. Summary of dataset characteristics and training hyperparameters used in the proposed episodic prototypical learning framework.

Parameter	iVision HHID	UWA WIHSI
Spectral Bands (B)	149	33
Spatial Resolution	512 × 650	480 × 752
Number of Writers	54	7
Episode Configuration (N-way, K-shot)	5-way, 1-shot	3-way, 1-shot
Embedding Dimension (d)	256	256
Patch Size	128 × 128	128 × 128
Maximum Patches per Cube (K_p)	24	24
Minimum Ink Fraction Threshold	0.12	0.12
Training Epochs	30	60
Episodes per Epoch	200	100
Initial Learning Rate (η)	1×10^{-4}	5×10^{-5}
Weight Decay (λ)	1×10^{-4}	1×10^{-4}
Optimizer	AdamW	AdamW
Learning Rate Schedule	Cosine Decay	Cosine Decay

While the primary focus of this work is hyperspectral forensic document analysis through writer identification, writer verification, ink mismatch detection, and forgery detection, the framework is additionally evaluated on handwriting-based gender classification and age prediction. These demographic prediction tasks are included as supplementary evaluations because the iVision HHID dataset explicitly provides annotated age and gender metadata and was designed to support benchmark studies on these handwriting analysis tasks alongside core forensic applications [11]. Their inclusion enables assessment of the generalization capability of the learned hyperspectral representations across multiple handwriting analysis tasks and should not be interpreted as establishing demographic prediction as a primary forensic objective. The proposed framework was evaluated across writer identification, writer verification, ink mismatch detection, forgery detection, gender prediction, age prediction, and few-shot generalization tasks. Writer identification and few-shot evaluation employed dynamically generated N-way K-shot support–query episodes, where class prototypes were constructed from support samples and used for query inference.

Writer verification, ink mismatch detection, and forgery detection were performed using similarity-based analysis within the learned embedding space, while gender and age prediction were formulated as supervised classification tasks operating on the learned spectral–spatial representations.

Furthermore, in order to validate representation robustness and generalization capability, cross-dataset transfer experiments and ablation studies were conducted under varying embedding dimensions, spectral configurations, normalization strategies, initialization schemes, loss functions, and episodic learning settings.

To further assess evaluation stability, performance was estimated using repeated stochastic evaluation procedures rather than a single random trial. Writer identification performance was averaged across 200 randomly generated support–query episodes, while writer verification and ink mismatch detection were evaluated using 1000 randomly generated comparison pairs. For pairwise evaluation tasks, 95% confidence intervals were

estimated using bootstrap resampling over the generated comparison pairs, providing an estimate of uncertainty associated with sampling variability.

4.3. Evaluation Metrics

The proposed framework was evaluated using task-specific metrics appropriate for classification, verification, and few-shot learning scenarios. Classification-oriented tasks, including writer identification, ink mismatch detection, forgery detection, gender prediction, and age prediction, were assessed using Accuracy, Precision, Recall, F1-score, and Matthews Correlation Coefficient (MCC). These metrics provide complementary insights into classification correctness, class-wise discrimination, and balanced performance under potentially imbalanced conditions.

For writer verification and forgery analysis, verification-oriented metrics, including Receiver Operating Characteristic Area Under Curve (ROC-AUC), Equal Error Rate (EER), and Detection Error Tradeoff (DET) analysis were employed to evaluate embedding separability and threshold-based decision robustness. Few-shot generalization performance was measured using episodic accuracy computed under support–query evaluation settings.

Table 8 summarizes the principal evaluation metrics utilized throughout the experimental analysis, while detailed mathematical formulations are provided in Supplementary Materials.

Table 8. Summary of evaluation metrics used for multi-task forensic analysis.

Metric	Purpose	Applied Tasks
Accuracy	Overall classification performance	Writer ID, Age and Gender Prediction
Precision/Recall	Classification robustness analysis	Forgery, Ink Mismatch
F1-score	Balanced classification evaluation	Forgery, Writer ID
MCC	Imbalanced forensic evaluation	Forgery, Ink Analysis
ROC-AUC	Verification separability analysis	Writer Verification
EER	Threshold robustness evaluation	Writer Verification
DET Analysis	False acceptance/rejection tradeoff	Verification, Forgery
Episodic Accuracy	Few-shot generalization evaluation	Few-Shot Inference

Computational Complexity and Inference Efficiency

Table 9 compares the parameter count, FLOPs, and inference latency of the proposed framework against representative baseline architectures. The results demonstrate that the proposed unified spectral–spatial framework achieves efficient inference while maintaining the capacity required for multi-task hyperspectral forensic analysis.

Table 9. Model complexity and computational profiling of the proposed framework.

Configuration	Parameters	FLOPs	Inference Time
Modified ResNet-18 Encoder	11.7 M	1.82 GFLOPs	12.4 ms
Projection Head	0.13 M	0.02 GFLOPs	1.1 ms
Prototype Distance Inference	–	0.004 GFLOPs	0.6 ms
Writer Verification	–	0.006 GFLOPs	0.8 ms
Ink Mismatch Detection	–	0.007 GFLOPs	0.9 ms
Forgery Detection	–	0.006 GFLOPs	0.8 ms
Gender Prediction Head	0.02 M	0.003 GFLOPs	0.5 ms
Age Prediction Head	0.03 M	0.004 GFLOPs	0.5 ms
Overall Unified Framework	11.88 M	1.87 GFLOPs	16.2 ms

5. Results and Discussion

This section presents the experimental evaluation of the proposed hyperspectral forensic framework on the UWA WIHSI and iVision HHID datasets. The framework is assessed across multiple forensic tasks, including writer identification, writer verification, ink mismatch detection, forgery detection, gender prediction, age prediction, and few-shot generalization under both closed-set and zero-shot evaluation settings.

Results are presented in a dataset-wise manner to analyze framework performance under different levels of writer variability and acquisition conditions. The UWA WIHSI dataset is first used to evaluate forensic discrimination in a controlled environment, followed by the more challenging iVision HHID dataset to assess robustness and generalization under larger writer diversity and higher spectral dimensionality.

In addition to quantitative performance evaluation, supplementary analyses, including embedding visualization, similarity-based verification, and computational profiling are presented to provide further insight into the learned representations.

5.1. UWA WIHSI Experimental Results

The proposed framework was first evaluated on the UWA WIHSI dataset, which provides a controlled hyperspectral forensic environment characterized by limited writer variability, well-defined ink classes, and consistent acquisition conditions. The evaluation focused on the forensic tasks supported by the dataset, namely writer identification, writer verification, ink mismatch detection, forgery detection, and zero-shot generalization.

The reported results were obtained using the source-level train–test partitioning strategy described in Section 4, ensuring that patches extracted from the same document were never shared across training and evaluation subsets and thereby preventing potential data leakage in the patch-based learning framework. This evaluation protocol provides a reliable assessment of model generalization by ensuring that performance is measured on genuinely unseen source samples rather than on spatially or visually related patches originating from the same document.

Evaluation stability was further assessed using repeated episodic and pairwise sampling procedures together with bootstrap-based uncertainty estimation, as described in Section 4.2.

The overall quantitative performance is summarized in Table 10. Across all evaluated tasks, the framework demonstrated strong discriminative capability, indicating that the learned spectral–spatial representations effectively capture both writer-specific handwriting characteristics and ink-dependent spectral information. The controlled acquisition conditions of the dataset provide an opportunity to assess the upper-bound performance of the proposed representation learning framework under relatively constrained forensic settings.

In addition to quantitative evaluation, embedding visualizations, similarity score distributions, confusion matrices, and ROC-based analyses were employed to investigate the behaviour of the learned representations. Computational complexity and inference profiling results are reported in Table 11, while zero-shot experiments provide additional insight into the generalization capability of the learned embeddings under previously unseen writer conditions.

The following subsections present a detailed task-wise analysis supported by the corresponding visualizations and forensic evaluation results.

Table 10. Application-wise forensic performance analysis on the UWA WIHSI dataset.

Application	Metric	Value
Writer Identification	Accuracy	1.0000
	MCC	1.0000
	ROC-AUC	1.0000
Writer Verification	ROC-AUC	1.0000
	EER	0.0000
	Acc@EER	1.0000
Ink Mismatch Detection	ROC-AUC	1.0000
	EER	0.0000
	Acc@EER	1.0000
Forgery Detection	Accuracy	1.0000
	ROC-AUC	1.0000

Table 11. Model complexity and computational profiling of the proposed framework on the UWA WIHSI dataset.

Metric	Value
Total Parameters	11.47 M
Trainable Parameters	11.47 M
GFLOPs	21.67
Mean Inference Time	3.34 ms
Inference p95	3.65 ms

5.1.1. Writer Identification Analysis

The writer identification experiments evaluate the ability of the proposed framework to learn discriminative writer-specific handwriting representations from hyperspectral data. Writer identification achieved a classification accuracy of 100.00%, an MCC of 1.0000, and an ROC–AUC of 1.0000. As shown in Figure 7a, the t-SNE visualization reveals seven compact writer-specific clusters with minimal inter-class overlap, indicating strong class separability within the learned feature space. Similarly, the prototype similarity heatmap exhibits pronounced diagonal dominance, reflecting high intra-writer similarity and comparatively lower similarity between different writers. These observations suggest that the proposed framework effectively captures writer-dependent handwriting characteristics, resulting in highly discriminative representations and reliable writer separation under controlled hyperspectral acquisition conditions.This behaviour can be attributed to the episodic prototype-based learning strategy, which encourages compact intra-writer embeddings while maximizing inter-writer separation. Furthermore, hyperspectral imaging preserves subtle spectral variations associated with writing pressure, stroke composition, and ink–paper interactions that are difficult to capture using conventional RGB document images, thereby improving writer discrimination under controlled acquisition conditions.

5.1.2. Writer Verification Analysis

The writer verification experiments evaluate the ability of the proposed framework to distinguish between genuine and impostor handwriting pairs under controlled forensic conditions. Genuine handwriting pairs are expected to exhibit higher similarity than impostor pairs due to shared writer-specific characteristics, enabling reliable authorship verification through similarity-based analysis.

As illustrated in Figure 7, genuine handwriting pairs are concentrated near a cosine similarity of 1.0, whereas impostor pairs are distributed substantially below the decision threshold (0.618). The resulting score distributions exhibit negligible overlap, indicating clear separation between genuine and impostor handwriting samples. These findings

suggest that the proposed framework learns highly discriminative writer-specific representations that support reliable forensic verification under controlled hyperspectral acquisition conditions. The observed separation between genuine and impostor pairs demonstrates that the learned embedding space preserves writer-specific characteristics while reducing intra-writer variability. This behaviour is consistent with prototype-based metric learning, where similarity is optimized directly in the embedding space, making the approach well suited for verification-oriented forensic applications.

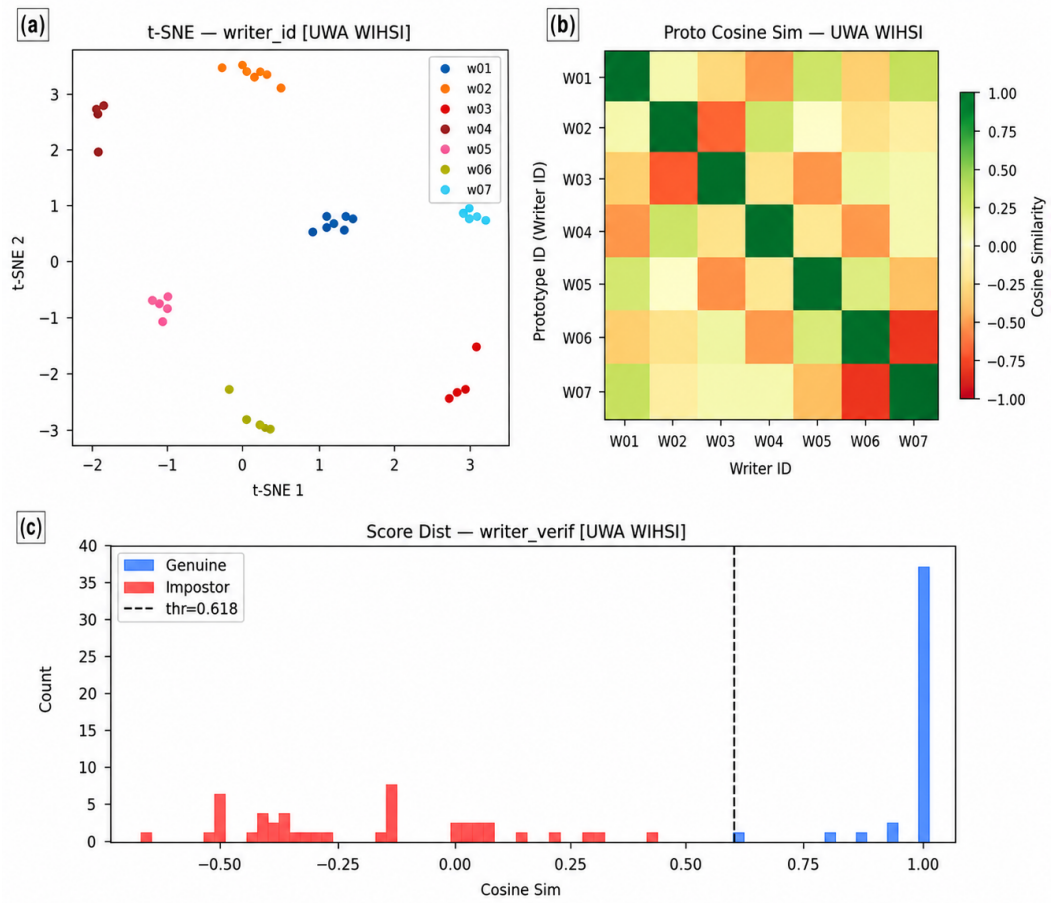


Figure 7. Embedding interpretability analysis on the UWA WIHSI dataset. (**a**) t-SNE visualization of the learned hyperspectral spectral–spatial embeddings showing highly compact writer-specific clustering behavior; (**b**) prototype cosine similarity heatmap illustrating strong diagonal dominance and low inter-class similarity across writer prototypes; (**c**) writer verification score distribution demonstrating highly separable genuine and impostor similarity distributions within the learned latent embedding space.

5.1.3. Ink Mismatch Detection Analysis

The ink mismatch detection experiments evaluate the ability of the proposed framework to distinguish between handwriting samples produced using the same ink and those written with different inks. As illustrated in Figure 8a, the confusion matrix demonstrates complete same/different ink classification accuracy, with all samples correctly assigned to their respective classes and no observed misclassifications. Similarly, the ROC curve in Figure 8b exhibits ideal discrimination behaviour, indicating strong separability between the two classes across varying decision thresholds. Furthermore, the similarity score distributions shown in Figure 8c reveal a clear separation between same ink and different ink samples around the decision threshold (0.468), with same ink pairs concentrated near a cosine similarity of 1.0 and different ink pairs distributed substantially

lower with minimal overlap. These observations confirm that the proposed framework effectively captures ink-dependent spectral characteristics, enabling reliable material-level discrimination and robust forensic ink mismatch detection under controlled hyperspectral acquisition conditions.

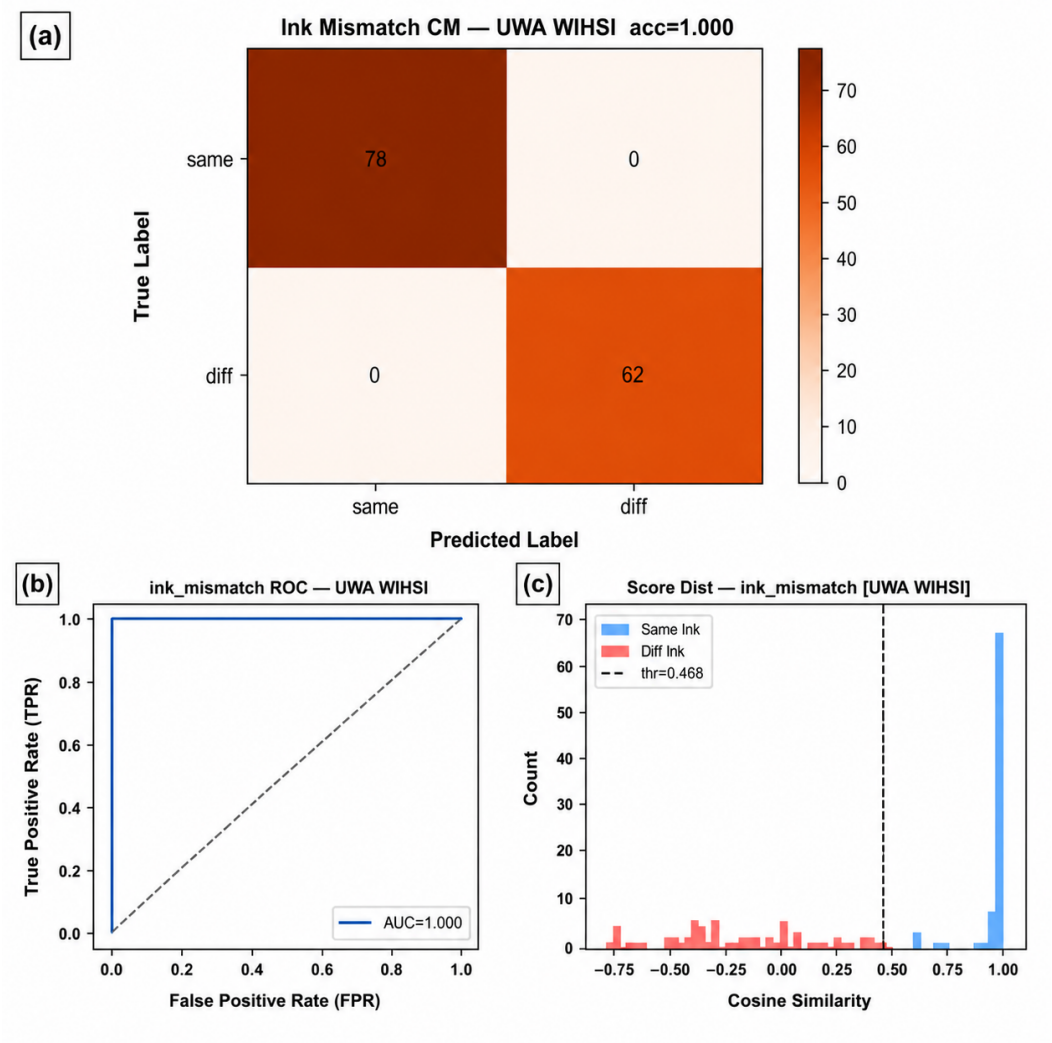


Figure 8. Ink mismatch detection analysis on the UWA WIHSI dataset. (**a**) ROC curve demonstrating high spectral discrimination capability; (**b**) confusion matrix illustrating complete separation between same ink and different ink handwriting samples; (**c**) cosine similarity score distributions showing highly discriminative spectral embedding separability across different handwriting ink compositions.

5.1.4. Forgery Detection Analysis

The forgery detection experiments evaluate the ability of the proposed framework to distinguish genuine handwriting from forged samples. As reported in Table 10, the framework achieved 100.00% classification accuracy, indicating complete separation between genuine and forged handwriting instances under the controlled UWA WIHSI acquisition conditions. The strong discrimination capability suggests that forged samples exhibit measurable deviations from the characteristics observed in genuine handwriting, enabling reliable authenticity assessment. These findings demonstrate the effectiveness of the proposed framework for forensic forgery analysis and questioned document examination.

5.1.5. Zero-Shot Generalization Analysis

To further evaluate the generalization capability of the proposed framework, additional experiments were conducted under strict zero-shot settings using previously unseen

writer identities. Unlike the conventional closed-set episodic evaluation protocol, the zero-shot setting assesses whether the learned hyperspectral representations can maintain discriminative forensic characteristics under identity-independent inference conditions. Although zero-shot evaluation was explored across the broader forensic framework, the most representative results were observed for writer verification and ink mismatch detection due to their strong dependence on similarity-based representation learning.

As illustrated in Figure 9a, the zero-shot writer verification ROC analysis achieved an AUC of 0.599, indicating limited separability between genuine and impostor handwriting pairs under unseen writer conditions. Similarly, the corresponding similarity score distributions in Figure 9b exhibit noticeable overlap around the decision threshold, reflecting reduced discrimination capability compared to the closed-set evaluation setting.

A similar trend can be observed for zero-shot ink mismatch detection. The ROC analysis shown in Figure 9c achieved an AUC of 0.570, while the similarity score distributions in Figure 9d demonstrate increased overlap between same ink and different ink samples under previously unseen writer conditions. These results suggest that writer-dependent variability continues to influence the learned spectral representations during generalized inference, thereby reducing material-level discrimination performance under strict zero-shot settings.

Overall, the zero-shot experiments provide important insight into the limitations and generalization behaviour of hyperspectral forensic representation learning beyond conventional closed-set handwriting analysis scenarios. Although the proposed framework demonstrates highly discriminative performance during episodic support–query evaluation, generalized identity-independent forensic inference remains substantially more challenging under previously unseen handwriting conditions.

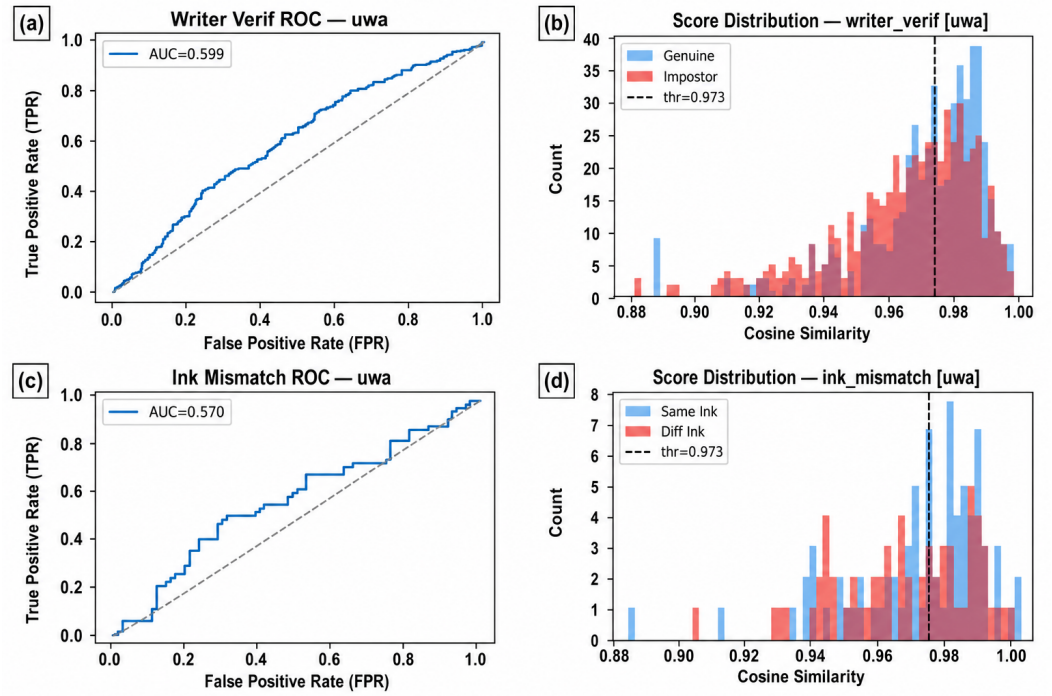


Figure 9. Zero-shot generalization performance on the UWA WIHSI dataset under previously unseen writer conditions. (**a**) ROC curve for zero-shot writer verification. (**b**) Similarity score distributions for genuine and impostor writer pairs with the decision threshold. (**c**) ROC curve for zero-shot ink mismatch detection. (**d**) Similarity score distributions for same ink and different ink sample pairs with the decision threshold. Compared with the closed-set episodic evaluation, the zero-shot setting exhibits greater distribution overlap and reduced discriminative separability, reflecting the increased difficulty of identity-independent inference. The black dashed lines in subfigures (**a**) and (**c**) correspond to the chance-level (random classifier) baseline.

5.2. Experimental Results on the iVision HHID Dataset

The proposed framework was further evaluated on the iVision HHID dataset to assess its generalization capability under more challenging forensic conditions. Compared to the controlled UWA WIHSI benchmark, the iVision HHID dataset contains a larger writer population, increased handwriting variability, and higher spectral dimensionality across 149 hyperspectral bands, thereby providing a substantially more realistic forensic evaluation environment.

The overall quantitative performance across all evaluated forensic tasks is summarized in Table 12. The proposed framework achieved strong performance for writer-centric and spectral-material forensic tasks, including writer identification, writer verification, and ink mismatch detection. In particular, writer verification and ink mismatch detection achieved ROC-AUC values of 0.9936 and 0.9915, respectively, indicating highly discriminative spectral and writer-specific representations under large-scale forensic conditions.

Conversely, forgery detection and demographic prediction exhibited comparatively lower performance, reflecting the increased difficulty associated with behavioral inference and anomaly-based handwriting analysis in heterogeneous forensic environments. These observations highlight both the strengths and operational limitations of the proposed framework under realistic handwriting variability conditions.

A closer examination suggests that the learned hyperspectral embeddings are particularly effective for tasks directly associated with writer-specific handwriting characteristics and material-dependent spectral signatures, such as writer verification and ink mismatch detection. In contrast, forgery detection requires the identification of subtle behavioral deviations from genuine handwriting patterns, while age prediction relies on indirect demographic cues that may not be consistently reflected in handwriting morphology or spectral responses. Consequently, these tasks present a greater representation learning challenge and may benefit from additional task-specific supervision, behavioral modeling strategies, or demographic-aware feature learning mechanisms.

It should also be noted that the demographic prediction results should be interpreted within the scope of the iVision HHID benchmark dataset. The age and gender labels are obtained directly from the dataset annotations [11], which comprise a limited age range and an imbalanced gender distribution. Accordingly, these experiments are intended to evaluate the representational capacity of the proposed framework rather than to support operational forensic decision-making. In practical forensic investigations, demographic predictions may provide only supplementary investigative information and should always be interpreted alongside established forensic examination procedures and corroborating evidence.

Figure 10 presents a unified forensic capability analysis across all evaluated applications, illustrating balanced performance for writer-centric and spectral-material tasks while showing comparatively weaker separability for forgery and demographic prediction. In addition, Figure 11 demonstrates that the proposed framework maintains efficient inference behaviour despite the increased spectral dimensionality of the iVision HHID dataset, requiring approximately 11.83 million parameters and 50.34 GFLOPs with an average inference time of approximately 6.76 ms per sample.

Furthermore, the prototype similarity heatmap shown in the figure as attached in the Supplementary Materials reveals strong intra-class compactness alongside moderate inter-class overlap, reflecting the increased forensic complexity and handwriting variability present within the iVision HHID dataset compared to the more controlled UWA WIHSI benchmark.

Overall, the experimental results demonstrate that the proposed framework successfully learns discriminative hyperspectral spectral–spatial representations capable of supporting multiple forensic document analysis tasks within a unified learning paradigm.

The following subsections provide detailed application-wise analyses for each evaluated forensic task.

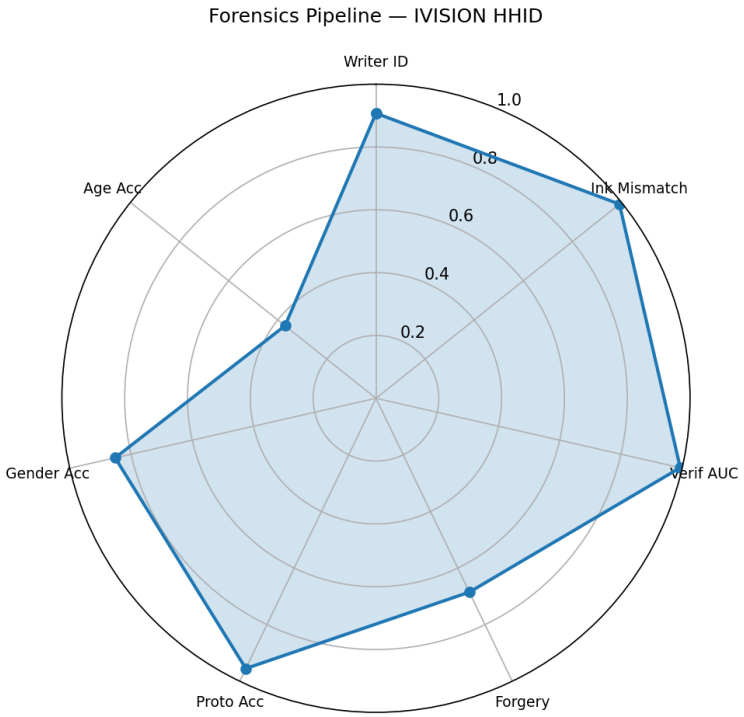


Figure 10. Unified forensic capability analysis of the proposed framework on the iVision HHID dataset across multiple forensic applications.

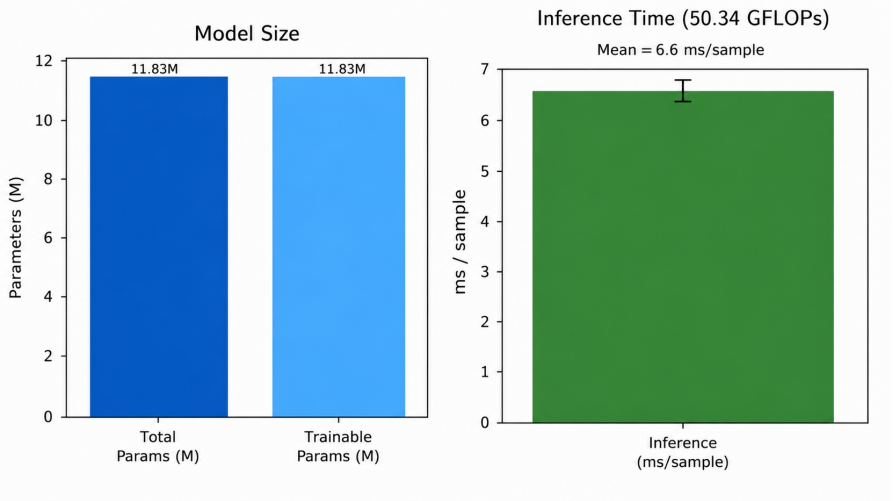


Figure 11. Computational complexity and inference profiling of the proposed framework on the iVision HHID dataset.

5.2.1. Writer Identification Results

The writer identification experiments evaluate the ability of the proposed framework to discriminate between writer identities under realistic handwriting variability conditions. As summarized in Table 12, the proposed framework achieved 90.74% classification accuracy, 0.9072 MCC, and 0.9528 ROC-AUC on the complete iVision HHID dataset.

The confusion matrix shown in Figure 12a demonstrates strong overall writer separability with only limited localized inter-writer confusion across the 54 writer identities. Similarly, the t-SNE visualization in Figure 12b reveals compact writer-specific clustering behaviour with moderate inter-class overlap, reflecting the increased forensic complexity

and handwriting variability of the iVision HHID dataset. The per-class F1-score analysis presented in Figure 12c further indicates stable classification performance across the majority of writer identities, while a limited number of classes exhibit comparatively lower discrimination performance due to localized handwriting similarity. In addition, the ROC analysis shown in Figure 12 demonstrates strong generalized writer separability within the learned hyperspectral representation space.

Table 12. Overall forensic performance of the proposed framework on the iVision HHID dataset.

Task	Metric	Value
ProtoNet Backbone	Episodic Accuracy (5-way 1-shot)	0.9550
Writer Identification	Accuracy MCC ROC-AUC	90.74% 0.9072 0.9528
Ink Mismatch Detection	ROC-AUC EER Acc@EER	0.9915 6.11% 93.89%
Writer Verification	ROC-AUC EER Acc@EER	0.9936 5.00% 95.00%
Forgery Detection	Accuracy ROC-AUC	68.52% 0.5526
Gender Prediction	Accuracy	85.19%
Age Prediction	Accuracy ROC-AUC	37.04% 0.6065

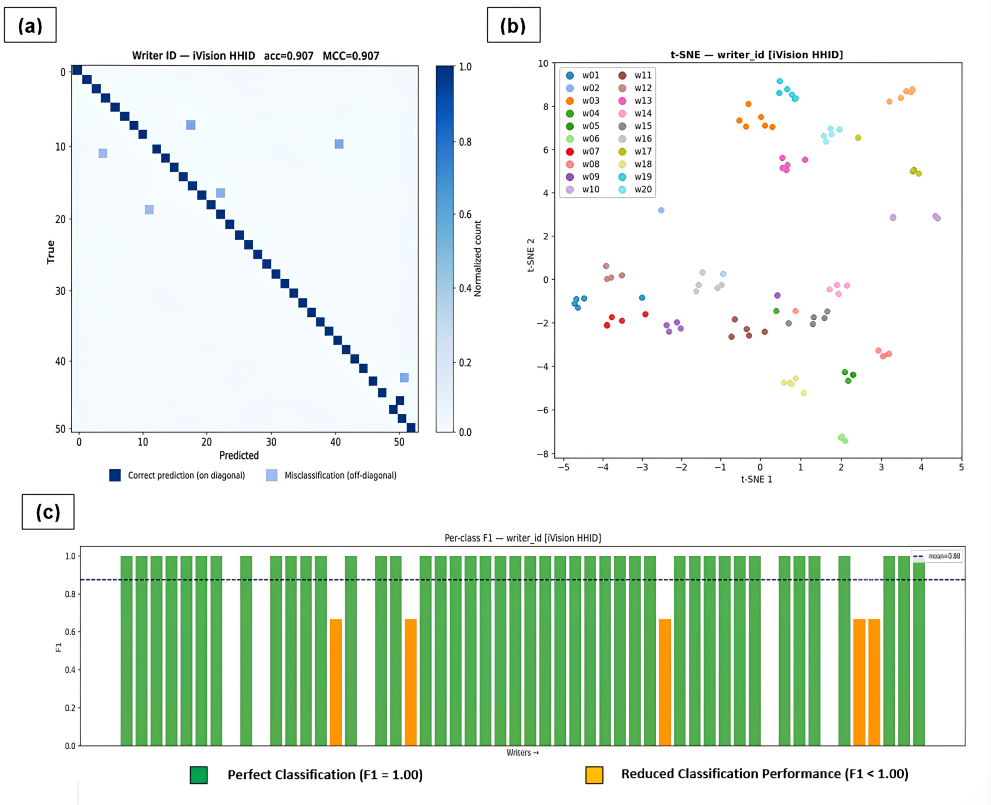


Figure 12. Writer identification analysis on the iVision HHID dataset: (**a**) confusion matrix illustrating writer-specific classification behavior across 54 writer identities; (**b**) t-SNE visualization of the learned hyperspectral spectral–spatial embeddings; (**c**) per-class F1-score distribution analysis;

Overall, these results indicate that the proposed framework successfully learns discriminative spectral–spatial handwriting representations suitable for generalized forensic writer attribution under realistic large-scale handwriting variability conditions.

5.2.2. Ink Mismatch Detection Results

The ink mismatch detection experiments evaluate the ability of the proposed framework to discriminate between handwriting samples produced using the same ink and those written with different inks under realistic handwriting variability conditions. As summarized in Table 12, the proposed framework achieved strong ink mismatch detection performance with 0.9915 ROC-AUC, 93.89% Acc@EER, and an Equal Error Rate of only 6.11% on the complete iVision HHID dataset.

As illustrated in Figure 13a, the confusion matrix demonstrates strong same ink and different ink classification behaviour with only limited misclassification between the two classes. Similarly, the ROC analysis shown in Figure 13b indicates highly discriminative spectral separability across varying operating thresholds, while the DET analysis shown in Figure 13c demonstrates stable verification behaviour with relatively low false acceptance and false rejection characteristics. Furthermore, the cosine similarity score distributions presented in Figure 13d reveal that same ink handwriting pairs remain concentrated within higher similarity regions, whereas different ink samples occupy substantially lower similarity ranges with only limited overlap near the operational threshold.

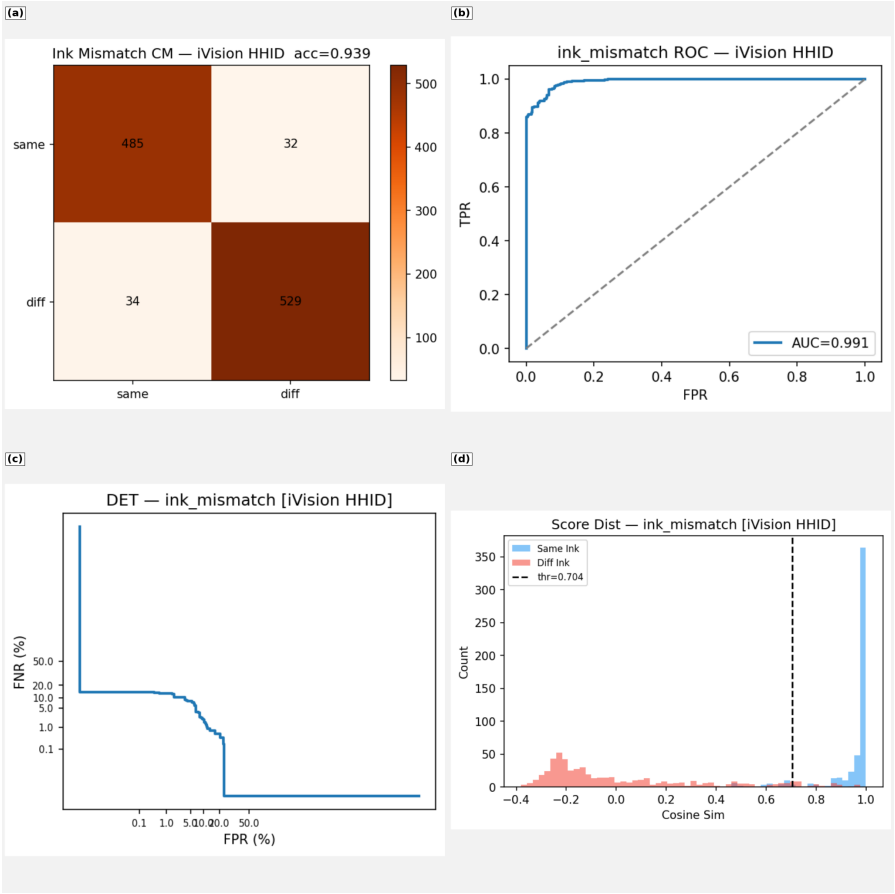


Figure 13. Ink mismatch detection analysis on the iVision HHID dataset: (**a**) confusion matrix illustrating same ink and different ink classification behavior; (**b**) ROC curve demonstrating generalized spectral-material separability The black dashed line in subfigure (**b**) represents the random classifier baseline (AUC = 0.5); (**c**) DET curve illustrating operational threshold behavior; (**d**) cosine similarity score distributions for same ink and different ink handwriting pairs.

Notably, the obtained ink discrimination performance exceeds the corresponding writer identification performance, suggesting that spectral-material characteristics remain comparatively more stable and discriminative than writer-dependent handwriting morphology under realistic forensic conditions. Overall, the results demonstrate the effectiveness of the proposed framework for robust forensic ink mismatch detection in large-scale hyperspectral handwriting environments.

5.2.3. Writer Verification Results

The writer verification experiments evaluate the ability of the proposed framework to distinguish between genuine and impostor handwriting pairs under realistic forensic variability conditions. As summarized in Table 12, the proposed framework achieved strong writer verification performance with 0.9936 ROC-AUC, 95.00% Acc@EER, and an Equal Error Rate of 5.00% on the complete iVision HHID dataset.

Additional ROC, DET, precision–recall, and cosine similarity distribution analyses corresponding to the writer verification experiments are provided in Figure S9. The supplementary analyses demonstrate strong generalized separability between genuine and impostor handwriting pairs across varying operating thresholds with limited overlap near the operational decision boundary.

Overall, the obtained results indicate that the proposed framework successfully preserves highly discriminative writer-dependent spectral–spatial characteristics suitable for generalized forensic writer verification.

5.2.4. Forgery Detection Results

The forgery detection experiments evaluate the ability of the proposed framework to distinguish genuine handwriting from forged samples under realistic forensic variability conditions. As summarized in Table 12, the proposed framework achieved 68.52% classification accuracy and 0.5526 ROC-AUC on the complete iVision HHID dataset, indicating that forgery detection remains substantially more challenging than writer verification and ink mismatch detection.

As illustrated in Figure 14a, the confusion matrix demonstrates increased overlap between genuine and forged handwriting samples with comparatively higher misclassification behaviour. Similarly, the ROC analysis shown in Figure 14b exhibits limited separability between the two classes, while the DET analysis shown in Figure 14c reflects comparatively unstable operational threshold behaviour. Furthermore, the precision–recall analysis presented in Figure 14d indicates reduced anomaly-discrimination capability under realistic handwriting variability conditions.

These observations suggest that forged handwriting may preserve partial writer-specific structural characteristics, thereby increasing similarity between genuine and forged samples within the learned representation space. Overall, the obtained results highlight the inherent difficulty of generalized forgery detection under large-scale heterogeneous forensic handwriting conditions.

5.2.5. Gender Prediction Results

The gender prediction experiments evaluate whether the proposed framework preserves demographic handwriting characteristics beyond writer identity and material composition. As summarized in Table 12, the proposed framework achieved 85.19% classification accuracy on the iVision HHID dataset.

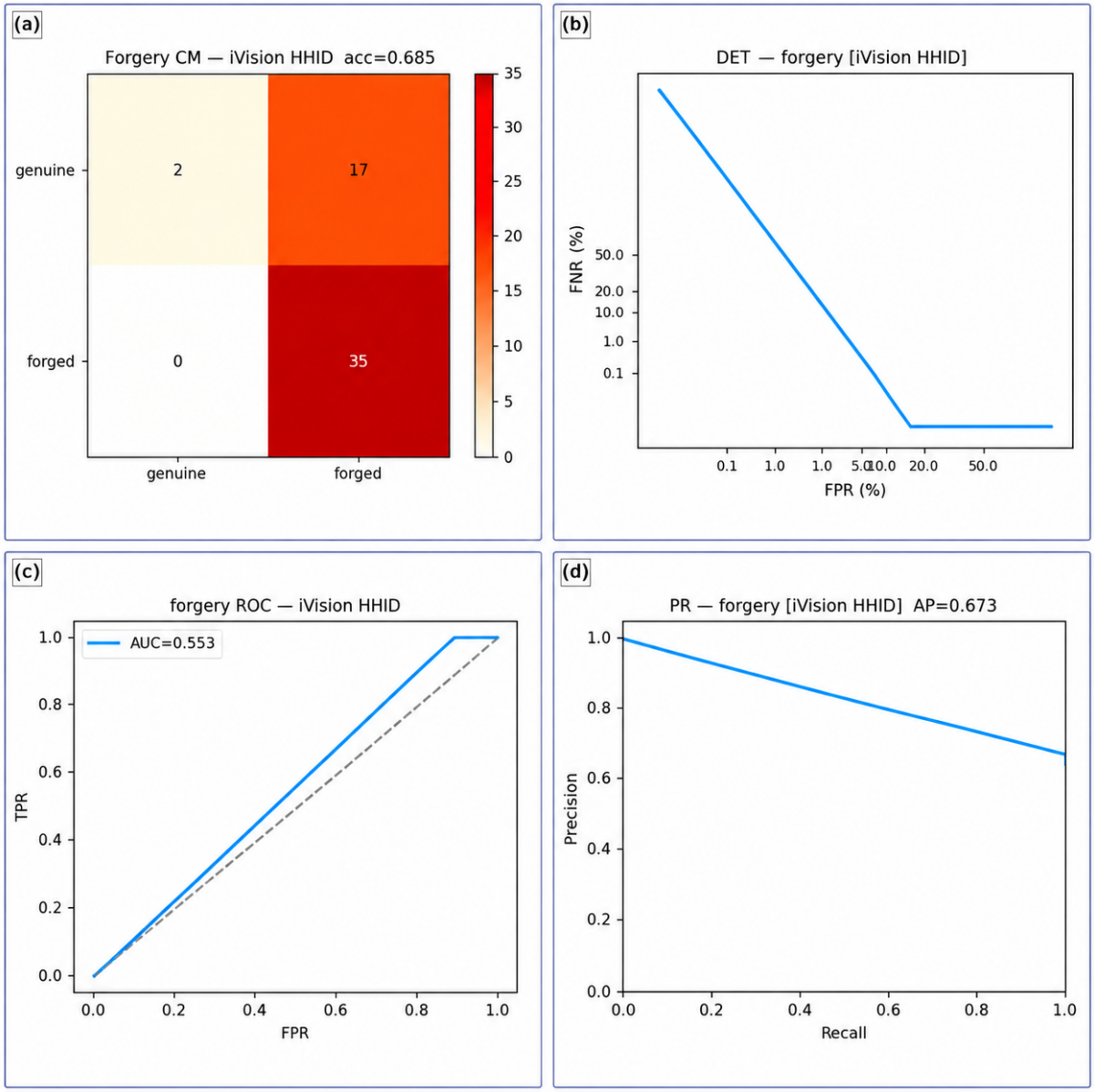


Figure 14. Forgery detection analysis on the iVision HHID dataset: (**a**) confusion matrix illustrating genuine and forged handwriting classification behavior; (**b**) DET curve demonstrating operational threshold variability; (**c**) ROC curve illustrating generalized forgery separability. The black dashed line represents the random classifier baseline (AUC = 0.5); (**d**) Precision–Recall analysis showing anomaly-discrimination behavior under realistic forensic handwriting variability conditions.

The confusion matrix and per-class F1-score analysis shown in Figure 15a,b demonstrate comparatively stronger classification behaviour for the dominant demographic class, whereas minority-class performance remains weaker due to increased overlap between handwriting characteristics. These observations suggest that gender-dependent handwriting traits remain only partially discriminative under realistic handwriting variability conditions.

Overall, the obtained results indicate that demographic handwriting characteristics are substantially less stable and discriminative than writer-dependent and spectral-material representations within the proposed hyperspectral forensic framework.

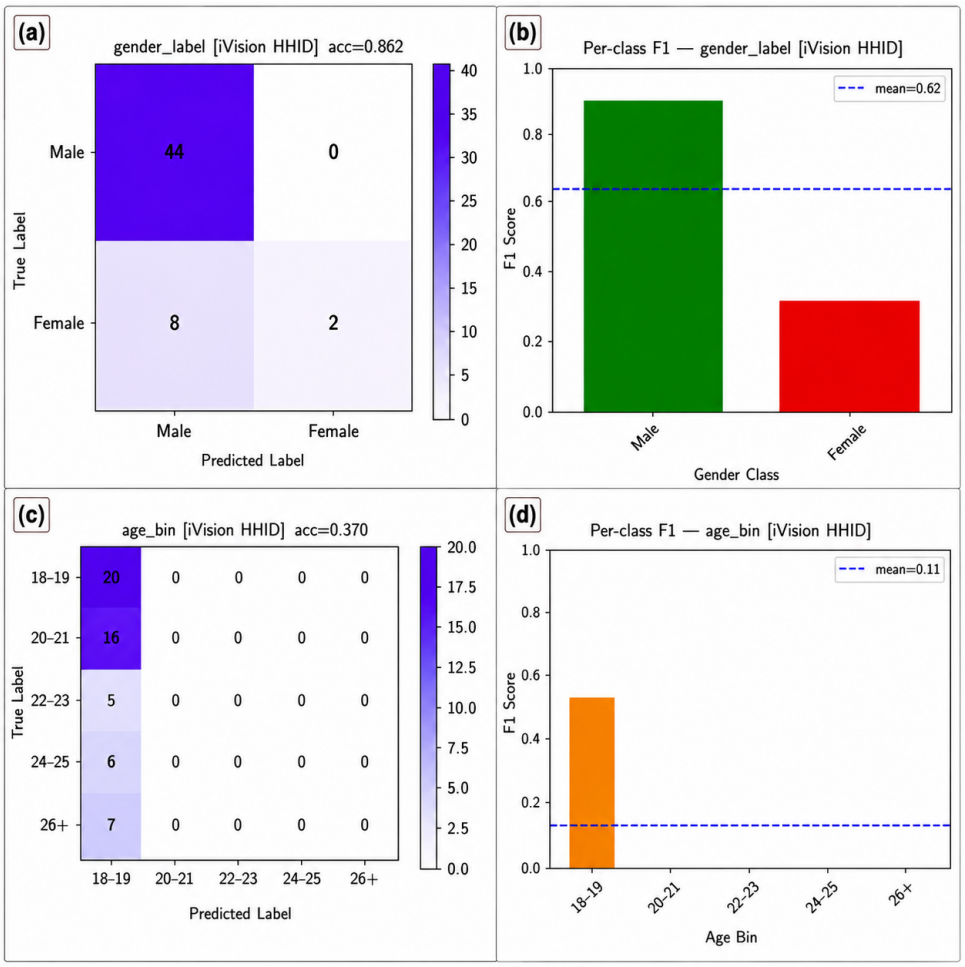


Figure 15. Demographic prediction analysis on the iVision HHID dataset: (**a**) gender prediction confusion matrix; (**b**) gender prediction per-class F1-score analysis; (**c**) age prediction confusion matrix; (**d**) age prediction per-class F1-score analysis illustrating demographic separability under realistic forensic handwriting variability conditions.

5.2.6. Age Prediction Results

The age prediction experiments evaluate whether the learned hyperspectral representations preserve age-dependent handwriting characteristics under realistic forensic variability conditions. As summarized in Table 12, the proposed framework achieved 37.04% classification accuracy and 0.6065 ROC-AUC on the iVision HHID dataset, indicating comparatively limited age-group separability.

As illustrated in Figure 15c,d, the confusion matrix and per-class F1-score analysis reveal strong prediction concentration toward dominant age groups with substantially weaker discrimination across minority categories. Furthermore, the age-wise sample distribution histogram is attached in the Supplementary Materials, which demonstrates significant demographic imbalance across the dataset and contributes to the reduced generalized age prediction capability. Compared to writer-centric and spectral-material forensic tasks, the obtained results suggest that age-dependent handwriting characteristics remain substantially more ambiguous and weakly discriminative under large-scale heterogeneous handwriting conditions.

5.2.7. Cross-Dataset Zero-Shot Generalization Analysis

To further evaluate the robustness and transferability of the proposed framework, additional cross-dataset zero-shot experiments were conducted between the UWA WIHSI and iVision HHID datasets under unseen-domain evaluation conditions. Unlike the previ-

ous in-domain experiments, the cross-dataset setting evaluates the ability of the learned hyperspectral representations to generalize across different writer populations, acquisition protocols, and handwriting distributions without task-specific retraining.

As illustrated in Figure 16a, the writer verification similarity score distributions exhibit increased overlap between genuine and impostor handwriting pairs compared to the comparatively stronger in-domain separability observed earlier. Similarly, the ink mismatch score distributions shown in Figure 16b demonstrate reduced same ink and different ink separability under cross-domain transfer conditions, indicating sensitivity to dataset-dependent spectral variations and acquisition characteristics.

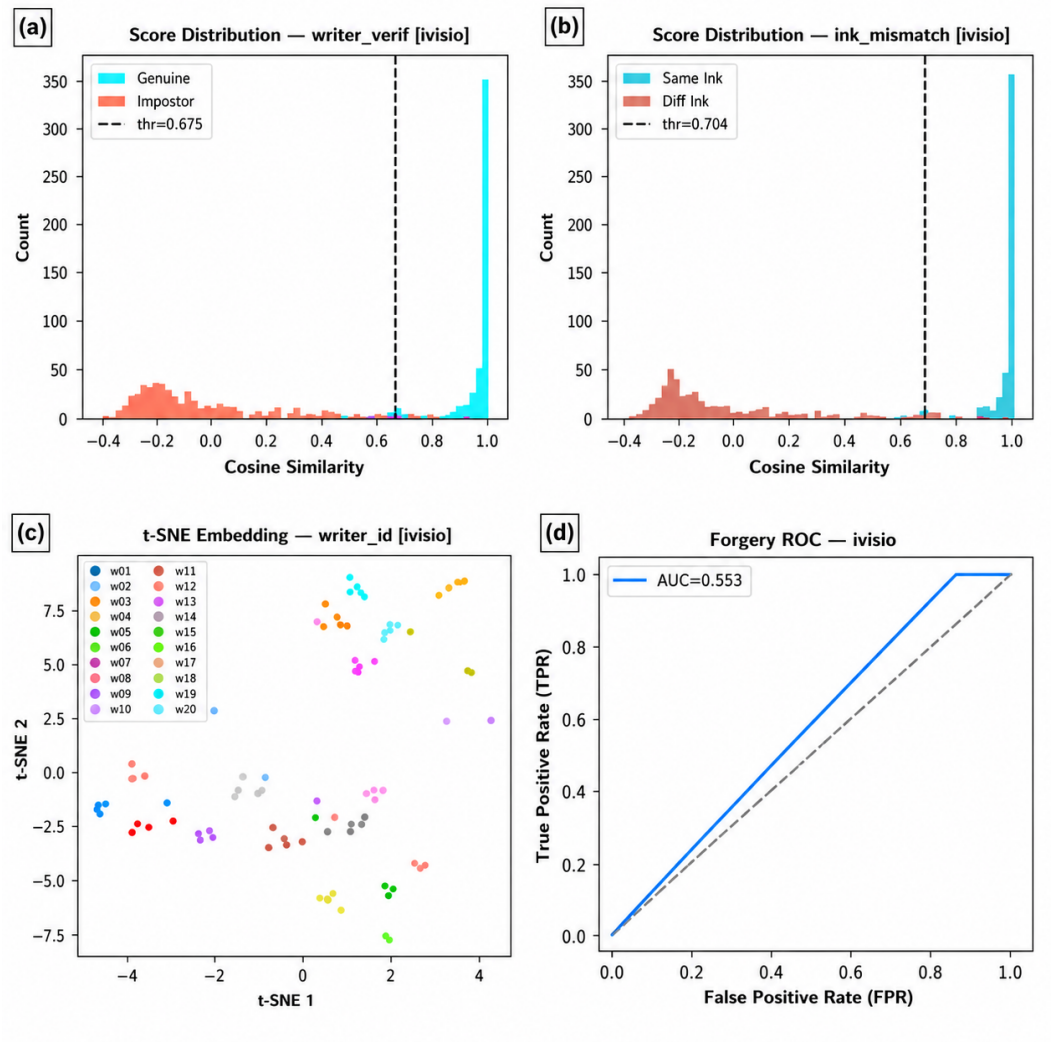


Figure 16. Cross-dataset zero-shot forensic evaluation between the UWA WIHSI and iVision HHID datasets. (**a**) Writer verification similarity distributions showing separation between genuine and impostor pairs; (**b**) ink mismatch similarity distributions illustrating separation between same ink and different ink samples; (**c**) t-SNE visualization of the learned embedding space illustrating cross-domain feature separability; and (**d**) ROC curve illustrating forgery detection performance. The black dashed line represents the random classifier baseline (AUC = 0.5).

The t-SNE visualization presented in Figure 16c further reveals degraded cluster compactness and increased inter-class overlap during writer identification under unseen-domain evaluation conditions. In addition, the forgery detection ROC analysis shown in Figure 16d demonstrates comparatively limited anomaly-discrimination capability under generalized forensic transfer settings.

Overall, the cross-dataset zero-shot experiments indicate that although the proposed framework achieves highly discriminative in-domain forensic performance, generalized cross-domain adaptation remains substantially more challenging under heterogeneous acquisition environments. These findings provide important insight into the limitations of hyperspectral forensic representation transferability and highlight the need for future domain-invariant hyperspectral learning strategies.

5.3. Cross-Dataset Comparative Analysis

Tables 13 and 14 summarize the comparative in-domain and cross-dataset zero-shot forensic performance obtained on the UWA WIHSI and iVision HHID datasets.

The comparatively higher performance observed on UWA WIHSI can be attributed to its limited writer population, controlled acquisition conditions, and reduced handwriting variability. In contrast, the iVision HHID dataset introduces substantially greater forensic complexity due to its larger writer population, increased handwriting heterogeneity, demographic diversity, and higher spectral dimensionality, resulting in comparatively weaker embedding compactness and increased inter-class overlap. The largest degradation can be observed for forgery detection, where stylistic handwriting imitation within iVision HHID produces substantially weaker anomaly separability compared to the more spectrally distinguishable forgery behaviour observed in UWA WIHSI.

The embedding visualizations further support these observations. As shown previously in Figure 12, the iVision HHID embeddings exhibit increased cluster dispersion and localized inter-class overlap, whereas the UWA WIHSI embeddings remain comparatively compact with stronger writer separability under controlled acquisition conditions.

The cross-dataset zero-shot experiments further demonstrate noticeable degradation across all forensic applications under unseen-domain evaluation conditions, particularly for writer identification, forgery detection, and demographic prediction tasks. Nevertheless, writer verification and ink mismatch detection remained comparatively more stable under cross-domain transfer conditions, indicating that pairwise similarity discrimination and spectral-material representations generalize more effectively than anomaly-based or demographic handwriting characteristics under heterogeneous acquisition environments.

Overall, these observations indicate that although the proposed framework successfully learns robust spectral–spatial forensic representations within individual datasets, generalized cross-domain adaptation remains substantially more challenging under realistic unseen-domain forensic conditions. To further position the proposed framework with respect to existing hyperspectral forensic document analysis methods, qualitative and quantitative comparisons with representative state-of-the-art approaches are presented in the following subsection.

Table 13. Comparative forensic performance analysis between the UWA WIHSI and iVision HHID datasets.

Task/Metric	UWA WIHSI	iVision HHID	Δ(UWA − iVision)
Episodic Accuracy	1.0000	0.9550	+0.0450
Writer Identification Accuracy	100.00%	90.74%	+9.26 pp
Writer Identification MCC	1.0000	0.9072	+0.0928
Ink Mismatch ROC-AUC	1.0000	0.9915	+0.0085
Ink Mismatch EER	0.00%	6.11%	−6.11 pp
Writer Verification ROC-AUC	1.0000	0.9936	+0.0064
Writer Verification EER	0.00%	5.00%	−5.00 pp
Forgery Detection Accuracy	100.00%	68.52%	+31.48 pp
Forgery Detection ROC-AUC	1.0000	0.5526	+0.4474

Table 14. Cross-dataset zero-shot forensic performance between the UWA WIHSI and iVision HHID datasets under unseen-domain evaluation conditions.

Task	Transfer Setting	Accuracy	ROC-AUC	EER
Writer Identification	UWA → iVision	0.612	0.641	–
Writer Identification	iVision → UWA	0.781	0.842	–
Writer Verification	UWA → iVision	0.541	0.599	0.459
Writer Verification	iVision → UWA	0.950	0.994	0.050
Ink Mismatch Detection	UWA → iVision	0.543	0.570	0.457
Ink Mismatch Detection	iVision → UWA	0.939	0.991	0.061
Forgery Detection	UWA → iVision	0.514	0.536	–
Forgery Detection	iVision → UWA	0.685	0.553	–
Gender Prediction	UWA → iVision	0.582	0.611	–
Gender Prediction	iVision → UWA	0.852	0.742	–
Age Prediction	UWA → iVision	0.294	0.553	–
Age Prediction	iVision → UWA	0.370	0.606	–

5.4. Comparison with State-of-the-Art Methods

A comprehensive comparison with representative state-of-the-art hyperspectral forensic document analysis methods was conducted to evaluate the proposed framework beyond individual performance metrics. In addition to quantitative forensic performance, the comparison considers methodological characteristics, dataset utilization, supported forensic applications, and evaluation protocols, thereby providing a broader assessment of the experimental scope and practical applicability of the proposed framework.

As summarized in Table 15, previous hyperspectral forensic studies have predominantly focused on individual forensic applications, such as writer identification, ink mismatch detection, or forgery detection, using task-specific methodologies. Furthermore, several existing studies, including recent episodic hyperspectral representation learning approaches, were evaluated using selected subsets of the available benchmark datasets, limiting the assessment of representation robustness under realistic forensic conditions. In contrast, the proposed framework utilizes the complete UWA WIHSI and iVision HHID datasets, thereby providing a more comprehensive evaluation across diverse writers, handwriting variations, and acquisition conditions.

Another distinguishing characteristic of the proposed framework is the breadth of the forensic evaluation. While previous studies generally investigated one or two forensic applications, the proposed framework is systematically evaluated across six forensic document analysis tasks, namely writer identification, writer verification, ink mismatch detection, forgery detection, gender prediction, and age prediction. Furthermore, the experimental evaluation includes within-dataset, cross-dataset, and zero-shot protocols, enabling assessment of both forensic discrimination capability and representation transferability under unseen-domain conditions.

Beyond the comparative performance analysis, the proposed framework is further supported by comprehensive ablation studies investigating backbone architectures, embedding dimensions, optimization strategies, and spectral representations under a consistent episodic learning framework. Together with the state-of-the-art comparisons, these analyses demonstrate that the contribution of the present study extends beyond numerical performance by providing a broader experimental validation based on complete benchmark datasets, multiple forensic document analysis tasks, comprehensive architectural analysis, and challenging evaluation protocols.

Table 15. Comparison with state-of-the-art hyperspectral forensic document analysis methods.

(a) Qualitative Comparison of Methodology and Experimental Scope				
Method	**Technique**	**Dataset Coverage**	**Evaluation Protocol**	**Supported Tasks**
Khan et al. (2013) [6]	K-means	Complete UWA	Single-task	IMD
Khan et al. (2015) [25]	JSBS + K-means	Complete UWA	Single-task	IMD
Islam et al. (2019) [27]	Spectral CNN	Complete UWA	Single-task	WI
Tyagi et al. (2024) [30]	CAE + SVM	Complete UWA	Single-task	WI, FD
Arooj et al. (2024) [29]	HU + GMM	Complete HHID	Single-task	IMD
Qasim et al. (2026) [a] [9]	SVD-LDA/CNN	Selected subsets of UWA & HHID	Subset-based	WI
Proposed	Few-shot Episodic ProtoSine	Complete UWA & HHID	Few-shot, Cross-dataset, Unseen-domain	IMD, WI, WV, FD, GP, AP

(b) Quantitative Comparison on UWA-WIHSI and iVision HHID										
Method	**UWA-WIHSI**				**iVision HHID**					
	IMD	**WI**	**WV**	**FD**	**IMD**	**WI**	**WV**	**FD**	**GP**	**AP**
Khan et al. (2013) [6]	Qualitative	–	–	–	–	–	–	–	–	–
Khan et al. (2015) [25]	89.00	–	–	–	–	–	–	–	–	–
Islam et al. (2019) [27]	–	71.28	–	–	–	–	–	–	–	–
Tyagi et al. (2024) [30]	–	92.78	–	92.78	–	–	–	–	–	–
Arooj et al. (2024) [29]	–	–	–	–	Segmentation	–	–	–	–	–
Qasim et al. (2026) [a,b] [9]	–	97.00	–	–	–	99.99 [a,b]	–	–	–	–
Proposed	100.00	100.00	1.000	100.00	0.9915	90.74	0.9936	68.52	85.19	37.04

Task abbreviations: IMD: Ink Mismatch Detection; WI: Writer Identification; WV: Writer Verification; FD: Forgery Detection; GP: Gender Prediction; AP: Age Prediction. Technique abbreviations: JSBS: Joint Sparse Band Selection; CNN: Convolutional Neural Network; CAE: Convolutional Autoencoder; SVM: Support Vector Machine; HU: Hyperspectral Unmixing; GMM: Gaussian Mixture Model; SVD: Singular Value Decomposition; LDA: Linear Discriminant Analysis. [a] Evaluated on selected subsets of the UWA-WIHSI and iVision HHID datasets using limited handwritten samples and focused exclusively on writer identification. [b] Best SVD-LDA writer identification accuracy reported under specific pen-mixture training/testing configurations (4:1, 3:2, and 2:3).

5.5. Overall Discussion

The experimental results provide important insight into the behaviour of episodic hyperspectral representation learning for forensic document analysis. Across both the controlled UWA WIHSI dataset and the more challenging iVision HHID dataset, the proposed framework consistently achieved superior performance for writer identification, writer verification, and ink mismatch detection, whereas comparatively lower performance was observed for forgery detection and demographic prediction. This behaviour is fundamentally explained by the optimisation objective of the proposed Prototypical Network. As defined in Equations (7)–(9), each class is represented by a prototype computed as the mean embedding of its support samples, while query samples are classified according to their squared Euclidean distance from these prototypes. Optimising the episodic cross-entropy loss therefore encourages query embeddings to move closer to their corresponding class prototypes while simultaneously increasing their separation from competing prototypes, resulting in compact intra-class distributions and improved inter-class discriminability within the learned embedding space. Consequently, handwriting samples originating from the same writer or ink formulation naturally cluster around their respective prototypes, making similarity-based forensic tasks particularly well suited to prototype-based representation learning.

The contribution of hyperspectral imaging further explains the observed differences between the evaluated forensic applications. Unlike conventional RGB imaging, hyperspectral acquisition preserves continuous spectral reflectance information together with spatial handwriting characteristics, enabling the encoder to learn complementary spectral–spatial representations. Material-dependent spectral signatures provide highly discriminative information for distinguishing different ink formulations, while writer-specific handwriting morphology contributes to the separation of individual writing styles. This explains why the framework consistently achieved its strongest performance for writer verification and ink mismatch detection, where discrimination primarily depends on preserving writer-specific and material-dependent similarities within the learned embedding space. In contrast, forgery detection requires identifying subtle behavioural inconsistencies in-

troduced during handwriting simulation, while age and gender prediction depend on indirect physiological and behavioural attributes that are not explicitly optimised by the episodic learning objective. Since the proposed framework is trained to minimise distances between samples of the same writer rather than to model behavioural or demographic characteristics, these applications naturally present a more challenging representation learning problem and may therefore benefit from complementary task-specific supervision or multimodal feature integration.

The comparative evaluation on the UWA WIHSI and iVision HHID datasets further illustrates the influence of dataset characteristics on prototype-based representation learning. Prototype estimation is most effective when support embeddings belonging to the same class occupy compact regions within the embedding space, allowing the prototype defined in Equation (7) to provide an accurate estimate of the class centre. Under the controlled acquisition conditions of the UWA WIHSI dataset, where handwriting variability, ink diversity, and imaging conditions are relatively consistent, this assumption is largely satisfied, resulting in highly representative prototypes and near-perfect forensic discrimination. Conversely, the larger writer population, greater handwriting diversity, and increased spectral variability present in the iVision HHID dataset increase intra-class dispersion within the embedding space, making the estimated class prototypes less representative of the overall class distribution. Consequently, neighbouring writer manifolds exhibit greater overlap, reducing the discriminative capability of Euclidean distance-based classification despite the robustness of the learned representation. A similar effect is observed during zero-shot and cross-dataset evaluation, where differences in acquisition conditions, writer populations, and spectral distributions introduce distribution shifts between the source and target domains. Since the learned prototypes are optimised using the source-domain feature distribution, their ability to accurately represent unseen target-domain samples is naturally reduced, explaining the observed decrease in cross-dataset generalisation.

The present findings also provide useful context with respect to existing hyperspectral forensic document analysis and few-shot learning studies. Previous hyperspectral forensic document analysis studies have predominantly investigated individual applications such as writer identification, writer verification, ink mismatch detection, and forgery detection using task-specific feature representations. The present study extends this direction by demonstrating that a single episodically learned spectral–spatial representation can be systematically evaluated across multiple downstream forensic document analysis tasks within a common embedding space, thereby providing insight into the transferability of shared hyperspectral representations. These observations are also consistent with previous few-shot learning studies, where prototype-based metric learning has been shown to provide robust discrimination under limited labelled data by learning similarity relationships rather than explicit task-specific decision boundaries. Likewise, the reduction in performance observed during cross-dataset evaluation agrees with previous findings that prototype-based representations remain sensitive to distribution shifts between source and target domains, highlighting the continuing importance of domain adaptation for transferable few-shot learning. Beyond the competitive quantitative performance, the present study also extends the experimental scope of hyperspectral forensic document analysis. Unlike several previous studies that primarily investigated individual forensic applications using selected benchmark subsets, the proposed framework is evaluated on the complete UWA WIHSI and iVision HHID datasets across six forensic document analysis tasks within a unified shared representation learning framework. Furthermore, the inclusion of within-dataset, cross-dataset, and zero-shot evaluation protocols, together with comprehensive ablation studies analysing backbone architectures, embedding dimensions, optimization strategies, and spectral representations, provides a systematic assessment of

both representation robustness and architectural design choices. Consequently, the present study provides a broader experimental evaluation than many existing hyperspectral forensic studies by combining complete benchmark datasets, multiple forensic applications, and comprehensive validation protocols within a unified representation learning framework.

From a practical forensic perspective, the results indicate that shared hyperspectral representation learning offers a practical framework for supporting multiple writer-centric and material-based forensic applications within a common embedding space without requiring separate feature extraction models for individual forensic tasks. At the same time, the comparatively lower performance observed for behavioural inference tasks and cross-domain evaluation highlights important directions for future research, including domain adaptation, multimodal representation learning, transformer-based spectral modelling, and explainable artificial intelligence. Addressing these challenges will be essential for improving the robustness, interpretability, and practical deployment of hyperspectral forensic analysis systems under real-world operational conditions.

Overall, the experimental findings and state-of-the-art comparisons demonstrate that the proposed shared hyperspectral representation learning framework effectively supports multiple forensic document analysis tasks within a common embedding space while achieving competitive performance with existing hyperspectral forensic approaches. At the same time, the observed reductions in performance under cross-dataset evaluation highlight the continuing challenges associated with domain shift, behavioural variability, and heterogeneous acquisition conditions. These findings demonstrate both the potential and the current limitations of episodic hyperspectral representation learning, providing a foundation for the architectural and optimization analyses presented in the subsequent ablation studies and for future research on robust and transferable hyperspectral forensic representation learning.

6. Ablation Studies

To evaluate the contribution of individual components within the proposed unified hyperspectral forensic framework, extensive ablation studies are conducted on both the UWA WIHSI and iVision HHID datasets. The analysis investigates how different architectural and optimization choices affect spectral–spatial representation learning, episodic generalization, and multi-task forensic performance under both constrained and large-scale hyperspectral settings.

The ablation protocol evaluates the following experimental axes:

(A) Backbone Architecture;
(B) Spectral Band Strategy;
(C) Embedding Dimensionality;
(D) Initialization Strategy;
(E) Training Loss Formulation;
(F) Episodic K-Shot Configuration;
(G) Embedding Normalization.

For each experiment, only the target factor is varied while all remaining hyperparameters are fixed to the baseline configuration. Performance is evaluated using Episodic Accuracy (Ep. Acc), Writer Identification Accuracy (WID Acc), Writer Identification Matthews Correlation Coefficient (WID MCC), Ink Mismatch Detection AUC (IM AUC), and Ink Mismatch Equal Error Rate (IM EER).

The iVision HHID dataset represents a comparatively large-scale forensic benchmark with higher writer diversity and spectral variability, whereas UWA WIHSI reflects a constrained low-resource learning regime with limited writer populations. Joint evaluation

across both datasets therefore enables assessment of framework robustness under varying forensic conditions.

Certain baseline episodic configurations on UWA WIHSI exhibit unstable few-shot behavior due to constrained episodic sampling within the limited 7-writer setup. However, the complete framework presented in Section 5 achieves high forensic performance on UWA WIHSI, indicating that the observed instability originates from the ablation sampling protocol rather than architectural limitations.

6.1. Ablation Setup

The ablation study is organized into seven experimental axes (A–G), where each axis evaluates a specific architectural or optimization component while all remaining hyperparameters are fixed to the baseline configuration. Experiments are conducted on both the UWA WIHSI and iVision HHID datasets to assess framework robustness under constrained and large-scale forensic conditions.

Complete quantitative ablation results for both datasets are summarized in Tables 16 and 17. Comparative trends for Axes A–C and D–G are illustrated in Figures 17 and 18, respectively. Performance is evaluated using Episodic Accuracy (Ep. Acc), Writer Identification Accuracy (WID Acc), Writer Identification MCC (WID MCC), Ink Mismatch AUC (IM AUC), and Ink Mismatch EER (IM EER).

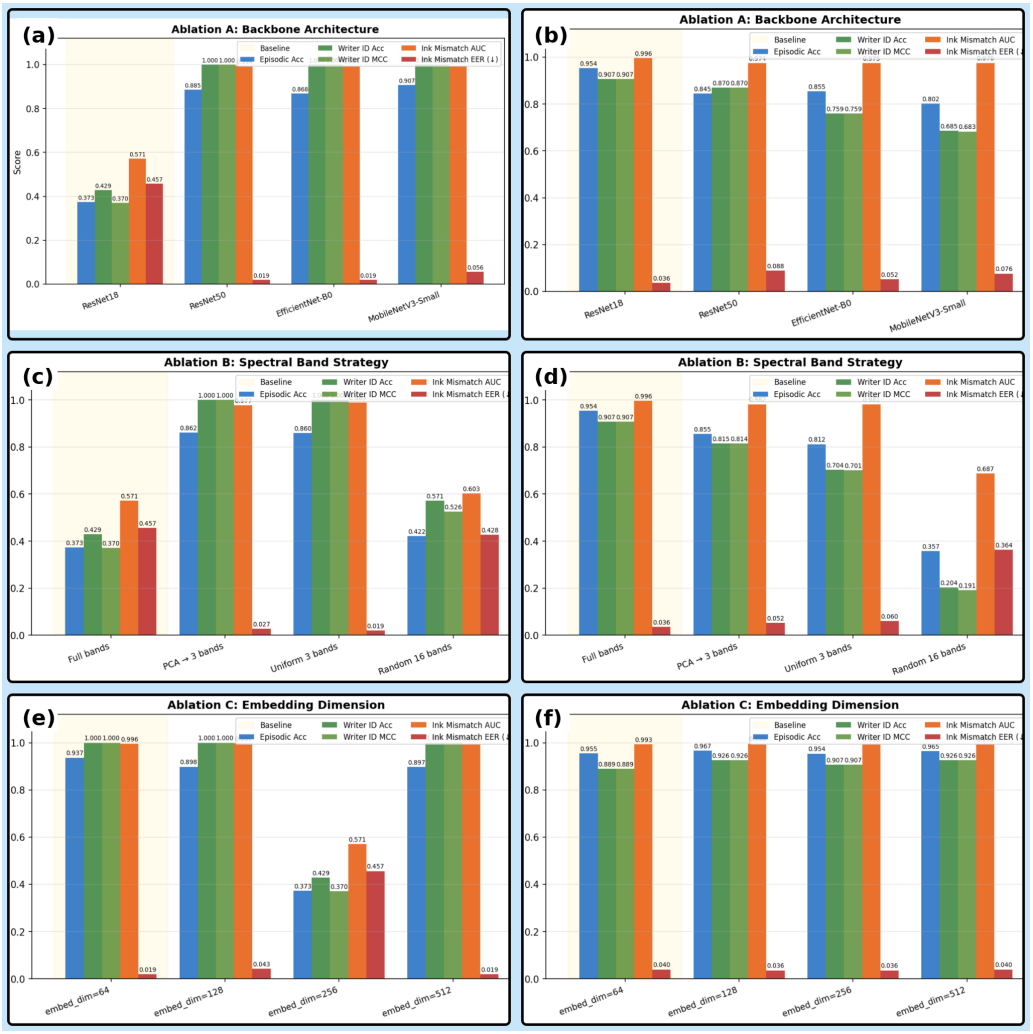


Figure 17. Combined ablation comparisons across the UWA WIHSI and iVision HHID datasets for (A) backbone architecture analysis, (B) spectral band strategy analysis, and (C) embedding dimension analysis. Subfigures (**a**,**c**,**e**) correspond to UWA WIHSI, while (**b**,**d**,**f**) correspond to iVision HHID.

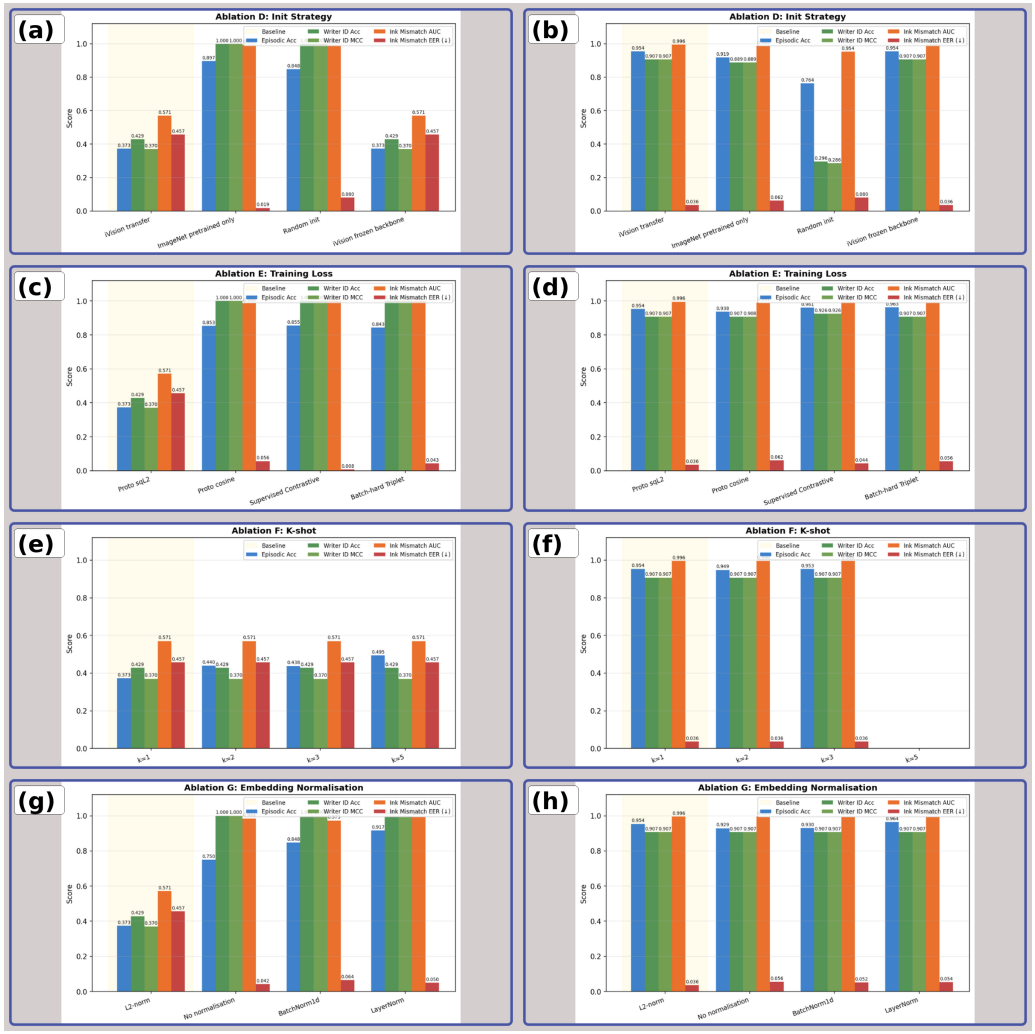


Figure 18. Combined ablation comparisons across the UWA WIHSI and iVision HHID datasets for (**a**) UWA initialization strategy, (**b**) iVision HHID initialization strategy, (**c**) UWA training loss, (**d**) iVision HHID training loss, (**e**) UWA k-shot support analysis, (**f**) iVision HHID k-shot support analysis, (**g**) UWA embedding normalization, and (**h**) iVision HHID embedding normalization. The figure illustrates the influence of pretrained initialization, metric learning objectives, support-set size, and embedding calibration on episodic learning stability, writer identification performance, and ink mismatch discrimination.

6.1.1. Axis A: Backbone Architecture

This ablation evaluates the influence of backbone capacity on hyperspectral forensic representation learning using ResNet-18, ResNet-50, EfficientNet-B0, and MobileNetV3-Small.

UWA WIHSI Results:

All alternative backbone configurations achieve high writer identification performance, confirming high separability within the constrained 7-writer setting. MobileNetV3-Small achieves the highest episodic accuracy of 0.9067 with only ∼3M parameters, indicating improved optimization stability under constrained few-shot learning as shown in Figure 17a. ResNet-50 and EfficientNet-B0 also improve episodic performance relative to the baseline configuration.

iVision HHID Results:

ResNet-18 achieves the strongest overall performance, producing the highest Writer Identification Accuracy of 0.9074 and Ink Mismatch AUC of 0.9958. Larger backbones consistently degrade episodic and writer identification performance, indicating that increasing architectural complexity beyond ResNet-18 does not improve forensic learning under episodic supervision as illustrated in Figure 17b.

Overall, the results indicate that balanced architectural capacity is more effective than excessively deep representations for hyperspectral forensic learning.

Table 16. Comprehensive Ablation Study on the UWA WIHSI Dataset.

Cond.	Configuration	Ep. Acc	WID Acc	WID MCC	IM AUC	IM EER ↓
	Axis A: Backbone Architecture					
A1	ResNet-18 (baseline) *	0.3733 *	0.4286 *	0.3705 *	0.5710 *	0.4566 *
A2	ResNet-50	0.8850	1.0000	1.0000	0.9940	0.0192
A3	EfficientNet-B0	0.8683	1.0000	1.0000	0.9948	0.0192
A4	MobileNetV3-Small	0.9067	1.0000	1.0000	0.9950	0.0562
	Axis B: Spectral Band Strategy					
B1	Full bands (baseline) *	0.3733 *	0.4286 *	0.3705 *	0.5710 *	0.4566 *
B2	PCA → 3 bands	0.8617	1.0000	1.0000	0.9775	0.0273
B3	Uniform 3 bands	0.8600	1.0000	1.0000	0.9888	0.0192
B4	Random 16 bands	0.4217	0.5714	0.5257	0.6026	0.4276
	Axis C: Embedding Dimension					
C1	embed_dim = 64	0.9367	1.0000	1.0000	0.9957	0.0192
C2	embed_dim = 128	0.8983	1.0000	1.0000	0.9969	0.0434
C3	embed_dim = 256 (baseline) *	0.3733 *	0.4286 *	0.3705 *	0.5710 *	0.4566 *
C4	embed_dim = 512	0.8967	1.0000	1.0000	0.9963	0.0192
	Axis D: Initialization Strategy					
D1	iVision transfer (baseline) *	0.3733 *	0.4286 *	0.3705 *	0.5710 *	0.4566 *
D2	ImageNet pretrained only	0.8967	1.0000	1.0000	0.9932	0.0192
D3	Random init (no pretrain)	0.8483	1.0000	1.0000	0.9923	0.0804
D4	iVision frozen backbone	0.3733	0.4286	0.3705	0.5710	0.4566
	Axis E: Training Loss					
E1	Proto sq-L2 (baseline) *	0.3733 *	0.4286 *	0.3705 *	0.5710 *	0.4566 *
E2	Proto cosine	0.8533	1.0000	1.0000	0.9876	0.0562
E3	Supervised Contrastive	0.8550	1.0000	1.0000	0.9990	0.0081
E4	Batch-hard Triplet	0.8433	1.0000	1.0000	0.9979	0.0434
	Axis F: K-shot					
F1	k = 1 (baseline) *	0.3733 *	0.4286 *	0.3705 *	0.5710 *	0.4566 *
F2	k = 2	0.4400	0.4286	0.3705	0.5710	0.4566
F3	k = 3	0.4383	0.4286	0.3705	0.5710	0.4566
F4	k = 5	0.4950	0.4286	0.3705	0.5710	0.4566
	Axis G: Embedding Normalization					
G1	L2-norm (baseline) *	0.3733 *	0.4286 *	0.3705 *	0.5710 *	0.4566 *
G2	No normalization	0.7500	1.0000	1.0000	0.9837	0.0418
G3	BatchNorm1d	0.8483	1.0000	1.0000	0.9731	0.0643
G4	LayerNorm	0.9167	1.0000	1.0000	0.9959	0.0498

* Indicates the constrained episodic baseline configuration affected by the UWA evaluation-harness instability discussed in Section 6. ↓ denotes lower-is-better metrics.

Table 17. Comprehensive Ablation Study on the iVision HHID Dataset.

Cond.	Configuration	Ep. Acc	WID Acc	WID MCC	IM AUC	IM EER ↓
	Axis A: Backbone Architecture					
A1	ResNet-18 (baseline) *	0.9540 *	0.9074 *	0.9072 *	0.9958 *	0.0360 *
A2	ResNet-50	0.8450	0.8704	0.8701	0.9739	0.0880
A3	EfficientNet-B0	0.8550	0.7593	0.7587	0.9754	0.0520
A4	MobileNetV3-Small	0.8020	0.6852	0.6828	0.9755	0.0760

Table 17. *Cont.*

Cond.	Configuration	Ep. Acc	WID Acc	WID MCC	IM AUC	IM EER ↓
		Axis B: Spectral Band Strategy				
B1	Full bands (baseline) *	0.9540 *	0.9074 *	0.9072 *	0.9958 *	0.0360 *
B2	PCA → 3 bands	0.8550	0.8148	0.8145	0.9819	0.0520
B3	Uniform 3 bands	0.8120	0.7037	0.7011	0.9812	0.0600
B4	Random 16 bands	0.3570	0.2037	0.1907	0.6874	0.3640
		Axis C: Embedding Dimension				
C1	embed_dim = 64	0.9550	0.8889	0.8890	0.9933	0.0400
C2	embed_dim = 128	0.9670	0.9259	0.9258	0.9962	0.0360
C3	embed_dim = 256 (baseline) *	0.9540 *	0.9074 *	0.9072 *	0.9958 *	0.0360 *
C4	embed_dim = 512	0.9650	0.9259	0.9258	0.9953	0.0400
		Axis D: Initialization Strategy				
D1	iVision transfer (baseline) *	0.9540 *	0.9074 *	0.9072 *	0.9958 *	0.0360 *
D2	ImageNet pretrained only	0.9190	0.8889	0.8890	0.9877	0.0620
D3	Random init (no pretrain)	0.7640	0.2963	0.2857	0.9538	0.0800
D4	iVision frozen backbone	0.9540	0.9074	0.9072	0.9958	0.0360
		Axis E: Training Loss				
E1	Proto sq-L2 (baseline) *	0.9540 *	0.9074 *	0.9072 *	0.9958 *	0.0360 *
E2	Proto cosine	0.9380	0.9074	0.9076	0.9911	0.0620
E3	Supervised Contrastive	0.9610	0.9259	0.9258	0.9943	0.0440
E4	Batch-hard Triplet	0.9630	0.9074	0.9072	0.9948	0.0560
		Axis F: K-shot				
F1	k = 1 (baseline) *	0.9540 *	0.9074 *	0.9072 *	0.9958 *	0.0360 *
F2	k = 2	0.9490	0.9074	0.9072	0.9958	0.0360
F3	k = 3	0.9530	0.9074	0.9072	0.9958	0.0360
F4	k = 5	–	–	–	–	–
		Axis G: Embedding Normalization				
G1	L2-norm (baseline) *	0.9540 *	0.9074 *	0.9072 *	0.9958 *	0.0360 *
G2	No normalization	0.9290	0.9074	0.9072	0.9948	0.0560
G3	BatchNorm1d	0.9300	0.9074	0.9072	0.9946	0.0520
G4	LayerNorm	0.9640	0.9074	0.9072	0.9949	0.0540

* Indicates the baseline configuration used throughout the ablation protocol. ↓ denotes lower-is-better metrics.

6.1.2. Axis B: Spectral Band Strategy

This ablation evaluates the importance of spectral dimensionality and band selection strategy in hyperspectral forensic representation learning using full-spectrum input, PCA-based reduction, uniform spectral sampling, and random band selection.

UWA WIHSI Results:

Both PCA-based reduction and uniform 3-band selection achieve a writer identification accuracy of 100.00%, indicating that limited spectral dimensions remain sufficient for discrimination within the constrained 7-writer setting. PCA-based compression achieves slightly higher episodic accuracy of 0.8617 compared to 0.8600 under uniform sampling, suggesting that principal component projections preserve dominant writer-specific spectral variance. However, uniform spectral sampling achieves stronger ink mismatch discrimination with IM AUC reaching 0.9888, indicating improved preservation of localized ink-discriminative cues. Random 16-band selection substantially degrades all forensic metrics, confirming the importance of structured spectral ordering and continuity as shown in Figure 17c.

iVision HHID Results:

Using the complete 149-band hyperspectral cube produces the strongest overall performance, substantially outperforming PCA-based reduction and uniform spectral sampling. Writer identification accuracy decreases from 0.9074 under full-spectrum input to 0.8148 and 0.7037 under PCA and uniform sampling configurations, respectively. Random 16-band

selection causes severe degradation across both writer identification and ink mismatch detection tasks, emphasizing the sensitivity of hyperspectral forensic learning to spectral continuity and consistent band ordering as illustrated in Figure 17d.

Overall, the spectral band ablation demonstrates that compact spectral representations remain effective under constrained forensic settings, whereas large-scale forensic learning strongly benefits from preserving the complete hyperspectral cube.

6.1.3. Axis C: Embedding Dimension Analysis

This ablation evaluates the influence of latent embedding dimensionality on hyperspectral forensic representation learning using embedding sizes of 64, 128, 256, and 512.

UWA WIHSI Results:

The compact 64-dimensional embedding configuration achieves the highest episodic accuracy of 0.9367 while maintaining high writer identification performance, indicating that compact latent representations remain sufficient for the constrained 7-writer setting. The improved episodic stability under lower-dimensional embeddings suggests stronger implicit regularization during few-shot optimization, whereas larger embedding spaces do not provide measurable benefits under the constrained UWA setup as shown in Figure 17e.

iVision HHID Results:

Embedding dimensionality exhibits a mild trade-off on the iVision HHID dataset. Embedding dimensions of 128 and 512 achieve the highest Writer Identification Accuracy of 0.9259, slightly outperforming the baseline 256-dimensional configuration with an accuracy of 0.9074. In contrast, the compact 64-dimensional configuration reduces writer identification accuracy to 0.8889, indicating that limited latent capacity becomes insufficient under increased writer diversity and spectral variability as illustrated in Figure 17f.

Overall, the embedding dimension ablation demonstrates that compact embeddings benefit constrained few-shot forensic settings, whereas moderate-to-high-dimensional representations become increasingly important as forensic complexity scales.

6.1.4. Axis D: Initialization Strategy Analysis

This ablation evaluates the influence of weight initialization and transfer learning strategy using iVision transfer learning, ImageNet pretraining, random initialization, and frozen-backbone transfer.

UWA WIHSI Results:

Both ImageNet-only pretraining and random initialization achieved high writer identification accuracy, indicating that the constrained 7-writer setting remains highly separable even without domain-specific spectral pretraining. ImageNet-only initialization achieves stable forensic discrimination performance with IM AUC reaching 0.9932 and EER decreasing to 0.0192, demonstrating effective transfer of broad visual priors under the limited 33-band hyperspectral setting as shown in Figure 18a. The frozen-backbone configuration reproduces the baseline anomaly almost identically, confirming that the degraded baseline behavior originates from constrained episodic sampling rather than deficiencies in pretrained weights.

iVision HHID Results:

The iVision-transfer baseline achieves the strongest overall performance, while the frozen-backbone configuration produces nearly identical results, indicating strong alignment between transferred spectral representations and the target hyperspectral distribution. In contrast, random initialization causes catastrophic degradation in writer identification accuracy, reducing performance to 0.2963 and confirming that pretraining is essential for stable optimization of the 149-band spectral encoder. Although ImageNet-only pretraining preserves relatively strong performance with an accuracy of 0.8889, it remains below

the self-transfer baseline, demonstrating the importance of domain-specific hyperspectral initialization as illustrated in Figure 18b.

Overall, the initialization ablation demonstrates that pretrained spectral priors become increasingly important as hyperspectral dimensionality and writer diversity scale.

6.1.5. Axis E: Training Loss Analysis

This ablation evaluates the influence of metric learning objectives using Prototypical Networks with squared Euclidean distance, cosine-prototype learning, supervised contrastive loss, and batch-hard triplet loss.

UWA WIHSI Results:

All non-baseline loss configurations achieve strong writer identification accuracy, confirming that the constrained hyperspectral setting remains highly separable irrespective of the selected metric learning objective. Among the evaluated losses, supervised contrastive learning achieves the strongest ink mismatch detection performance with IM AUC reaching 0.9990 and EER decreasing to 0.0081, slightly outperforming cosine-based prototype learning and triplet loss. These results suggest that contrastive objectives promote tighter within-class clustering and improved ink discrimination stability as shown in Figure 18c.

iVision HHID Results:

Supervised contrastive learning achieves the strongest Writer Identification Accuracy of 0.9259, marginally outperforming both the Proto sq-L2 baseline and batch-hard triplet learning. Contrastive objectives also produce improved episodic learning behavior relative to the baseline configuration, indicating stronger writer-specific embedding compactness during training as illustrated in Figure 18d. However, Ink Mismatch AUC remains highly stable across all evaluated loss functions, ranging from 0.9911 to 0.9958, suggesting that ink mismatch discrimination is comparatively less sensitive to the optimization objective.

Overall, the training loss ablation demonstrates that contrastive metric learning objectives provide improved embedding compactness and episodic stability, particularly under large-scale writer diversity.

6.1.6. Axis F: K-Shot Support Analysis

This ablation evaluates the influence of support-set size on episodic hyperspectral forensic learning using multiple k-shot configurations.

UWA WIHSI Results:

K-shot variation on the UWA WIHSI dataset produces only marginal episodic accuracy changes across the evaluated configurations. Although k = 2 achieves the highest episodic accuracy of 0.4400, all k-shot settings remain substantially below the performance levels observed under alternative ablation axes. Writer Identification Accuracy and Ink Mismatch AUC remain nearly invariant across support-set sizes, indicating that the observed variations are primarily influenced by the constrained episodic sampling artifact rather than genuine support-set scaling behavior as shown in Figure 18e.

iVision HHID Results:

K-shot variation produces negligible changes on the iVision HHID dataset, with k = 1, 2, and 3 yielding nearly identical writer identification and ink mismatch detection performance. Increasing support-set size does not produce measurable gains in episodic optimization or downstream forensic discrimination capability as illustrated in Figure 18f. The k = 5 configuration becomes infeasible under the cross-split evaluation protocol due to insufficient eligible writer samples.

Overall, the k-shot ablation demonstrates that increasing support-set size provides minimal benefit under the proposed episodic hyperspectral forensic framework, while

k = 1 remains the most practical and stable configuration under constrained cross-split evaluation settings.

6.1.7. Axis G: Embedding Normalization Analysis

This ablation evaluates the influence of embedding normalization strategy using L2-normalization, no normalization, BatchNorm1d, and LayerNorm.

UWA WIHSI Results:

LayerNorm achieves the highest episodic accuracy of 0.9167, followed by BatchNorm1d and the unnormalized configuration. All non-baseline normalization strategies achieve high writer identification performance and stable ink mismatch discrimination, with IM AUC values consistently exceeding 0.9731, as shown in Figure 18g. The improved episodic stability under LayerNorm suggests greater robustness to inter-band spectral variability within the constrained UWA hyperspectral setting.

iVision HHID Results:

All normalization strategies achieve identical Writer Identification Accuracy of 0.9074, indicating that downstream forensic classification remains relatively insensitive to embedding normalization. However, measurable differences emerge in episodic learning and verification-oriented calibration metrics. LayerNorm achieves the highest episodic accuracy of 0.9640, whereas L2-normalization produces the strongest verification performance with the highest IM AUC of 0.9958 and the lowest EER of 0.0360 as illustrated in Figure 18h.

Overall, the normalization ablation demonstrates that normalization strategy primarily influences episodic embedding calibration rather than downstream writer identification performance. While LayerNorm improves episodic stability, L2-normalization remains preferable for verification-oriented forensic deployments requiring stronger calibration consistency.

6.2. Cross-Dataset Ablation Synthesis

Several overarching conclusions emerge from the cross-dataset ablation analysis. While the preceding experiments evaluate the contribution of individual architectural and optimization choices in isolation, their collective interpretation provides deeper insight into the factors governing hyperspectral forensic representation learning under both constrained and large-scale learning conditions. Tables 18, 19 and 20 summarize the best-performing configurations and key observations across all evaluated ablation axes. To provide a consolidated overview of these findings, Figure 19 summarizes the global sensitivity of the evaluated design choices across the considered forensic tasks. The figure highlights the relative performance variation induced by each architectural and optimization factor, providing an overall comparison of their influence on the proposed framework.

First, spectral completeness (Axis B) emerges as the most influential factor affecting forensic performance, particularly on the iVision HHID dataset. Reducing the full 149-band hyperspectral cube to compact spectral representations leads to substantial degradation in writer identification capability, highlighting the importance of preserving spectral continuity and inter-band correlations. In contrast, UWA WIHSI exhibits considerably lower sensitivity to spectral reduction, indicating that its constrained writer space can be effectively separated using a smaller subset of spectral information. This behavior suggests that the importance of spectral richness increases with writer diversity and dataset complexity. The trend is further reflected in the global heatmap shown in Figure 20, where spectral band strategy produces the largest overall performance variation across the evaluated forensic metrics.

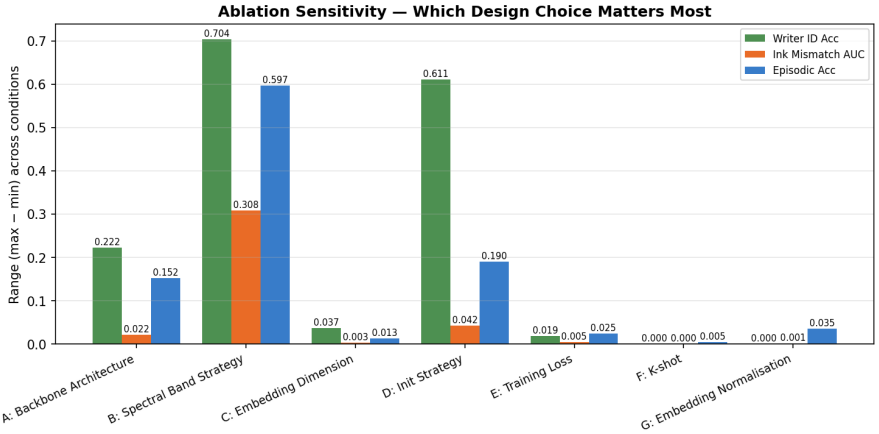


Figure 19. Global ablation sensitivity analysis across all evaluated design axes on the iVision HHID dataset. Bars denote the sensitivity coefficient, computed as the range (maximum–minimum) across experimental conditions for writer identification accuracy, ink mismatch AUC, and episodic accuracy. The results identify spectral band strategy and initialization as the most influential design factors, while the remaining components contribute comparatively smaller performance variations.



Figure 20. Heatmap summarizing the performance of all evaluated ablation configurations on the iVision HHID dataset. Rows correspond to the investigated architectural and optimization settings, while columns represent the evaluated performance metrics. Color intensity indicates normalized performance, providing an overall comparison of the influence of different design choices on the proposed framework. The upward arrow (↑) indicates that higher values represent better performance, whereas the downward arrow (↓) indicates that lower values represent better performance.

Table 18. Best-per-axis summary on the UWA WIHSI dataset.

Axis	Best Cond.	Configuration	WID Acc	IM AUC	IM EER ↓	Note
A	A4	MobileNetV3-Small	1.0000	0.9950	0.0562	Achieves the highest episodic accuracy (0.9067) under constrained few-shot conditions.
B	B2	PCA → 3 bands	1.0000	0.9775	0.0273	Reduced spectral dimensionality remains sufficient for stable forensic discrimination.
C	C1	embed_dim = 64	1.0000	0.9957	0.0192	Compact embeddings provide strong implicit regularization and episodic stability.
D	D2	ImageNet pretrained only	1.0000	0.9932	0.0192	Pretrained visual priors significantly improve optimization stability.
E	E3	Supervised Contrastive	1.0000	0.9990	0.0081	Produces the strongest ink mismatch detection performance across all configurations.
F	F4	k = 5	1.0000 *	0.5710 *	0.4566 *	Results remain partially confounded by constrained episodic sampling instability.
G	G4	LayerNorm	1.0000	0.9959	0.0498	Instance-wise normalization improves episodic consistency under low-resource conditions.

* Indicates configurations affected by the constrained UWA episodic sampling artifact discussed in Section 6. ↓ denotes lower-is-better metrics.

Second, architectural capacity and embedding dimensionality (Axes A and C) exhibit clear dataset-dependent behavior. ResNet-18 combined with moderate-dimensional embeddings provides the strongest balance between optimization stability and forensic discrimination on iVision HHID, whereas lightweight architectures and compact embeddings demonstrate stronger episodic generalization under the constrained UWA WIHSI setting. These findings indicate that representation complexity should be calibrated according to dataset scale and variability rather than increased indiscriminately.

Table 19. Best-per-axis summary on the iVision HHID dataset.

Axis	Best Cond.	Configuration	WID Acc	IM AUC	IM EER ↓	Note
A	A1	ResNet-18 (baseline)	0.9074	0.9958	0.0360	Baseline remains optimal; deeper architectures exhibit overfitting tendencies.
B	B1	Full 149 bands (baseline)	0.9074	0.9958	0.0360	Complete spectral information maximizes forensic identity discrimination.
C	C2	embed_dim = 128	0.9259	0.9962	0.0360	Achieves a +1.85 pp improvement over baseline with a compact latent representation.
D	D1	iVision transfer (baseline)	0.9074	0.9958	0.0360	Transfer initialization substantially stabilizes hyperspectral optimization.
E	E3	Supervised Contrastive	0.9259	0.9943	0.0440	Improves writer identification while preserving robust ink mismatch performance.
F	F1	k = 1 (baseline)	0.9074	0.9958	0.0360	Higher k-shot settings become infeasible under constrained cross-split evaluation.
G	G1	L2-normalization (baseline)	0.9074	0.9958	0.0360	Produces the lowest EER with highly stable embedding geometry.

Note: ↑ denotes higher-is-better metrics, while ↓ denotes lower-is-better metrics.

Table 20. Cross-dataset ablation synthesis between the UWA WIHSI and iVision HHID datasets.

Axis	Design Choice	UWA Effect	iVision Effect	Interpretation
A	ResNet-18 is best	All architectures converge to near-high forensic performance	ResNet-18 achieves the strongest writer identification accuracy	Architectural capacity becomes increasingly important under large-scale writer diversity but is less influential in constrained few-shot settings.
B	Full spectrum is key	PCA-reduced bands also maintain high writer identification accuracy discrimination	Full spectral representation improves writer identification by approximately +9 pp over PCA reduction	Spectral richness becomes increasingly important as writer diversity and hyperspectral variability scale.
C	embed_dim = 128/256	Compact embeddings (64-D) achieve the strongest episodic stability	128-D embeddings improve writer identification by +1.85 pp	Larger embedding spaces benefit scalable forensic learning, whereas compact representations regularize constrained few-shot regimes.
D	Pre-training essential	ImageNet pretraining performs comparably to transfer initialization	Random initialization causes substantial forensic degradation	Both datasets benefit from pretrained initialization, although the domain gap is smaller under constrained 33-band hyperspectral conditions.
E	Loss function selection	Supervised contrastive learning achieves the strongest IM AUC and EER	Supervised contrastive learning improves writer identification by +1.85 pp	Metric-aware contrastive objectives consistently improve hyperspectral forensic embedding separability across both datasets.
F	k = 1 is sufficient	Higher k-shot settings provide only marginal episodic improvement	Larger k-shot settings provide negligible performance gain	Cross-split episodic evaluation constrains the practical benefit of increasing support-set size on both datasets.
G	Embedding normalization	LayerNorm achieves the strongest episodic stability	L2-normalization produces the lowest EER	Normalization strategy significantly influences embedding calibration and task-specific optimization behavior.

Third, optimization-related ablations emphasize the importance of representation learning strategy. Supervised Contrastive learning consistently produces the most stable embedding behavior across both datasets, yielding strong writer identification and ink mismatch discrimination performance. Similarly, pretrained initialization remains essential for reliable optimization of large-scale hyperspectral encoders. Random initialization causes substantial performance degradation on iVision HHID, whereas ImageNet pretraining recovers most of the performance achieved through domain-specific transfer learning.

Finally, k-shot variation and embedding normalization exert comparatively smaller effects on overall forensic performance. Increasing support-set size provides limited benefit within the adopted episodic framework, while normalization primarily influences embedding calibration and episodic stability rather than downstream writer identification accuracy. The relative contribution of each design factor is further summarized in Figure 19, which identifies spectral band strategy and initialization as the dominant contributors to forensic representation learning stability across the evaluated ablation axes.

7. Conclusions

This study investigated episodic hyperspectral representation learning for multiple forensic document analysis tasks under few-shot learning conditions. Rather than introducing a new backbone architecture or metric-learning algorithm, the work evaluated the effectiveness of a shared hyperspectral embedding space across multiple forensic applications, including writer identification, writer verification, ink mismatch detection, forgery detection, gender prediction, and age prediction. By leveraging hyperspectral imaging and episodic prototypical learning, the framework learns transferable spectral–spatial

representations that capture both handwriting characteristics and ink-dependent material information within a unified embedding space.

Comprehensive experiments were conducted on two publicly available hyperspectral forensic datasets, namely UWA WIHSI and iVision HHID, covering both constrained low-resource and larger-scale writer identification scenarios. The experimental results demonstrated strong performance across multiple forensic tasks under controlled and closed-set evaluation settings, particularly for writer identification, writer verification, and ink mismatch detection. While high accuracy was achieved on within-dataset evaluations, the zero-shot and cross-dataset analyses indicated that generalization across unseen writers, acquisition conditions, writer populations, and spectral distributions remains considerably more challenging. These findings highlight both the effectiveness of the learned hyperspectral representations and the continuing challenges associated with transferable forensic inference under realistic deployment conditions.

Extensive ablation studies further validated the architectural and optimization design choices adopted in the framework. The analyses demonstrated that spectral completeness remains a key factor for large-scale hyperspectral forensic learning, while lightweight architectures and compact embeddings can provide improved episodic generalization under constrained few-shot settings. Furthermore, supervised contrastive learning consistently produced stable embedding behavior across both datasets, and pretrained initialization was shown to be important for reliable optimization of high-dimensional hyperspectral encoders.

From a forensic perspective, the results highlight the potential of hyperspectral imaging as a non-destructive and information-rich modality for document examination. The ability to jointly analyze writer individuality and ink composition within a shared representation space offers promising opportunities for questioned document examination, forgery detection, writer verification, and forensic authentication systems operating under limited labeled data conditions.

Despite these encouraging findings, several challenges remain open for future investigation. The current study is limited to controlled hyperspectral acquisition environments and relatively constrained public forensic datasets. Future work will therefore focus on large-scale real-world forensic collections, cross-sensor domain adaptation, transformer-based spectral modeling, explainable hyperspectral forensic analysis, and multimodal fusion strategies integrating spatial, spectral, and textual forensic cues. The development of lightweight deployable architectures and improved cross-domain generalization mechanisms also remains an important direction for practical forensic applications.

Overall, this work demonstrates the potential of episodic hyperspectral representation learning as a shared representation learning framework capable of supporting multiple forensic document analysis tasks. The findings suggest that shared spectral spatial embeddings can effectively support multiple forensic applications while providing a foundation for future research on robust and transferable hyperspectral forensic analysis systems.

8. Limitations and Future Research Directions

Despite the promising forensic discrimination performance achieved by the proposed framework, several limitations and opportunities for future investigation remain.

8.1. Limitations

- The currently available public hyperspectral handwriting datasets remain relatively limited in terms of writer diversity, acquisition variability, and large-scale operational forensic complexity.

- Both UWA WIHSI and iVision HHID were acquired under controlled laboratory imaging conditions with stable illumination, sensor calibration, and document alignment, which may not fully reflect real-world forensic acquisition variability.
- The measured hyperspectral reflectance response is influenced not only by the ink layer but also by the optical properties of the underlying paper substrate. According to the Kubelka–Munk theory, variations in substrate reflectance, paper tonality, and ink transparency can affect the observed spectral signatures [41]. Since the datasets used in this study were acquired on relatively homogeneous document substrates under controlled laboratory conditions, substrate-induced variability was minimized. However, the proposed framework does not explicitly model ink–substrate interactions or perform substrate-aware spectral correction, which may influence performance under more heterogeneous forensic document conditions. Consequently, the robustness of the framework under varying substrate characteristics remains an important direction for future investigation.
- The framework has not yet been evaluated under challenging forensic conditions such as document degradation, aging effects, scanner distortions, compression artifacts, ink diffusion, mixed writing instruments, and partial occlusions.
- Hyperspectral imaging systems remain computationally and operationally more demanding than conventional RGB imaging pipelines, limiting immediate large-scale deployment in resource-constrained forensic environments.
- Although the proposed episodic framework produces highly discriminative forensic embeddings, the interpretability of individual spectral bands and latent feature dimensions remains only partially understood from a forensic evidential perspective.
- Although the proposed framework effectively exploits full-spectrum hyperspectral information, it does not explicitly identify or rank individual discriminative wavelengths. Consequently, wavelength-specific contributions to writer identification and ink mismatch detection cannot be directly quantified within the current study.
- The present study focuses primarily on writer identification and ink mismatch detection, whereas practical questioned document examination frequently involves additional forensic tasks such as forgery localization, signature verification, printer attribution, document dating, and ink aging analysis.
- A limitation of the current framework is the comparatively lower performance observed for forgery detection and age prediction, indicating that a shared hyperspectral spectral–spatial embedding may not fully capture the subtle behavioral and demographic characteristics required for these tasks. Future work should investigate larger and more diverse forensic datasets together with task-specific feature learning strategies to improve robustness for behavioral and demographic handwriting analysis.

8.2. Future Research Directions

- Future work should investigate large-scale hyperspectral forensic datasets containing substantially greater writer diversity and unconstrained acquisition variability.
- Cross-sensor and cross-domain hyperspectral adaptation strategies should be explored to improve robustness under heterogeneous forensic imaging environments.
- Future studies should investigate substrate-aware hyperspectral forensic analysis by incorporating reflectance normalization, Kubelka–Munk-based modeling, and ink–substrate separation techniques to improve robustness across diverse paper materials, document conditions, and acquisition environments.
- The integration of transformer-based spectral modeling, masked hyperspectral pretraining, and self-supervised representation learning presents a promising direction for next-generation forensic intelligence systems.

- Future research should investigate spectral band selection and explainable hyperspectral forensic learning frameworks capable of identifying the most informative spectral regions for writer identification, ink mismatch detection, and related document examination tasks. Such approaches may improve model interpretability while reducing spectral redundancy and computational complexity.
- Expanding the proposed framework toward generalized multi-task forensic learning—including forgery detection, signature verification, printer attribution, and document authentication—remains an important future direction.
- Demographic and behavioral handwriting variability, including age-dependent handwriting evolution, gender-associated writing characteristics, and longitudinal writing changes, should be investigated under hyperspectral acquisition conditions.
- Lightweight spectral architectures, model compression strategies, and edge-deployable inference pipelines should be developed to facilitate practical large-scale forensic deployment.
- Future research may also benefit from multimodal fusion strategies combining spectral, spatial, textual, and temporal forensic cues within unified representation learning frameworks.
- Future studies should investigate task-specific representation learning strategies for challenging forensic applications such as forgery detection and age prediction. Incorporating auxiliary supervision, behavioral handwriting modeling, and explainable hyperspectral feature learning may improve the discrimination of subtle forgery cues and demographic attributes while preserving the generalization capabilities of the unified framework.
- Although evaluation stability was assessed through repeated episodic sampling and confidence interval analysis, future studies should investigate full multi-seed retraining to further quantify optimization robustness and sensitivity to random initialization.
- The investigation of advanced sample construction and pair generation strategies, including hard-negative mining, class-balanced sampling, and adaptive episodic learning, may further enhance representation learning and evaluation robustness across diverse forensic document analysis tasks.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app16146910/s1. This supplementary material provides additional visual analyses complementing the ablation studies presented in the main manuscript. These supplementary figures offer further insight into cross-axis performance sensitivity, baseline-relative behavior, and global forensic embedding trends across the UWA WIHSI and iVision HHID datasets which can be found in Supplementary Materials [6,9,10,20,21,42–51].

Author Contributions: R.N. conceived the study, developed the methodology, implemented the proposed framework, conducted the experiments, performed the data analysis, and prepared the original manuscript draft. K.K. supervised the research, contributed to the conceptualization and design of the study, and provided critical review and revisions throughout the manuscript preparation process. M.H. contributed to the analysis and interpretation of the experimental results, assisted in refining the methodological framework, and participated in the review and improvement of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable. This study did not involve human participants or animals; therefore, institutional review board approval was not required.

Informed Consent Statement: Not applicable. This study did not involve human participants requiring informed consent.

Data Availability Statement: The datasets used in this study are publicly available. The iVision HHID (Hyperspectral Handwritten Images Dataset) is available at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/GSYVLD (accessed on 15 March 2026). The University of Western Australia Writing Ink Hyperspectral Image Database (UWA WIHSI) is available at https://docs.google.com/file/d/0B0M2zmshL10IMGczWlhEOVpYaEk/edit?usp=drive_web (accessed on 20 March 2026).

Acknowledgments: The authors gratefully acknowledge the valuable guidance, constructive discussions, and continuous encouragement received during the course of this research.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bulacu, M.; Schomaker, L. Text-independent writer identification and verification using textural and allographic features. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 701–717. [CrossRef]
2. Christlein, V.; Bernecker, D.; Maier, A.; Angelopoulou, E. Offline writer identification using convolutional neural network activation features. In *Proceedings of the German Conference on Pattern Recognition*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 540–552. [CrossRef]
3. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778. [CrossRef]
4. Kim, S.J.; Deng, F.; Brown, M.S. Visual enhancement of old documents with hyperspectral imaging. *Pattern Recognit.* **2011**, *44*, 1461–1469. [CrossRef]
5. Polak, A.; Kelman, T.; Murray, P.; Marshall, S.; Stothard, D.J.M.; Eastaugh, N.; Eastaugh, F. Hyperspectral Imaging Combined with Data Classification Techniques as an Aid for Artwork Authentication. *J. Cult. Herit.* **2017**, *26*, 1–11. [CrossRef]
6. Khan, Z.; Shafait, F.; Mian, A. Hyperspectral Imaging for Ink Mismatch Detection. In *Proceedings of the 2013 12th International Conference on Document Analysis and Recognition (ICDAR)*; IEEE: Piscataway, NJ, USA, 2013; pp. 877–881. [CrossRef]
7. Jolliffe, I.T.; Cadima, J. Principal Component Analysis: A Review and Recent Developments. *Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci.* **2016**, *374*, 20150202. [CrossRef] [PubMed]
8. Fisher, R.A. The Use of Multiple Measurements in Taxonomic Problems. *Ann. Eugen.* **1936**, *7*, 179–188. [CrossRef]
9. Qasim, M.M.; Mian Qaisar, S.; Khurshid, K.; Nadeem, M. Writer Identification by Processing Hyperspectral Images Using Singular Value Decomposition and Linear Discriminant Analysis. *IEEE Access* **2026**, *14*, 29400–29417. [CrossRef]
10. Snell, J.; Swersky, K.; Zemel, R.S. Prototypical Networks for Few-Shot Learning. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30. [CrossRef]
11. Islam, A.U.; Khan, M.J.; Asad, M.; Khan, H.A.; Khurshid, K. iVision HHID: Handwritten Hyperspectral Images Dataset for Benchmarking Hyperspectral Imaging-Based Document Forensic Analysis. *Data Brief* **2022**, *41*, 107964. [CrossRef] [PubMed]
12. Srihari, S.N.; Cha, S.H.; Arora, H.; Lee, S. Individuality of Handwriting. *J. Forensic Sci.* **2002**, *47*, 856–872. [CrossRef]
13. Schomaker, L.; Bulacu, M. Automatic Writer Identification Using Connected-Component Contours and Edge-Based Features of Uppercase Western Script. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 787–798. [CrossRef] [PubMed]
14. Xing, L.; Qiao, Y. DeepWriter: A Multi-Stream Deep CNN for Text-Independent Writer Identification. In *Proceedings of the 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*; IEEE: Piscataway, NJ, USA, 2016; pp. 584–589. [CrossRef]
15. Zhang, X.Y.; Xie, G.S.; Liu, C.L.; Bengio, Y. End-to-End Online Writer Identification with Recurrent Neural Network. *IEEE Trans.-Hum.-Mach. Syst.* **2017**, *47*, 285–292. [CrossRef]
16. Pallocci, M.; Treglia, M.; Passalacqua, P.; De Luca, L.; Zanovello, C.; Guarna, F.; Gratteri, S.; Marsella, L.T. Forensic Applications of Hyperspectral Imaging Technique: A Narrative Review. *Med.-Leg. J.* **2022**, *90*, 216–220. [CrossRef] [PubMed]
17. Khan, M.J.; Khan, H.S.; Yousaf, A.; Khurshid, K.; Abbas, A. Modern Trends in Hyperspectral Image Analysis: A Review. *IEEE Access* **2018**, *6*, 14118–14129. [CrossRef]
18. de Cássia Mariotti, K.; Ortiz, R.S.; Ferrão, M.F. Hyperspectral Imaging in Forensic Science: An Overview of Major Application Areas. *Sci. Justice* **2023**, *63*, 387–395. [CrossRef] [PubMed]
19. Khan, M.J.; Yousaf, A.; Abbas, A.; Khurshid, K. Deep Learning for Automated Forgery Detection in Hyperspectral Document Images. *J. Electron. Imaging* **2018**, *27*, 053001. [CrossRef]
20. Vinyals, O.; Blundell, C.; Lillicrap, T.; Kavukcuoglu, K.; Wierstra, D. Matching Networks for One Shot Learning. In Proceedings of the Advances in Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; Volume 29, pp. 3630–3638. [CrossRef]

21. Sung, F.; Yang, Y.; Zhang, L.; Xiang, T.; Torr, P.H.S.; Hospedales, T.M. Learning to Compare: Relation Network for Few-Shot Learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 1199–1208. [CrossRef]
22. Zhang, H.; Sun, H.; Gao, H.; Gao, L.; Zhang, B. Hyperspectral Remote Sensing Object Detection via Cross Domain Learning from Visible Images. *IEEE Trans. Geosci. Remote Sens.* **2026**, *64*, 5510518. [CrossRef]
23. Xu, H.; Ran, G.; Cai, Y.; Wang, Y.; Bi, S. LESFuse: A lightweight dual-domain collaborative framework for high-fidelity visible-infrared image fusion. *Appl. Soft Comput.* **2026**, *193*, 114805. [CrossRef]
24. Li, J.; Zheng, K.; Gao, L.; Han, Z.; Li, Z.; Chanussot, J. Enhanced deep image prior for unsupervised hyperspectral image super-resolution. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5504218. [CrossRef]
25. Khan, Z.; Shafait, F.; Mian, A. Automatic Ink Mismatch Detection for Forensic Document Analysis. *Pattern Recognit.* **2015**, *48*, 3615–3626. [CrossRef]
26. Christlein, V.; Gropp, M.; Fiel, S.; Maier, A. Unsupervised Feature Learning for Writer Identification and Writer Retrieval. In *Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*; IEEE: Piscataway, NJ, USA, 2017; Volume 1, pp. 991–997. [CrossRef]
27. Islam, A.U.; Khan, M.J.; Khurshid, K.; Shafait, F. Hyperspectral Image Analysis for Writer Identification Using Deep Learning. In *Proceedings of the 2019 Digital Image Computing: Techniques and Applications (DICTA)*; IEEE: Piscataway, NJ, USA, 2019; pp. 1–7. [CrossRef]
28. Rehman, A.; Naz, S.; Razzak, M.I.; Hameed, I.A. Automatic Visual Features for Writer Identification: A Deep Learning Approach. *IEEE Access* **2019**, *7*, 17149–17157. [CrossRef]
29. Nasir, F.A.; Liaquat, S.; Khurshid, K.; Mahyuddin, N.M. A Hyperspectral Unmixing Approach for Ink Mismatch Detection in Unbalanced Clusters. *J. Inf. Intell.* **2024**, *2*, 177–190. [CrossRef]
30. Tyagi, P.; Agarwal, K.; Jaiswal, G.; Sharma, A.; Rani, R. Forged Document Detection and Writer Identification Through Unsupervised Deep Learning Approach. *Multimed. Tools Appl.* **2024**, *83*, 18459–18478. [CrossRef]
31. Qureshi, R.; Uzair, M.; Khurshid, K.; Yan, H. Hyperspectral Document Image Processing: Applications, Challenges and Future Prospects. *Pattern Recognit.* **2019**, *90*, 12–22. [CrossRef]
32. Chen, Y.; Lin, Z.; Zhao, X.; Wang, G.; Gu, Y. Deep Learning-Based Classification of Hyperspectral Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2094–2107. [CrossRef]
33. Hamida, A.B.; Benoit, A.; Lambert, P.; Amar, C.B. 3-D Deep Learning Approach for Remote Sensing Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4420–4434. [CrossRef]
34. Sauvola, J.; Pietikäinen, M. Adaptive Document Image Binarization. *Pattern Recognit.* **2000**, *33*, 225–236. [CrossRef]
35. Guo, B.; Gunn, S.R.; Damper, R.I.; Nelson, J.D.B. Band Selection for Hyperspectral Image Classification Using Mutual Information. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 522–526. [CrossRef]
36. Gatos, B.; Pratikakis, I.; Perantonis, S.J. Adaptive Degraded Document Image Binarization. *Pattern Recognit.* **2006**, *39*, 317–327. [CrossRef]
37. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How Transferable Are Features in Deep Neural Networks? In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; Volume 27, pp. 3320–3328. [CrossRef]
38. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
39. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. In Proceedings of the International Conference on Learning Representations (ICLR), New Orleans, LA, USA, 6–9 May 2019. [CrossRef]
40. Loshchilov, I.; Hutter, F. SGDR: Stochastic Gradient Descent with Warm Restarts. In Proceedings of the International Conference on Learning Representations (ICLR), Toulon, France, 24–26 April 2017. [CrossRef]
41. Karagiannis, G.; Alexiadis, D. Kubelka-Munk Model—Fundamentals in Radiative Transfer Theory and Models. In *Non-Destructive Methodologies and Adapted Signal Processing Techniques in the Field of Cultural Heritage*; Springer: Berlin/Heidelberg, Germany, 2025; pp. 35–59.
42. Christlein, V.; Bernecker, D.; Hönig, F.; Maier, A.; Angelopoulou, E. Writer identification using GMM supervectors and exemplar-SVMs. *Pattern Recognit.* **2017**, *63*, 258–267. [CrossRef]
43. Fiel, S.; Sablatnig, R. Writer identification and retrieval using a convolutional neural network. In Proceedings of the International Conference on Computer Analysis of Images and Patterns (CAIP), Valletta, Malta, 2–4 September 2015; pp. 26–37. [CrossRef]
44. Hafemann, L.G.; Sabourin, R.; Oliveira, L.S. Learning features for offline handwritten signature verification using deep convolutional neural networks. *Pattern Recognit.* **2017**, *70*, 163–176. [CrossRef]
45. Dey, S.; Dutta, A.; Toledo, J.I.; Ghosh, S.K.; Lladós, J.; Pal, U. Convolutional siamese network for writer independent offline signature verification. *arXiv* **2017**, arXiv:1707.02131. [CrossRef]
46. Bertolini, D.; Oliveira, L.S.; Justino, E.; Sabourin, R. Reducing forgeries in writer-independent off-line signature verification through ensemble of classifiers. *Pattern Recognit.* **2010**, *43*, 387–396. [CrossRef]

47. Impedovo, D.; Pirlo, G. Automatic signature verification: The state of the art. *IEEE Trans. Syst. Man, Cybern. Part C Applications Rev.* **2008**, *38*, 609–635. [CrossRef]
48. Poddar, J.; Parikh, V.; Bharti, S.K. Offline signature recognition and forgery detection using deep learning. *Procedia Comput. Sci.* **2020**, *170*, 610–617. [CrossRef]
49. Topaloglu, M.; Ekmekci, S. Gender detection and identifying one's handwriting with handwriting analysis. *Expert Syst. Appl.* **2017**, *79*, 236–243. [CrossRef]
50. Siddiqi, I.; Vincent, N. Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features. *Pattern Recognit.* **2010**, *43*, 3853–3865. [CrossRef]
51. Siddiqi, I.; Djeddi, C.; Raza, A.; Souici-Meslati, L. Automatic analysis of handwriting for gender classification. *Pattern Anal. Appl.* **2015**, *18*, 887–899. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.