Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 335 (2025) 125916

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.journals.elsevier.com/spectrochimica-acta-part-a-molecular-and-biomolecular-spectroscopy

Ink classification in historical documents using hyperspectral imaging and machine learning methods

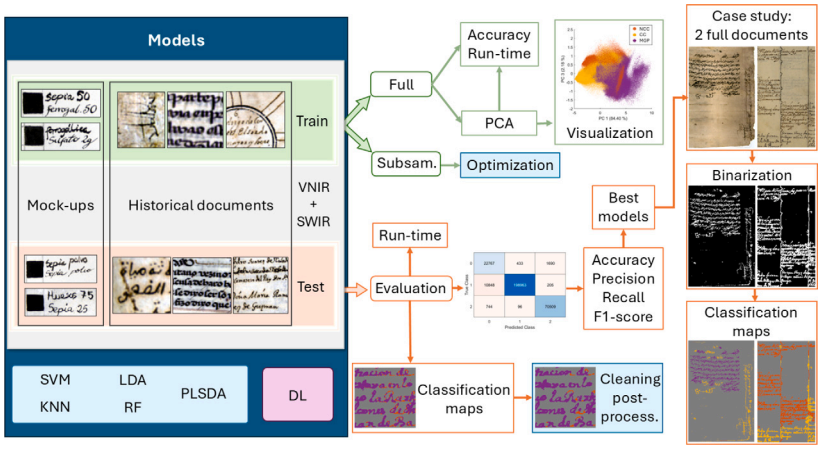
Ana Belén López-Baldomero [a,*], Marco Buzzelli [b], Francisco Moronta-Montero [a], Miguel Ángel Martínez-Domingo [a], Eva María Valero [a]

[a] Department of Optics, University of Granada, Faculty of Sciences, Campus Fuentenueva, s/n, Granada, 18071, Spain
[b] Department of Informatics, Systems and Communication, University of Milano-Bicocca, Viale Sarca, 336, Milan, 20126, Italy

HIGHLIGHTS

- Hyperspectral data from visible to near-infrared enable accurate ink classification via machine learning.
- Three types of ink are classified: metallo-gallate, carbon-containing, and non-carbon.
- A workflow for traditional and deep learning algorithms for ink identification and mapping is proposed.
- Good performance is obtained for all models, reaching 98% of macro-averaged F1 for the deep learning algorithm.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Ink classification
Historical documents
Hyperspectral imaging
Material identification
Machine learning approach
Data fusion
Cultural heritage

ABSTRACT

Ink identification using only spectral reflectance information poses significant challenges due to material degradation, aging, and spectral overlap between ink classes. This study explores the use of hyperspectral imaging and machine learning techniques to classify three distinct types of inks: pure metallo-gallate, carbon-containing, and non-carbon-containing inks. Six supervised classification models, including five traditional algorithms (Support Vector Machines, K-Nearest Neighbors, Linear Discriminant Analysis, Random Forest, and Partial Least Squares Discriminant Analysis) and one Deep Learning-based model, were evaluated. The methodology integrates data fusion from different imaging systems, sample extraction, ground truth creation, and a post-processing step to increase uniformity. The evaluation was performed using both mock-up samples and historical documents, achieving micro-averaged accuracy above 90% for all models. The best performance was obtained using the DL-based model (98% F1-score), followed by the Support Vector Machine model. In the case study documents, the overall performance of the traditional model was better. This study highlights the potential of hyperspectral imaging combined with machine learning for non-invasive ink identification and mapping, even under challenging conditions, contributing to the conservation and analysis of historical manuscripts.

* Corresponding author.
E-mail address: anabelenlb@ugr.es (A.B. López-Baldomero).

https://doi.org/10.1016/j.saa.2025.125916
Received 10 October 2024; Received in revised form 11 February 2025; Accepted 15 February 2025
Available online 27 February 2025

1. Introduction

The identification of materials used in tangible cultural heritage is vital for selecting appropriate restoration and preservation strategies [1]. In particular, the analysis of inks in manuscripts and historical documents enhances our understanding of the artistic and historical context [2,3], aids in dating documents [4], determining authorship, detecting falsifications or undocumented restorations, and identifying causes of deterioration [5,6]. This makes ink analysis a key tool for codicologists and historians who explore both the content and material composition of manuscripts to gather this information [7].

Historically, different types of inks have been used across cultures and periods. The oldest preserved documents were written with carbon inks, created by mixing soot with a binder and water. Metallo-gallate or iron gall inks, which result from the reaction of tannins with iron salts, became more prominent during the medieval period, particularly in Europe. Their identification is crucial for selecting optimal conservation strategies, as their composition can cause corrosion of the support material, a phenomenon known as "iron-gall ink burn" [5]. In general, the study of inks can reveal much about the sociocultural and technological shifts in historical document production [8].

The increasing attention to the material composition of manuscripts and ancient artifacts, in general, reflects a broader recognition of their importance. To obtain compositional information while preserving the integrity and value of these objects, non-invasive analytical techniques are predominantly employed. Among these, X-ray Fluorescence (XRF) [9,10], X-ray Diffraction (XRD) [11,12], Fourier Transform Infrared (FTIR) spectroscopy [9,10,13], and Raman spectroscopy [12,14] are widely utilized. In recent years, hyperspectral imaging (HSI) has gained prominence in this field [6,15]. This technique combines spectroscopy and spatial imaging to provide images at different wavelengths, capturing the spectral reflectance at each pixel of the image. It produces a hypercube containing three-dimensional data: two spatial coordinates and a spectrum for each pixel of the image [16]. The primary advantage of HSI over other methods is its ability to provide spatial information, which enables the retrieval of material distribution within a document, crucial for historical studies and the evaluation of conservation status [17]. In addition, its non-contact and rapid data acquisition capabilities make it particularly suitable for the on-site analysis of historical artifacts at locations such as museums, libraries, or heritage institutions. This avoids the need for transporting fragile or valuable items to a laboratory, reducing the risk of damage, preserving the artifacts, and enabling real-time analysis in their original context. However, the spectral data obtained are not directly correlated with the chemical composition of the materials present in the document. This makes additional processing necessary to gain access to this information, generally by recurring to classification algorithms.

Different spectral ranges can be captured using this technique, each providing distinct data from the artworks. While the short-wave infrared (SWIR) range can be used for revealing *pentimenti* and underdrawings [18], as well as identifying binders [19], the ultraviolet (UV) range is employed to study varnishes [20], and the visible and near infrared (VNIR) range is primarily used for the identification of pigments and dyes [21–28].

The high dimensionality of HSI data makes it particularly suitable for integration with Machine Learning (ML) techniques. The increasing use of ML techniques coupled to HSI data is shown in recent studies for different purposes [16,29–31]. In particular, supervised classification algorithms have been proposed for pigment classification, including Support Vector Machines (SVM) [32–35], Partial Least Squares Discriminant Analysis (PLS-DA) [10,36], Decision Trees (DT) [35], and even Deep Learning (DL) techniques when sufficient data is available [25,33,37]. However, most automatic algorithms developed for material classification in artworks have been trained and tested primarily on mock-ups, *i.e.*, controlled samples created by researchers. As a result, these algorithms may not always perform accurately when applied to real works of art, which can exhibit significant variability due to aging, different states of conservation, and different compounds incorporated into the recipe for a given material class.

In the context of document analysis, several studies have focused on the identification of contemporary inks using HSI in the field of forensic analysis [38–42], as well as the analysis of pigments in illuminated manuscripts [17,25,27]. However, to our knowledge, only one study has addressed the classification of historical inks by means of spectral metrics and a reference library [43]. Thus, the classification of historical inks using non-invasive techniques has received limited attention, particularly when relying exclusively on HSI data.

To date, no study has investigated the automatic classification of historical inks by using ML methods and HSI data. Therefore, the objective of this study was to train and validate six state-of-the-art supervised ML models to automatically classify three types of inks: (1) pure metallo-gallate inks (MGP), (2) carbon-containing inks (CC), which include pure carbon-based inks like ivory black or bone black, as well as mixtures of carbon-based and metallo-gallate or sepia inks, and (3) non-carbon-containing inks (NCC), which can be pure sepia or a mixture of MGP and sepia. Throughout this study, the six algorithms will be divided into two groups: five in the group of traditional techniques, including SVM, K-Nearest Neighbors (KNN), Linear Discriminant Analysis (LDA), Random Forest (RF), and PLS-DA, and one in the group of DL techniques. Given that all inks appear brownish or black in the visible spectrum, hyperspectral images in the visible to near-infrared (VNIR, 400 to 1000 nm) and short-wave infrared (SWIR, 900 to 1700 nm) ranges were captured, and low-level fusion was performed in the spectral dimension to enhance classification accuracy. Both mock-ups and historical documents were included in the training and test sets.

In addition to the primary objective, Principal Component Analysis (PCA) was used prior to classification for visualization of the separability of the classes and dimensionality reduction, comparing the classification accuracy and running time with and without PCA. Parameter optimization for the different models was also tackled, and for the traditional algorithms, a novel post-processing step to increase the local consistency of the results was developed. The full workflow for several traditional algorithms made their performance comparable (only slightly worse) to that of the DL model. The decision to use traditional or DL-based algorithms will then be made depending on the available resources for computation at a given site and the availability of training data. The results of our study show that it is possible to perform ink identification and mapping using only spectral information, thus adding to the evidence validating HSI as a key non-invasive technology in the domain of historical document characterization and preservation.

2. Materials and methods

2.1. Mock-up and historical samples

Both mock-up and historical samples were used in this study to train and test different machine learning models.

The mock-up samples were extracted from a set of modern synthetic samples presented in [44]. These included metallo-gallate inks, sepia, and carbon-based inks, along with their mixtures, prepared according to different traditional recipes from the 13th to 17th centuries [45], and using materials of varying provenance to ensure a wide range of variability. All of them were bound with an Arabic gum solution in water. Each ink was used to fill a 1 × 1 cm square, and a few words were written by hand. Two substrates were used: parchment and hand-crafted cotton-linen paper.

The historical documents were obtained from three different collections of documents preserved in the Provincial Historical Archive and the Royal Chancellery Archive of Granada, Spain.

The first set comprises notarial documents from 1488 to 1494 and a religious text of an undetermined date. Previous analyses using optical microscopy, Scanning Electron Microscopy (SEM), and Fourier-transform infrared (FTIR) spectroscopy have identified various types of ink, including pure carbon-based and iron gall inks, on linen paper [46, 47].

The second set is a family tree book from the 16th and 17th centuries, containing both handwritten and stamped samples of two ink types applied on cotton-linen paper: a mixture of iron gall and sepia, and a pure carbon-based ink. The ink types were identified using SEM by the conservators in charge.

Finally, the third set includes handwritten texts in pure iron gall ink on parchment from various lawsuits of nobility dated between 1459 and 1608. The inks and substrate were identified using XRF and optical microscopy, respectively [48,49].

Additional details about the materials present in the mock-up and historical samples are provided in Appendix A.

2.2. Hyperspectral image acquisition

Two hyperspectral line-scan cameras (Resonon Ltd.) coupled to a linear stage were used to capture all documents. The first camera (Pika L) covers the spectral range from 380 to 1080 nm (VNIR range) with a sensor size of 900 pixels per line [50]. The second camera (Pika IR+) covers from 888 to 1732 nm (SWIR range) with a sensor size of 640 pixels per line [51]. The distance from the camera to the samples was 60 cm for the VNIR camera and 51 cm for the SWIR camera. The field of view (FOV) was 17.3 cm and 14.2 cm, respectively, resulting in an estimated spatial resolution of 192 μm and 221 μm per pixel for the VNIR and SWIR ranges.

Due to the low signal-to-noise ratio, the spectra were cropped at the extremes of the spectral range. This resulted in 121 spectral bands in the VNIR range from 400 to 1000 nm, and 161 bands in the SWIR range, from 900 to 1700 nm, with a sampling interval of 5 nm for both ranges. Before each capture, the devices were calibrated to perform dark subtraction and flat field correction with a white reference surface (the 90% reflectance patch from the Sphere Optics Zenith Lite Multistep of size 20 × 20 cm). A set of four stabilized halogen lamps oriented to avoid specular reflection from the documents and placed at 40 cm from the documents was used as the light source for spectral image capture.

2.3. Data pre-processing - registration, sample extraction, ground truth images, and data fusion

2.3.1. Image registration

In this study, spectral data from both the VNIR and SWIR ranges were fused for classifier model input. To achieve this, it is necessary to first pre-process the spectral images (captured as explained in Section 2.2) so that the corresponding pixels in the VNIR and SWIR images align. This alignment was achieved by spatially registering the VNIR image onto the SWIR image. The VNIR image was chosen as the image to transform because it has a higher spatial resolution, which minimizes artifacts in the final registered image. The registration was performed using only one band of the VNIR (700 nm) and one band of the SWIR hypercubes (1000 nm), which were selected using the criteria of sufficient sharpness according to visual perception and being below 1200 nm for the SWIR. The latter condition is set to avoid proximity to the beginning of the high reflectance region of metallo-gallate inks in the SWIR range, which would lead to a lack of key points needed for proper registration.

Feature-based image registration with SURF features [52] within the MATLAB Registration Estimator App (release R2023a, The MathWorks, Inc., Natick, MA, USA) was performed along with either an affine or a projective spatial transform. The registration quality was visually assessed using overlay images and the Structural Similarity Index Measure (SSIM) [53] after trying different features or spatial transforms to ensure that a satisfactory registration was obtained. The final registration transformation was then applied to all spectral bands within the VNIR hypercube. An example of this process is shown in the first row of Fig. 1.

2.3.2. Sample extraction

The registered VNIR and SWIR hypercubes were cropped multiple times to obtain representative areas containing substrate and one or two different types of ink. These cropped areas, which will be referred to from now on as *minicubes*, were extracted using identical spatial coordinates in both spectral ranges. Fig. 1 (d) and (e), show the false-color images of the minicubes in the VNIR (R = 645 nm, G = 565 nm, and B = 440 nm) and SWIR (R = 1600 nm, G = 1200 nm, and B = 1000 nm) ranges.

2.3.3. Ground truth images

For each minicube, a Ground Truth image (GT) was created using a semi-automatic method. This involved four steps: selecting a band with high contrast between the ink and background using the Signal-to-Noise ratio (SNR) metric, extracting the skeleton of the ink using MATLAB R2023a function *bwskel*, based on Lee et al.'s medial surface axis thinning algorithm [54], and adjusting the skeleton width until the intensity of surrounding pixels matched the average of the Canny edge detector borders. This is a variation on the method proposed in [55], in which the skeleton was manually corrected and then forced to grow until it met the borders. In the fourth step, manual correction is performed after obtaining the automatic GT by visually comparing the result with a false RGB image of the minicube. This procedure will generate a binary image for each pair of registered minicubes. The colors of this binary image are then altered according to the materials identified in the region. The GT image in Fig. 1 (f) has been generated using this procedure and shows the presence of two different inks: metallo-gallate ink mixed with sepia (orange) and carbon-based ink (yellow). The background pixels are middle gray (the three RGB values are set to 128).

2.3.4. Data fusion

After extracting the minicubes and building the GT, the spectral data in both ranges were fused. The fusion process is performed by integrating different data sources to produce more useful and accurate information than any individual data source [56]. In our case, a low-level fusion is performed, where the VNIR and SWIR spectra are concatenated using the bands 400–950 and 955–1700 nm, respectively, without further pre-processing [57]. This resulted in fused spectra with 261 bands. Although PCA could have been applied before data fusion for a mid-level approach, this would carry the risk of discarding potentially important information during dimensionality reduction [57], which could be critical for distinguishing ink classes with overlapping spectral properties. Fig. 1 (g) shows the data fusion result for a single pixel of metallo-gallate mixed with sepia ink (NCC group), illustrating the contribution of each range.

2.4. Training and testing sets

The training and test sets were extracted from a full dataset comprising 44 registered pairs of SWIR and VNIR documents described in Section 2.1. From these, 145 minicube pairs are extracted following the procedure described in Section 2.3.2.

For our experiments, the documents were partitioned into two sets, so that the corresponding minicubes cover 75% of the total for training and 25% for test. Partitioning at document level allows different minicubes extracted from the same document to fall into the same subset. Such precaution prevents the introduction of bias in the training-test split, avoiding test minicubes from having training counterparts coming from the same document. In fact, this would create

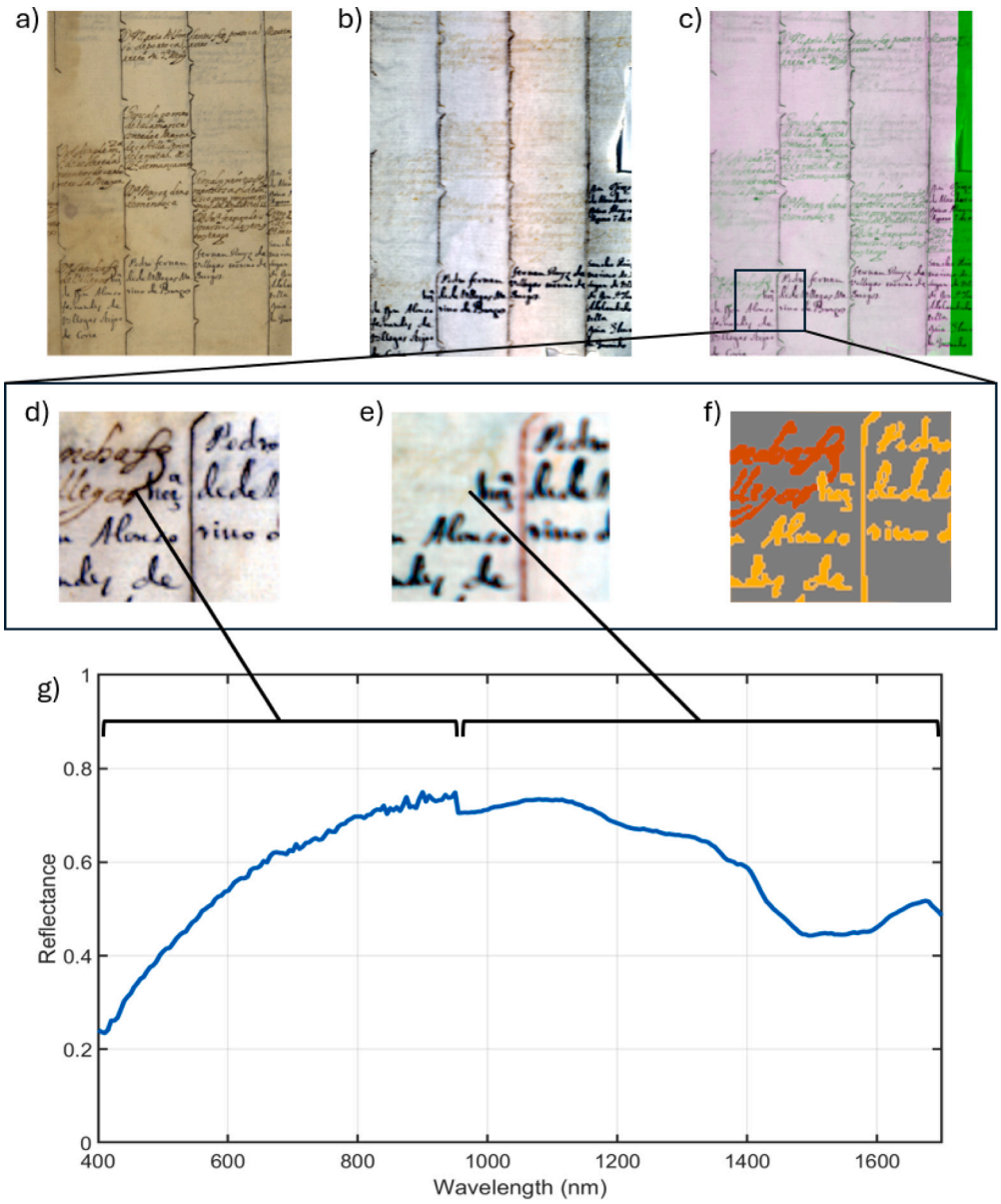


Fig. 1. Registration process and outcome. (a) False color image of the VNIR hypercube (R = 645 nm, G = 565 nm, and B = 440 nm). (b) False color image of the SWIR hypercube (R = 1600 nm, G = 1200 nm, and B = 1000 nm). (c) Overlay of both images. Green represents the areas belonging only to the SWIR capture. (d) False color image of registered VNIR minicube. (e) False color image of registered SWIR minicube. (f) Semi-automatic classification Ground Truth. (g) Full spectrum of an ink pixel after data fusion.

an unrealistic evaluation scenario, where test performance does not reflect real-world performance.

For the training and test sets, only ink spectra (according to the GT images) were selected. This includes pixels from square regions with higher ink deposition in mock-ups, as well as traces with variable amounts of ink, allowing the models to account for these variations during classification, ensuring robustness in the results. This does not exclude the possibility that a particularly dark substrate pixel may have been included in the training data incorrectly, but this will not happen often because the GTs were carefully revised.

Following this procedure, we obtained the data distribution shown in Table 1. Due to computational workload limitations, it was necessary to reduce the number of pixels in the training set for parameter optimization of traditional algorithms, as will be explained in Section 2.7. After the subsampling step, the number of pixels per class was reduced without altering the class imbalance, as shown in Table 1.

Hence, the three classes are not balanced, with 44% of the spectra (pixels) in the CC class, 31% in the MGP class, and 25% in the NCC class. It is worth noting that the sum of the samples containing each class exceeds the total number of samples in the table, as some samples include two types of ink instead of just one. For traditional algorithms, class imbalance compensation by reducing the majority classes did not improve the classification results in preliminary experiments.

2.5. Principal component analysis (PCA)

In this study, PCA is employed for both dimensionality reduction and visualization purposes. First, it is used to reduce the amount of data introduced to the classifiers to improve efficiency, comparing the

Table 1
Training and test data distribution for the three ink classes: pure metallo-gallate inks (MGP), carbon-containing inks (CC), and non-carbon-containing inks (NCC).

Class	Train			Test		Total	
	N. samples	N. pixels	N. pixels subsam.	N. samples	N. pixels	N. samples	N. pixels
MGP	49	232534	46510	14	71349	63	303883
CC	45	330236	66114	17	210016	62	540252
NCC	28	188497	37833	7	24890	35	213387
Total	109	751267	150457	36	306255	145	1057522

accuracy and running time for the training phases of the models with and without PCA, to determine whether PCA is useful for reducing training time without compromising accuracy. To do that, the optimal number of PCs was selected, and their projection coefficients were used to train the classifiers, as explained in Section 2.7.

Additionally, PCA is applied to the hyperspectral data of the full training set for visualization, projecting the first principal components of each spectrum onto a 2D graph. This approach facilitates a quick assessment of the separability of the data, allowing for the visualization of the three separate classes and helping to identify whether distinct clusters can be found, as explained in Section 3.

2.6. Classification models

All models used in this study are supervised classification techniques, so prior information about the data is required. These algorithms automatically identify spectral signatures corresponding to various types of inks, facilitating the classification of unknown samples through the use of a reference or training dataset. As explained before, two groups of algorithms are considered in this study: the traditional and the DL-based. For the implementation of the traditional models, MATLAB software (release R2023a, The MathWorks, Inc., Natick, MA, USA) was used. For the implementation of the DL-based model, Python 3.10.12 was used with the PyTorch deep learning framework at version 1.11.0.

Table 2 provides a summary of the five traditional algorithms (SVM, KNN, LDA, RF, and PLS-DA) and the DL-based model, highlighting their fundamentals, advantages, limitations, and hyperparameters used for each. For details on hyperparameter optimization, please refer to Section 2.7.

PLS-DA was implemented using the PLS_Toolbox (Eigenvector Research, Wenatchee, US). For the DL-based algorithm, neural parameters were initialized through pretraining on a subset of the Microsoft COCO dataset [63]. Although this dataset depicts a different type of visual content, scientific literature suggests that exposing the model to diverse visual data can improve training speed and reduce the amount of required training data [62]. To adapt the model architecture to our problem, we replaced the first convolutional layer, which was originally designed to process 3-channel RGB images, with a new convolutional layer capable of processing 261 channels (111 VNIR + 150 SWIR channels).

2.7. Optimization and post-processing for traditional algorithms

A k-fold cross-validation method with k = 5 was employed to optimize some of the traditional model parameters using a subsampled training set (see Table 1). This technique divides the dataset into k equal-sized folds, where each fold serves as a validation set while the remaining folds are used for training. This process rotates so that each fold is used for validation once, and the model's generalization performance is estimated by averaging the accuracy across all folds.

For KNN optimization, six distance metrics were evaluated: city-block, Chebychev, correlation, cosine, Euclidean, and Minkowski. First, the optimal distance metric was determined using a number of neighbors K = 1. After that, a different number of neighbors were tested with the optimal distance metric: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, and 100. A low K value can lead to overfitting, where the model memorizes the training data too closely and performs poorly on new, unseen data. On the other hand, a high K value can result in underfitting, where the model fails to capture the underlying patterns in the data adequately [59]. The performance metrics data obtained with the subsampled training set in cross-validation (k = 5) can be found in Appendix B. From these results, the cosine distance and K = 1 were selected as the final hyperparameters.

For the SVM model, the box constraint was optimized using the following values: 0.8, 1, 1.2, 1.4, 1.6, 2, 4, 10, 30, 100, 200, and 300. In this case, to select the best value, both micro-accuracy and training time were considered. Increasing the box constraint results in the SVM classifier assigning fewer support vectors, which leads to stricter data separation, but also to longer training times [64]. After evaluating the training time in cross-validation (k = 5) and micro-averaged accuracy (see Appendix C), a box constraint of 10 was selected.

After obtaining the classification maps as explained in Section 2.8, a post-processing cleaning procedure was applied. Considering the assumption that a continuous stroke is composed of the same ink type, each pixel in the classification map was reassigned to the most prevalent class within its surrounding neighborhood. Neighborhood size was defined as 5% of the smallest dimension of the minicube. The cleaning process was repeated over 10 iterations. These parameters were selected based on preliminary tests that indicated optimal performance with minimum computational time. The post-processing step is one of the contributions of this study, and its impact on the performance of traditional models is described in Section 3.

2.8. Performance evaluation

To evaluate the performance on the test set, the confusion matrix was first employed to compute pixel-level performance metrics. The number of True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) were used to calculate precision, recall, accuracy, and F1-score. Precision indicates the reliability of positive predictions (see Eq. (1)), recall assesses the model's ability to identify all the positive instances in the dataset (see Eq. (2)), accuracy measures the proportion of correct predictions out of all predictions (see Eq. (3)), and F1-score is the harmonic mean of precision and recall (see Eq. (4)) [65].

$$precision = \frac{TP}{TP + FP} \tag{1}$$

$$recall = \frac{TP}{TP + FN} \tag{2}$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$F1 - score = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{4}$$

Table 2
Summary of classification models used in the study: fundamentals, advantages, limitations, and hyperparameters.

Model	Fundamentals	Advantages	Limitations	Hyperparameters	Ref.
Support Vector Machines (SVM)	Finds the optimal hyperplane that separates classes with the maximum margin. Handles both linear and nonlinear problems using kernel functions.	– Robust theoretical framework. – Efficient for high-dimensional data. – Strong generalization. – Minimizes the risk of overfitting.	– Requires careful parameter tuning. – Computationally expensive for large datasets. – Originally designed for binary classification.	– Kernel function = Gaussian – Box constraint level = 10 – Kernel scale = Automatic – Multiclass coding = One-vs-One – Standardize data = Yes	[16,31,32]
K-Nearest Neighbors (KNN)	Classifies based on the majority vote of K nearest neighbors; no training phase required.	– Transparent and interpretable. – Well-suited for multi-class problems. – No assumptions about data distribution.	– Sensitive to noise and spectral variability. – High computational cost for large datasets. – Performance highly dependent on K value and distance metric.	– Number of neighbors = 1 – Distance metric = Cosine – Distance weight = Equal – Standardize data = Yes	[31,58,59]
Linear Discriminant Analysis (LDA)	Finds a linear combination of features that best separates two or more classes by maximizing the ratio of between-class variance to within-class variance. Projects the data onto a lower-dimensional space.	– Computationally efficient for large datasets. – Easy to implement and interpret. – Robust to overfitting. – Few hyperparameters. – Well-suited for multi-class problems.	– Assumes linear separability, normal distribution, and equal covariance matrices for all classes. – Performance degrades with high-dimensional or noisy data.	– Discriminant type = Linear – Covariance structure = Full	[30,60]
Random Forest (RF)	An ensemble of decision trees trained on random subsets of data and features. Uses majority voting for classification.	– Robust to noise and overfitting. – Handles high -dimensional data effectively. – No assumptions about data distribution. – Handles collinearity well.	– Sensitive to hyperparameter choices. – Increased complexity with many trees. – Less interpretable compared to simpler models. – High computational cost.	– Ensemble method = Bag – Learner type = Decision tree – Maximum number of splits = 751266 – Number of learners = 30 – Number of predictors to sample = Select All	[31,59]
Partial Least Squares Discriminant Analysis (PLS-DA)	Combines PLS regression and LDA to enhance class separation by creating discriminant functions from input variables that provide better separation than individual variables alone.	– Handles collinearity. – Effective for small sample sizes. – Provides robust class separation.	– Struggles with noisy or highly complex data. – Requires careful preprocessing.	– Preprocessing = Autoscale – Number of latent variables = 3 – Orthogonalize = Off – Algorithm = SIMPLS – Sample weighting = None	[10,29]
Deep Learning (DL)	Uses the DeepLabV3 semantic segmentation model for pixel-wise classification, employing dilated convolutions to efficiently integrate multiscale information.	– Excellent for high-dimensional data. – Models complex relationships and spectral variability. – Benefits from transfer learning.	– High sensitivity to noise without proper training. – Requires large computational resources and datasets to train up to millions parameters. – Increased risk of overfitting if not carefully tuned.	– Epochs = 25 (total of 3300 iterations) – Loss function = categorical cross-entropy – Adam optimizer – Learning rate = 0.0001.	[61,62]

To evaluate the performance for the multi-class problem, two approaches were considered. The Micro-average approach treats all individuals (in our case, the reflectance spectra per pixel) equally, not taking into account differences between the number of instances per class. Micro-average accuracy, precision, recall, and F1-score are exactly the same, so only Micro-average accuracy is computed in this study. The Macro-average approach gives each class equal weight in the average, which ensures that performance is balanced across all classes. Macro-average is computed as the arithmetic mean of the metrics for single classes [65]. The Micro-average approach weights classes according to their frequency in the dataset, which gives more importance to larger classes. Therefore, poor performance on smaller classes is less impactful as they represent a smaller portion of the overall dataset. In contrast, high Macro-average values indicate that the algorithm performs well across all classes, regardless of their frequency. This ensures that each class is considered equally, making it a better measure of performance for imbalanced datasets.

Additionally, classification maps were generated for each minicube based on the prediction results, with each class represented by a distinct color group to facilitate the quick identification of misclassifications. The color code used is: purple for MGP, yellow for CC, and orange for NCC. These maps were then visually inspected to assess the consistency of the classifications.

2.9. Case study: binarization and classification of inks in two full historical documents

After evaluating the different models, a practical application is presented to demonstrate a complete classification process of a historical document using the best-performing models (either traditional or DL-based).

For this purpose, two documents, one from the Royal Chancellery Archive and another from the Provincial Historical Archive of Granada, were selected. The first document is a page from a family tree book dating from between the 16th and 17th centuries. This page is entirely handwritten, and the hands of two different people can be identified in it, each using a particular ink. In this case, the reason for the existence of two authors is unknown. In previous analyses conducted by the conservators in charge, it has been verified the presence of a CC ink and another ink consisting of a mixture of MGP and sepia (NCC). Two minicubes extracted from this document were used as part of the training set samples.

The second document is an Arabic notarial manuscript dated to 1499, detailing a certificate of ownership for irrigated land in the Hotallar village. Such documents are particularly valuable for classification studies, as they typically contain the handwriting of two individuals: a notary who writes the document and a judge who validates it, adding a few words to indicate his agreement, each using different inks. In this manuscript, the text and marginal note are written with a MGP-based ink (MGP), and the judge's validation with a pure carbon-based ink (CC). No minicubes from this document were included in the training set.

For these two documents, an additional pre-processing step was required: the binarization of the document. This step consists of separating the background (substrate) and foreground (inks), so that we can use only the ink pixels as input to the classifiers. The spectral band with the highest contrast was selected using the SNR metric as explained in Section 2.3.3, and then Bradley's Local Image Thresholding algorithm [66] was applied. This method chooses a threshold T for each pixel based on its surroundings:

$$T = \mu \cdot \left(1 - \frac{t}{100}\right) \tag{5}$$

where μ represents the local mean intensity within the chosen window, and t is the percentage of intensity values to be considered as foreground. In our case, we used a window size of $\frac{1}{3}$ of the image height times $\frac{1}{3}$ of the image width, and t is set to 10. We have selected this algorithm and parameters since they obtained the best results in a previous study with similar documents [67].

After classifying the pixels that Bradley's method selected as ink, we performed the same cleaning post-processing as described in previous sections for the traditional algorithms.

To facilitate the evaluation of the classification maps, a GT was manually created using GIMP 2.10.38 software based on the binarized images. It is important to note that these GTs are intended to provide a general overview of the inks in the areas rather than a precise pixel-by-pixel identification as performed for the minicubes, as minor manual errors may be present. Therefore, our focus was on verifying that the number of ink types and their relative spatial positions were consistent with the findings from previous analyses of the documents.

Fig. 2 presents the workflow of the methodology followed in this study, providing a visual representation of the procedures outlined in the preceding subsections. Optimization and cleaning post-processing are shaded in blue as they were only performed for traditional algorithms.

3. Results and discussion

3.1. Average spectral reflectance of inks

In Fig. 3, the average spectral reflectance and standard deviation of the three ink classes from 400 to 1700 nm are presented, along with the average reflectance of two of the substrates, parchment and cotton-linen paper. The ink spectra were extracted from the pixels marked as ink in the GT images of the full training set. A gap can be observed in the 950 to 955 nm range, where data fusion occurred. This is common when different sensors are used for capture and is caused by several factors, including differences in spectral bandwidths, low signal-to-noise ratios, and misalignments in the image setup, which slightly affect the Bidirectional Reflectance Distribution Function (BRDF) [68]. In the visible range, the reflectance patterns of the three inks are similar, showing very low values and a flat shape (with a trend toward reddish color for the MGP ink). However, as the spectrum extends into the near-infrared region, the reflectance of MGP ink diverges, increasing notably as seen in previous studies [69,70]. This divergence is particularly evident starting at approximately 1300 nm, where pure MGP ink becomes nearly transparent (*i.e.* it lets the infrared radiation pass through almost completely, and what one sees in the reflectance curve is the radiation reflected by the substrate). This near transparency distinguishes MGP ink from other inks, including CC ink and sepia, as well as mixtures of inks with or without carbon content. Specifically, CC inks absorb a significant amount of infrared radiation. However, pure sepia and the mixtures of sepia and MGP ink (included in the NCC class) allow slightly more infrared radiation to pass through but do not reach the near-total transparency seen in the MGP class spectra.

In the spectral range from 400 to 1700 nm, the infrared region is particularly interesting as molecular overtone and combination vibrations can be studied from it. Specific absorption bands within this range are associated with distinct chemical bonds: the 1460–1570 nm range corresponds to N-H bond absorptions, the 1100–1400 nm and approximately 1700 nm bands are attributed to C-H bond absorptions, and the 1450 nm band is linked to O-H bond absorptions [71]. Considering the shape of the reflectance curves in Fig. 3, MGP and NCC inks have two peaks in the infrared range: one around 1300–1400 nm, and the other at 1650 nm approximately. In CC inks, the latter peak can also be seen, but much less pronounced. Both peaks could be reasonably assumed to correspond to C-H bond absorptions. However, in this spectral range the absorption bands are weaker and more complex than those in the mid-infrared region, so the application of chemometric techniques is highly suitable to achieve a higher level of confidence in the classification of ink spectra.

3.2. PCA for visualization

In Fig. 4, score plots of principal components (PCs) 1, 2 and 3 are presented. Three PCs were selected by analyzing the Variance Accounted For (VAF) curve and identifying the inflection point at which the curve flattens out. As seen in Fig. 4, 84.4% of the total variance is explained by PC1, 12.1% by PC2, and 2.2% by PC3, achieving a total of 98.6% VAF with just three components. In this Figure, the point cloud for MGP inks, represented in purple, seems to form a separable cluster from that of CC inks, represented in yellow. However, the point cloud for the NCC class is situated between the two previous groups. It should be noted that the CC group contains pure carbon-based inks as well as mixtures with metallo-gallate and sepia inks. Similarly, the NCC group includes pure sepia ink and its mixture with MGP ink. This may be the reason why no clear clustering pattern is observed among the three groups, from which it is concluded that the data are not clearly separable in the PCA components space.

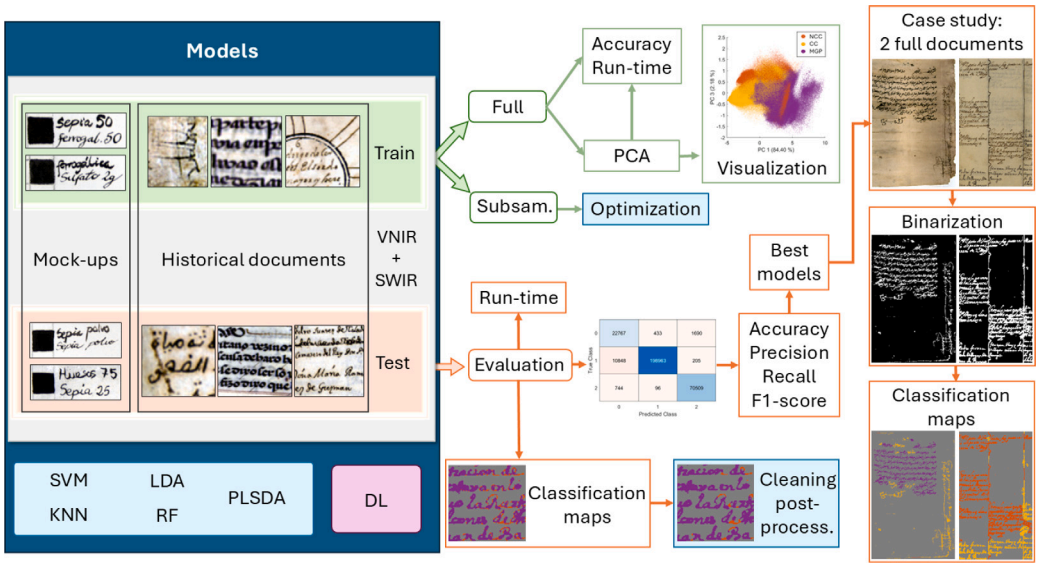


Fig. 2. Workflow outlining the steps followed during the different phases of this study.

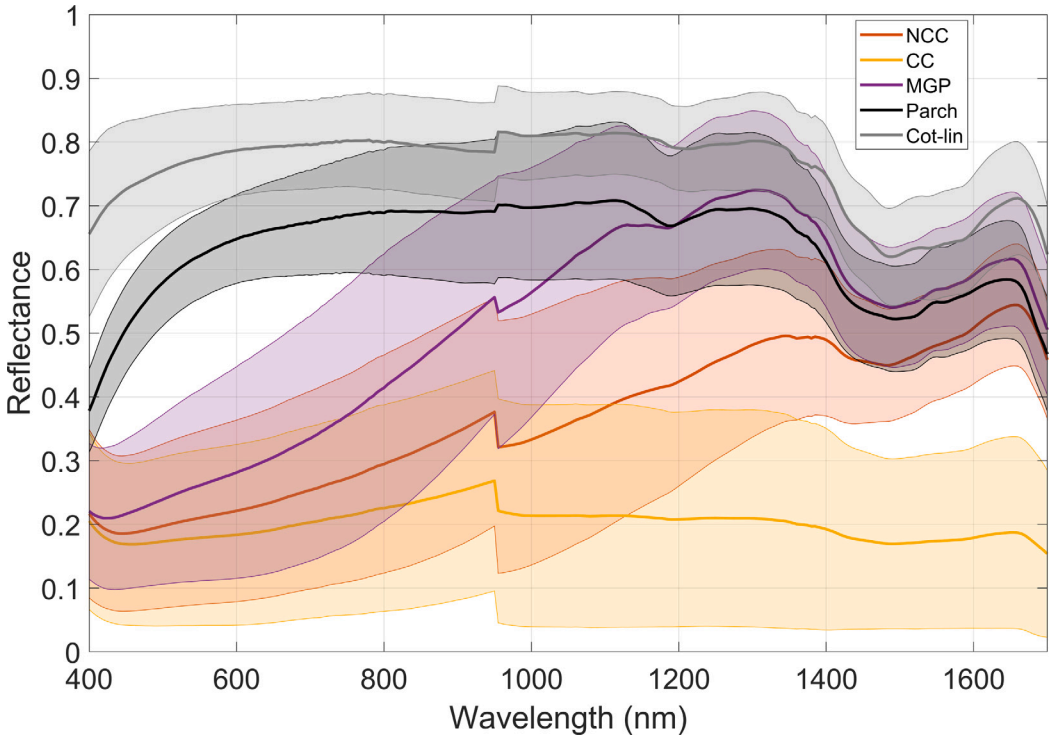


Fig. 3. Average spectral reflectance and standard deviation for the three ink classes in the full training set and two substrates: non-carbon-containing inks (NCC), carbon-containing inks (CC), pure metallo-gallate inks (MGP), parchment (Parch), and cotton-linen paper (Cot-lin).

3.3. Classification maps and performance metrics

In Table 3, the performance metrics for all traditional classification models evaluated on the test set before applying the cleaning post-processing step are presented. Among different models, SVM provided the highest performance across all the metrics studied, including both micro- and macro-averaged results, achieving over 95% in micro and macro-averaged accuracy and recall. This superior performance could be attributed to its efficiency in handling high-dimensional data, its ability to model non-linear relationships, and its robustness to over-fitting (see Table 2). In contrast, PLS-DA is the model providing the lowest values for all the metrics. However, it should be noted that the values of all the metrics are above 72%, reaching almost 87% in micro-averaged accuracy. This means that even the worst of the models

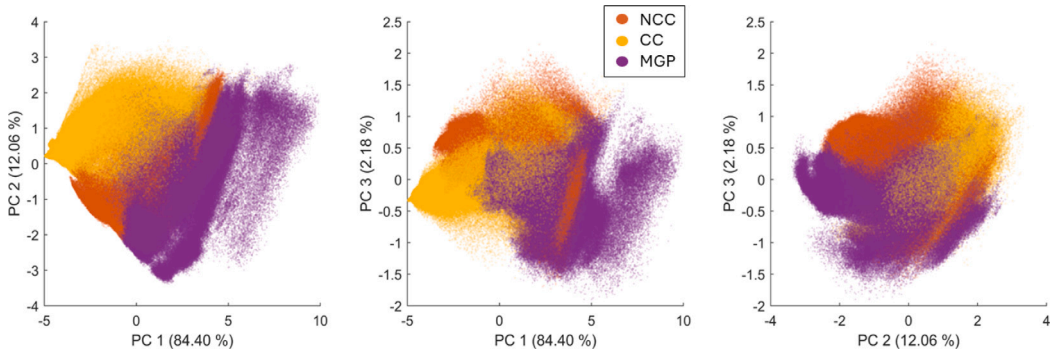


Fig. 4. Score plots of principal components (PC) 1, 2 and 3 for the three different classes used in the study: non-carbon-containing inks (NCC), carbon-containing inks (CC), pure metallo-gallate inks (MGP).

Table 3
Performance metrics in the test set for all traditional models before cleaning post-processing. The color shades represent a gradient from best (dark green) to worst (dark red).

Model	Micro-accuracy	Macro-accuracy	Macro-precision	Macro-recall	Macro-F1
SVM	96.06	95.38	89.17	95.38	91.79
KNN	93.22	93.39	84.09	93.39	87.43
LDA	95.27	93.02	87.90	93.02	89.73
RF	94.77	94.33	86.72	94.33	89.70
PLSDA	86.82	78.33	72.42	78.33	74.36

Table 4
Performance metrics in the test set for all models after cleaning post-processing. The color shades represent a gradient from best (dark green) to worst (dark red).

Model	Micro-accuracy	Macro-accuracy	Macro-precision	Macro-recall	Macro-F1
SVM	98.35	97.29	95.08	97.29	96.15
KNN	95.23	95.79	87.36	95.79	90.63
LDA	97.42	95.44	92.40	95.44	93.71
RF	97.25	97.13	91.78	97.13	94.14
PLSDA	89.80	81.56	76.09	81.56	78.23
DL	99.20	99.13	97.40	99.13	98.22

tested here provides what could be considered good results in this classification task (in comparison to random class assignment, which would yield only about 33% accuracy). All other models provide micro- and macro-averaged accuracy and recall above 93%.

Table 4 presents the performance metrics for the six models evaluated on the test set, after applying the cleaning post-processing step for the five traditional models. When compared to Table 3, a 2 to 3% improvement in micro and macro-averaged accuracy and recall can be observed after applying post-processing. In addition, between 3 and 5% improvement in macro-precision and F1 is achieved. These results indicate that cleaning post-processing is beneficial for the problem we are addressing in the context of traditional models, since in the same stroke (and, therefore, in contiguous pixels in the hyperspectral image) it is not normal to find different types of inks. The post-processing helps improving the results obtained for all the performance metrics studied.

The DL model is included in the post-processing set of results since, by design, it exploits pixel neighborhood information to inform the final class prediction. It outperforms all traditional models with the post-processing step included, having both micro-accuracy and macro-recall above 99%. However, it requires specialized hardware in order to efficiently complete the training and inference phases.

A comparative analysis of performance by class based on the confusion matrices (see Fig. 5) reveals that RF, SVM, and KNN provide the lowest macro-average performance metrics for the NCC class. These models provide also the highest accuracy and recall for the MGP class, and the highest precision and F1 for the CC class. The higher precision but lower recall for the CC class indicates that CC inks are more likely to be classified as MGP and NCC, than NCC are likely to be classified as CC. This misclassification can be attributed to the presence of mixtures of pure carbon with MGP or pure carbon with sepia (which is included in the NCC class). In contrast, both LDA and PLS-DA demonstrate the highest macro-average metrics for the CC class. PLS-DA, in particular, performs poorly for the NCC class, with a precision of 42.7% and an F1 score of 50%. This may be due to the inherent assumptions of the PLS-DA algorithm, which make it less effective at handling the complexity of this class, as PLS-DA is not well-suited for highly complex datasets (see Table 2 fifth row). In contrast, the metrics for the remaining two classes exceed 85%. This performance degradation may be due to class imbalance, as the reduction occurs in the least-represented class. Another explanation is that the spectrum of the NCC class is right between those of MGP and CC (see Fig. 3), leading to increased misclassification between NCC and these two classes, compared to direct misclassification between CC and MGP.

The DL model performs favorably throughout all classes. The lowest recall is at 98.1% for MGP, which tends to be misclassified as NCC,

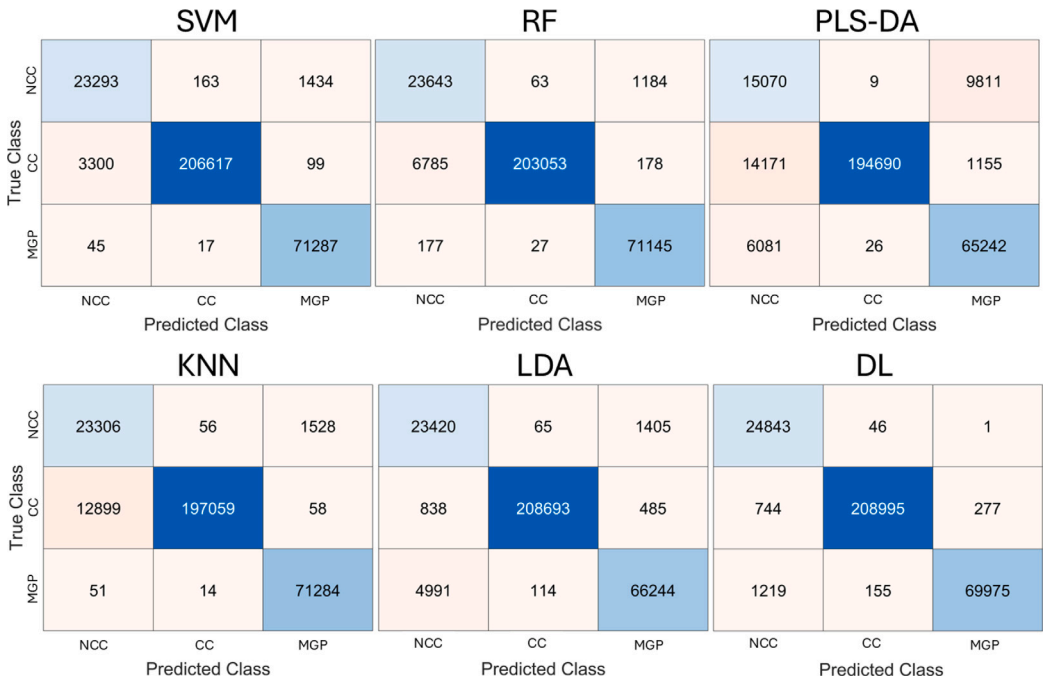


Fig. 5. Confusion matrices of all classification models after cleaning post-processing for the test set. Darker blue in the diagonal cells indicates a higher number of true positive (TP) spectra.

in turn lowering its precision to 92.68%. Of all models tested, the DL achieves the best results for the NCC and CC classes, but not for the MGP class.

In general, for traditional models, NCC pixels tend to be misclassified as MGP and vice versa, while misclassification as CC is less frequent for both classes. However, for all the models when CC pixels are misclassified, they are more likely to be assigned to the NCC class than to MGP. This makes sense, as CC group includes some mixtures of sepia and carbon-based ink, but MGP only includes pure metallo-gallate inks with no mixtures with sepia. In addition, we have seen in Fig. 3 that spectrally, NCC is more similar to CC inks than MGP.

Comparing different performance metrics, the lowest values were always obtained for macro-precision, due to the increased number of false positives for the NCC class. The highest values were obtained for the micro-averaged accuracy metric, which makes sense as MGP and CC are the most represented classes (as seen in Table 1), and provide a high accuracy value.

In a previous study, the classification of historical inks was performed using a library of reference spectra and different spectral metrics for pixel-by-pixel classification. Two spectral ranges, VNIR and SWIR, were studied separately, achieving a maximum F1-score of 58.1% for the VNIR range and 44.3% for the SWIR range using the Spectral Angle Mapper (SAM) metric. However, the ink classes in this study differ from those in the present research, as carbon-based inks from different sources (*e.g.*, vine black, ivory black, bone black, lamp black) were considered separately [43] In another study, hyperspectral analysis combined with Least Squares SVM classification was used for ink analysis and pen verification in handwritten documents. This approach achieved an 87.5% accuracy in discriminating between 25 different pens with modern inks [42] A further study analyzed 70 hyperspectral images of handwritten notes by 7 subjects, comparing 5 varieties of blue ink and 5 varieties of black ink, with a focus on ink mismatch detection [39]. However, these studies are not directly comparable to the present work, as they involve modern inks and are designed for forensic purposes.

In Table 5, training run-time and micro-averaged accuracy for the full training set with and without applying PCA are presented. By applying PCA, the number of features was reduced from 261 to 3, decreasing the training set to 1.15% of the original size. This resulted in a reduction of training run-time to 2%–3% of the original duration for most models, with the exception of PLS-DA, which showed a reduction to 11%, SVM that presented a reduction to less than 50%, and DL which introduced a negligible reduction of training run-time. However, applying PCA results in a 5%–13% decrease in micro-averaged accuracy for traditional models, while the DL model is much more robust to the dimensionality reduction, likely due to the learned ability to extract complex relationships among the input features, and to the possibility of accessing neighborhood data directly during the inference phase. This trade-off between reduced training time and decreased accuracy should be considered when using traditional models for ink classification: if minimizing training time is prioritized over accuracy, then PCA can be applied. On the other hand, the reduction in time is much less significant for the DL model, which might not make the use of dimensionality reduction worthwhile.

The run-time values of traditional models and DL cannot be directly compared because the DL model was run on a GPU, while the traditional models used CPU resources. Besides, the computers used for the two kinds of models were different (see Appendix D for details). However, comparing the training run-time between traditional models is possible: the fastest model to train was PLS-DA, followed by RF. The slowest model was KNN. However, it should also be clarified that run-time in KNN is related to the time required by the program to store the training dataset in the model, since this model does not have a training step as such (see Table 2). In this study, the full training dataset with all 261 bands was selected as the preferred set-up, as it provided the highest accuracy.

Additional insights into the model performances can be gathered from the classification maps. In Fig. 6, some of these maps for selected mock-ups and historical samples are presented for SVM, PLS-DA, and DL models.

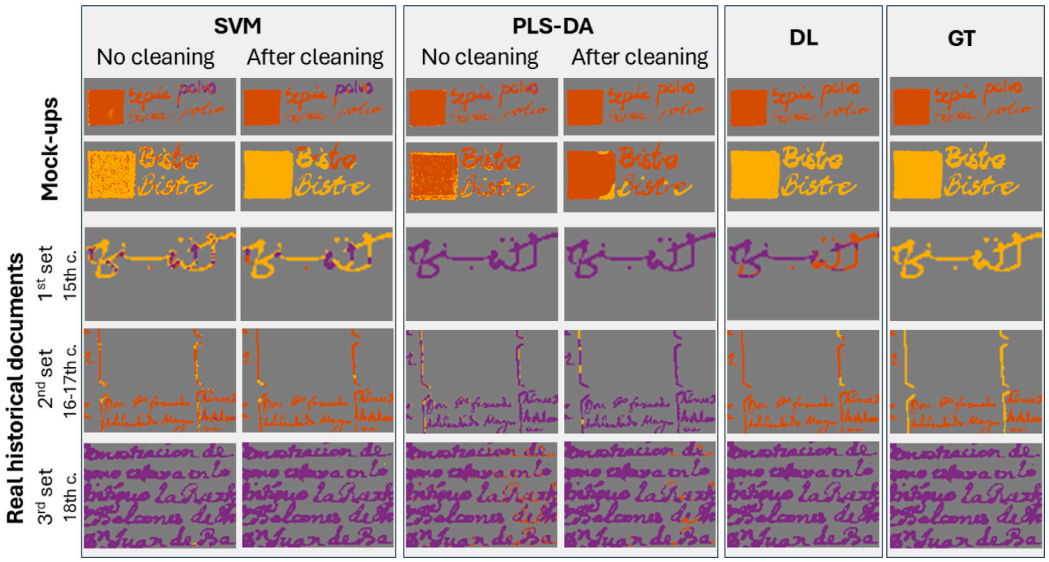


Fig. 6. Examples of classification maps using the SVM model (columns 1 and 2), the PLS-DA model (columns 3 and 4), and DL (column 5). The Ground Truth (GT) images are shown in column 6. Purple: metallo-gallate ink (MGP); yellow: carbon-containing ink (CC); orange: non-carbon-containing ink (NCC).

Table 5
Training run-time and micro-averaged accuracy comparison of different classifiers on the full dataset, with and without PCA.

	Full set (no PCA)		Full set (PCA)	
	Micro-accuracy	Training (seconds)[a]	Micro-accuracy	Training (seconds)[a]
SVM	99.51	27476	92.33	15564
KNN	99.17	90439	85.61	2935
LDA	96.73	24741	86.49	12
RF	99.03	14442	92.68	1538
PLSDA	91.34	137	86.26	15
DL	99.85	3890	99.32	3718

[a] Computational environment used for experiments available in Appendix D.

For the mock-up samples (first and second rows in Fig. 6), some problems persist for the SVM model when differentiating between pure sepia ink and MGP (1st row). For the PLS-DA model, this problem was mostly solved after applying cleaning post-processing (first row, columns three and four). In the case of CC ink (2nd row), SVM misclassifies some pixels in the strokes as NCC, while PLS-DA struggles significantly, incorrectly classifying most pixels as NCC. The DL results (column five) are totally correct for the mock-up samples.

For the historical samples, the example in the 3rd row of the figure was difficult for most models. This sample, composed of CC ink (yellow color coding) on linen paper, is classified by SVM as containing all three ink classes, while PLS-DA incorrectly identifies it as MGP. DL mistakenly identifies the sample as a partial mixture of MGP and NCC: given the better performance observed qualitatively and quantitatively in other samples (see Appendix E for additional examples of classification maps), a possible explanation is that of the model having learned an incorrect bias by relying on stroke structure. Even after the cleaning post-processing is applied for the SVM and PLS-DA models, there are still some or all pixels that are misclassified. However, if the number of pixels classified into the three classes is considered, SVM provides a more accurate classification by correctly identifying the majority of pixels belonging to the CC class. This classification challenge may be attributed to the sample's age, as the 15th-century manuscript exhibits ink fading due to aging, which increases the influence of the substrate on the final ink spectra and raises the reflectance (a wider explanation is given in Section 3.4).

In the second historical sample (4th row), two types of ink can be found: a mixture of MGP and sepia (NCC) in the text, and CC ink in the braces. SVM and DL models have problems with the identification of carbon in the braces, correctly classifying only a few pixels, although correctly performing on the text. For the case of DL in particular, neural architectures for semantic segmentation are known to struggle on isolated thin structures, as typically demonstrated on pole lights in automotive applications [72]: this is due to a combination of learned neighborhood bias (which otherwise helps in correctly identifying large chunks of text) and neural structure limitations (already significantly improved by the DeepLabV3 architecture adopted in this work). PLS-DA misclassifies the entire sample as MGP ink, with only a few carbon pixels correctly identified in the braces.

Finally, the third historical sample (5th row), which is made entirely of iron gall ink, was correctly classified by the SVM and DL models, and nearly correctly classified by PLS-DA.

3.4. Case study: binarization and classification of inks in two full historical documents

In this section, two historical documents with higher complexity and size than the minicubes were tested using the best-performing traditional algorithm (SVM, according to Section 3.3) and the DL model.

The hyperspectral data cube dimensions were [344 × 197 × 261] for the family tree document and [426 × 311 × 261] for the Arabic manuscript. The prediction times for both documents were 3.1 and 3.7 s, respectively, for the SVM model, demonstrating that it can provide near real-time classification, which is highly valuable for restorers, conservators, and researchers interested in the material composition of historical documents.

The DL model took 12.3 and 6.3 s for inference with a sliding window of 35 × 35 pixels, which increased to 30 and 15 s when accounting for data loading and transfer into GPU memory. This computational overhead is significant and should be considered for the implementation of any final application on systems with lower compu-

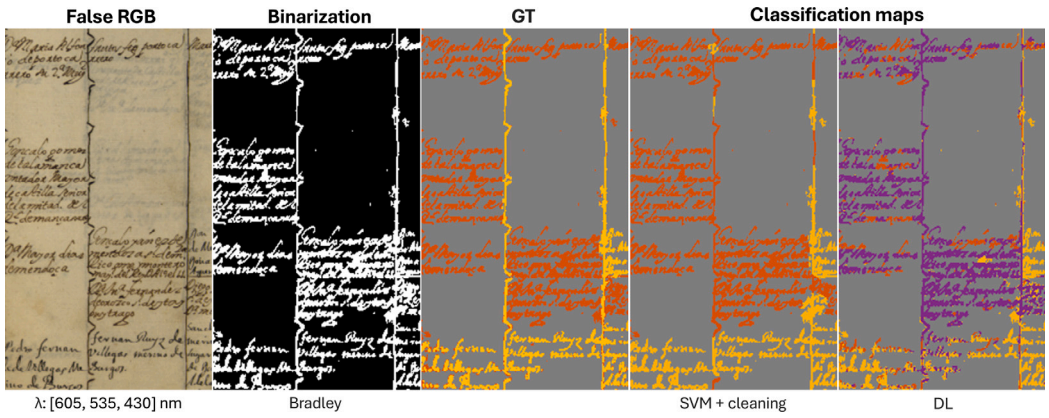


Fig. 7. Family tree document. From left to right: false RGB image, binarization, GT, and classification maps using SVM model after cleaning post-processing and DL model. Predicted MPG pixels are shown in purple, NCC pixels in orange and CC pixels in yellow.

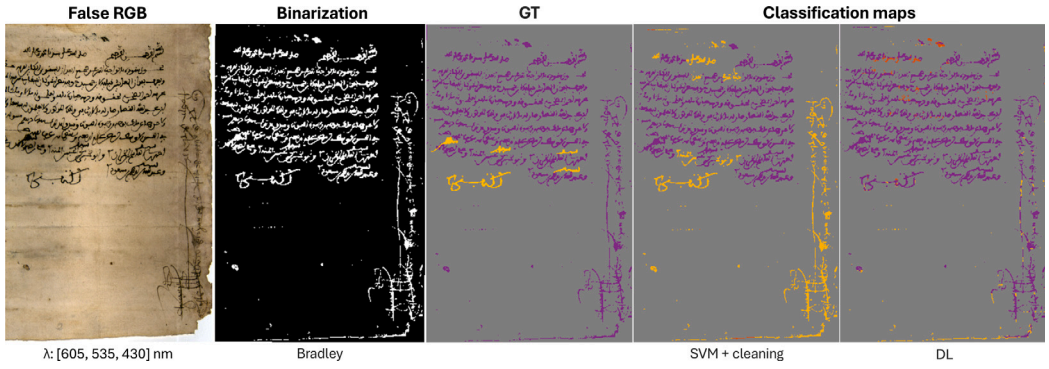


Fig. 8. Arabic notarial manuscript. From left to right: false RGB image, binarization, GT, and classification maps using SVM model after cleaning post-processing and DL model. Predicted MPG pixels are shown in purple, NCC pixels in orange and CC pixels in yellow.

tational capabilities (see Appendix D for details on the computational environment).

In the family tree document (see Fig. 7), the binarization step achieved good visual separation between the ink and support, with only some artifacts present in the lower right portion of the document along the right brace. A false RGB image was generated using the VNIR spectral bands at [605, 535, 430] nm. The SVM model successfully classified most parts of the text as NCC (orange color), while the DL model classified most of the pixels as MGP (purple color), and the remaining pixels as NCC. In addition, all carbon-based text was accurately located, although the DL model has some misclassified pixels as either NCC or MGP in the lower left and central parts. Some additional challenges arise due to the thin traces of the braces for both models, with SVM correctly identifying more pixels as CC ink (yellow color) in this part of the document. Overall, this document is classified more effectively by the traditional SVM-based model.

The analysis of the Arabic notarial manuscript (see Fig. 8) presented more challenges. The binarization results were visually acceptable on the whole, with some bleed-through and stains in the upper and lower part of the document. The manuscript contains two types of ink: the main text and marginal note, both written with metallo-gallate ink with added earth (the correct class would be MGP), and the judge's validations and signature, which are composed of pure carbon ink (CC).

Compared to the family tree document, the classification results were less consistent, as both types of ink (MGP and CC for SVM or MGP and NCC for DL) were found in the same lines of text or words, which does not make sense in a document. However, after applying post-processing techniques for the SVM model, the visual results improved significantly. The judge's signature, located below the main text, is correctly classified as carbon-containing ink (CC, yellow color). This signature was included as well in the test set (see Fig. 6), and it is misclassified as MGP (purple color) by the DL model. The main text is in most pixels correctly identified as MGP by both models. However, classification errors arise in other areas: the judge's validations are incorrectly labeled as MGP for both models, and the marginal note is mistakenly identified as CC for SVM, while correctly classified in most pixels by the DL model.

Further analysis of the document's reflectance in different regions (see Appendix F), reveals potential explanations for the misclassifications. When comparing the spectra, two distinct groups emerge: one containing the main text (MGP) and the judge's validations (CC), and another containing the signature (CC) and the marginal note (MGP). This explains the misclassification of the judge's validations as MGP and the marginal note as CC for the SVM model. Additionally, all inks become transparent in the SWIR range, complicating the classification further. Most of the mock-up CC training samples exhibit low reflectance in the SWIR range, which is crucial for accurately

classifying samples in this class. The behavior of the inks in this document may be attributed to ink degradation, aging, and discoloration, which significantly alter the spectral properties of the ink and complicate classification. Similar spectral changes have been reported in previous studies, particularly in offset inks on paper subjected to artificial aging [73]. In contrast, the family tree document does not present these issues, likely due to better preservation and the fact that it is two centuries younger.

4. Conclusions

In this study, six classification models, including five traditional models (SVM, KNN, LDA, RF, and PLS-DA), and one DL-based model, were implemented for ink classification, and their performance was compared using both mock-ups and historical samples (test set), as well as two full pages extracted from historical documents (case study).

All studied models provided micro-averaged accuracy over 89.8% for the test set. The best results were obtained from the DL model, with micro- and macro-averaged accuracy and recall above the 99% threshold. Nevertheless, among the traditional models, SVM emerged as the best option with all metrics above the 95% threshold and micro- and macro-averaged accuracy and recall above 97%. In both case studies, neither model achieved perfect results. The SVM misclassified fewer pixels and identified key features like the judge's signature in the Arabic notarial manuscript. This document was not included in the training of the model and presented notable challenges for accurate classification due to degradation, aging, and fading of CC inks in the SWIR range.

The choice between a traditional or a DL model can then be based mostly on the available computational resources and how pushing is the need for slightly better accuracy, since the training and hyperparameter tuning of the DL model require a considerable amount of processing resources and the prediction times for higher sized documents are longer. While traditional models could be trained and tested on a personal computer, the same machine could not tackle the training of the DL model. On the other hand, DL does not require a post-processing step that considers the spatial continuity of the classification maps, while traditional models benefit considerably from such post-processing.

The use of supervised classification models with HSI data has proven relevant for the material characterization of documents of historical interest. This can be related to the fact that reflectance imaging can provide indirect information about the molecular structure of the materials employed in the ink recipes, as highlighted in Section 3.1 and mentioned in previous studies [19]. Unlike XRF mapping, reflectance imaging offers the capability to map both inorganic and organic materials or their mixtures. In this respect, it is important to consider data fusion of different spectral ranges as a pre-processing step to highlight distinctive features of the materials like fading in the SWIR range for MGP inks. A key limitation of the proposed approach, compared to other analytical techniques, is the need for a large, annotated training dataset, which requires prior knowledge of the inks used in the documents. However, once the training phase is completed and the performance is evaluated with documents not included in the training set, this methodology eliminates the need for additional techniques to characterize new documents.

Although the identification of written areas in this study is achieved through binarization, this method may prove less effective in cases of poorly preserved texts or high variability, such as interference from complex backgrounds, fading and degradation of ink, stains on the paper, bleeding, paper transparency, or the presence of multi-colored inks. Future research could explore the use of automatic text zone identification schemes (*e.g.*, bounding box-type approaches) or the integration of advanced deep learning architectures designed to handle these complexities and effectively separate text from the substrate.

Classification of inks in the Arabic notarial manuscript has been challenging due to spectral changes, which are likely associated with aging and discoloration. To address this issue, several strategies can be implemented: expanding training datasets with additional historical samples, though this is not always possible due to their fragility and restricted access imposed by conservation policies, and the use of unknown recipes in the materials present; using virtual aging simulations to model spectral shifts resulting from ink degradation; applying accelerated artificial aging to mock-ups in controlled environments (heat, humidity, and radiation) to study spectral changes, although this method may not fully replicate natural aging processes; and using microfading, which, while faster, is less comprehensive than artificial aging, as it only studies the effects of light exposure. These approaches could improve ink classification accuracy in historical materials.

The three classes used for this study provide very useful information for restorers and historians interested in ink characterization of historical documents, since, for instance, MGP tends to show corrosion at the border of the trace, while CC will be more prone to fading. Nevertheless, future work will be focused on tackling a more detailed classification in which the subclasses present in the CC and NCC groups can be separated. One potential approach to address this, given that some inks are mixtures of different components, is the application of unmixing techniques. These methods can provide a more interpretable analysis of individual components and their concentrations in mixtures compared to deep learning (DL) or machine learning (ML) approaches. However, their effectiveness depends on the choice of mixing model, the accuracy of the extracted endmembers (spectra of pure components), and the availability of a comprehensive reference library. Incorporating substrate separation into the traditional model workflow will also be considered.

CRediT authorship contribution statement

Ana Belén López-Baldomero: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Marco Buzzelli:** Conceptualization, Methodology, Software, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Francisco Moronta-Montero:** Conceptualization, Software, Writing – original draft, Writing – review & editing. **Miguel Ángel Martínez-Domingo:** Software, Resources, Data curation, Writing – review & editing. **Eva María Valero:** Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by Ministry of Universities (Spain) [grant number FPU2020-05532], by MCIN/AEI/10.13039/501100011033, by "ERDF A way of making Europe" [grant number PID2021-124446NB-100], and by "ESF Investing in your future" [grant number PRE2022-101352].

This work was partially supported by the Italian Ministry for Universities and Research (MUR) (Italy) under the grant "Dipartimenti di Eccellenza 2023-2027" of the Department of Informatics, Systems and Communication of the University of Milano-Bicocca, Italy.

Table A.6
Details of the substrates, ink types, and corresponding labels for the samples used in the study.

Set	Inks		Label	Substrates
Mock-up samples	Metallo-gallate - ferrous sulfate in different proportions		MGP	Parchment Cotton-linen
	Metallo-gallate - ferrous sulfate + copper sulfate		MGP	
	Metallo-gallate - ferrous sulfate + zinc sulfate		MGP	
	Metallo-gallate - ferrous sulfate + pomegranate juice		MGP	
	Metallo-gallate - ferrous sulfate + pomegranate juice + myrtle leaves infusion		MGP	
	Metallo-gallate - ferrous sulfate + earth pigment[a]		MGP	
	Atramentum[a]		MGP	
	Ivory black[a]		CC	
	Bone black[a]		CC	
	Lamp black[a]		CC	
	Grape seed black[a]		CC	
	Cherry black[a]		CC	
	Bistre[a]		CC	
	Metallo-gallate - ferrous sulfate + lamp black[a] in different proportions		CC	
	Metallo-gallate - ferrous sulfate + bone black[a] in different proportions		CC	
	Lamp black[a] + earth pigment[a]		CC	
	Lamp black[a] + sepia in different proportions		CC	
	Bone black[a] + sepia in different proportions		CC	
	Sepia (from the ink sac of the animal) in different proportions		NCC	
	Sepia in powdered form[a]		NCC	
	Metallo-gallate - ferrous sulfate + sepia in different proportions		NCC	
Historical documents	1st set (notarial documents)	Metallo-gallate	MGP	Linen
		Metallo-gallate + earth	MGP	
		Carbon-based ink	CC	
		Carbon-based ink + earth	CC	
	2nd set (family tree book)	Metallo-gallate + sepia	NCC	Cotton-linen
		Carbon-based ink	CC	
	3rd set (lawsuits of nobility)	Metallo-gallate	MGP	Parchment

[a] From Kremer Pigmente GmbH.

Appendix A. Material information in mock-up and historical samples

See Table A.6.

More details about the recipes used in the mock-up samples can be found in [44]. Note that the exact recipes for the historical documents are unknown.

Appendix B. KNN optimization

See Table B.7.

Appendix C. SVM optimization

See Fig. C.9.

Appendix D. Computational environment

All experiments were conducted on a personal computer with the following hardware configuration for the traditional algorithms:

Table B.7
Micro-averaged accuracy of KNN with different distance metrics and numbers of neighbors.

Model	Distance metric	Neighbors	Micro-accuracy
KNN	cityblock	1	0.9817
KNN	chebychev	1	0.9764
KNN	correlation	1	0.9794
KNN	cosine	1	0.9845
KNN	euclidean	1	0.9826
KNN	minkowski	1	0.9826
KNN	cosine	1	0.9841
KNN	cosine	2	0.9817
KNN	cosine	3	0.9832
KNN	cosine	4	0.9825
KNN	cosine	5	0.9823
KNN	cosine	6	0.9816
KNN	cosine	7	0.9813
KNN	cosine	8	0.9808
KNN	cosine	9	0.9802
KNN	cosine	10	0.9798
KNN	cosine	20	0.9759
KNN	cosine	50	0.9685
KNN	cosine	100	0.9613

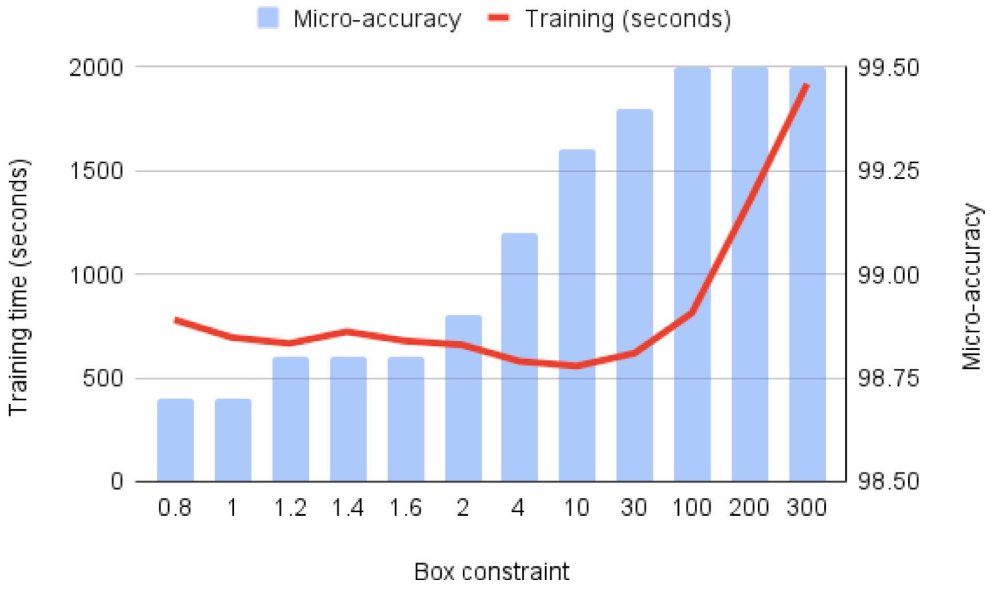


Fig. C.9. Training time in seconds (bar charts) and micro-averaged accuracy (red line) after cross-validation (k = 5) for different values of the box constraint in the SVM model.

- **Processor (CPU)**: Intel(R) Core(TM) i7-8700 CPU @ 3.20 GHz (12 CPUs), 3.19 GHz
- **Memory (RAM)**: 16 GB
- **Storage**: 512 GB NTFS SSD
- **Operating System**: Windows 11 Pro, v. 23H2, 64-bit

The DL-based method required a specialized hardware (GPU), run on a machine with the following configuration:

- **Processor (CPU)**: Intel(R) Core(TM) i7-7700 CPU @ 3.60 GHz (8 CPUs)
- **Memory (RAM)**: 32 GB
- **Graphics Card (GPU)**: NVIDIA Titan X, 12 GB
- **Storage**: 3 TB ext4 SSD
- **Operating System**: Ubuntu 22.04.3 LTS, 64-bit

Appendix E. Supplementary classification maps

See Fig. E.10.

Appendix F. Average spectral reflectance - historical document

See Fig. F.11.

Data availability

Data will be made available on request.

Fig. E.10. Classification maps obtained using all the models studied (SVM, KNN, LDA, RF, PLS-DA, and DL) after cleaning post-processing. The Ground Truth (GT) images are shown in the last row. Purple: metallo-gallate ink (MGP); yellow: carbon-containing ink (CC); orange: non-carbon-containing ink (NCC).

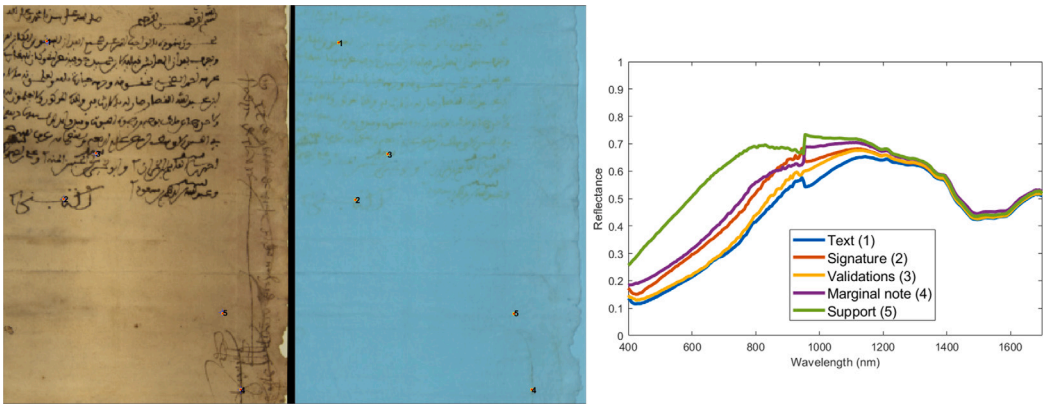


Fig. F.11. False RGB images of the Arabic notarial manuscript in the VNIR ([605, 535, 430] nm) and SWIR ([1300, 1100, 900] nm) spectral ranges. Regions of interest (1–5) were averaged to generate the spectral reflectance plot on the right.

References

[1] J.M. Madariaga, Analytical chemistry in the field of cultural heritage, Anal. Methods 7 (12) (2015) 4848–4876.

[2] S. González-García, A. López-Montes, T. Espejo-Arias, The use of writing inks in 12th–19th century arabic manuscripts: A study of the collection of the School of Arabic Studies, Granada (Spain), in: Science, Technology and Cultural Heritage, CRC Press, 2014, pp. 121–126.

[3] F. Lucarelli, P.A. Mando, Recent applications to the study of ancient inks with the florence external-PIXE facility, Nucl. Instruments Methods Phys. Res. Sect. B: Beam Interactions Mater. Atoms 109 (1996) 644–652.

[4] E.O. Omayio, S. Indu, J. Panda, Historical manuscript dating: traditional and current trends, Multimedia Tools Appl. 81 (22) (2022) 31573–31602.

[5] M.J. Melo, V. Otero, P. Nabais, N. Teixeira, F. Pina, C. Casanova, S. Fragoso, S.O. Sequeira, Iron-gall inks: a review of their degradation mechanisms and conservation treatments, Herit. Sci. 10 (1) (2022) 145.

[6] C. Cucci, A. Casini, Hyperspectral imaging for artworks investigation, in: Data Handling in Science and Technology, vol. 32, Elsevier, 2019, pp. 583–604.

[7] R. Córdoba de la Llave, Interdisciplinary exploration of medieval technical manuscripts from the Iberian Peninsula, J. Mediev. Iber. Stud. 14 (1) (2022) 96–108.

[8] G. Nehring, O. Bonnerot, M. Gerhardt, M. Krutzsch, I. Rabin, Looking for the missing link in the evolution of black inks, Archaeol. Anthr. Sci. 13 (2021) 1–10.

[9] F. Rosi, A. Burnstock, K.J. Van den Berg, C. Miliani, B.G. Brunetti, A. Sgamellotti, A non-invasive XRF study supported by multivariate statistical analysis and reflectance FTIR to assess the composition of modern painting materials, Spectrochim. Acta Part A: Mol. Biomol. Spectrosc. 71 (5) (2009) 1655–1662.

[10] A.L. de Queiroz Baddini, J.L.V. de Paula Santos, R.R. Tavares, L.S. de Paula, H. da Costa Araújo Filho, R.P. Freitas, PLS-DA and data fusion of visible reflectance, XRF and FTIR spectroscopy in the classification of mixed historical pigments, Spectrochim. Acta Part A: Mol. Biomol. Spectrosc. 265 (2022) 120384.

[11] M. Eveno, B. Moignard, J. Castaing, Portable apparatus for in situ X-ray diffraction and fluorescence analyses of artworks, Microsc. Microanal. 17 (5) (2011) 667–673.

[12] L.B. Brostoff, S. Centeno, P. Ropret, P. Bythrow, F. Pottier, Combined X-ray diffraction and Raman identification of synthetic organic pigments in works of art: From powder samples to artists' paints, Anal. Chem. 81 (15) (2009) 6096–6106.

[13] D. Buti, F. Rosi, B.G. Brunetti, C. Miliani, In-situ identification of copper-based green pigments on paintings and manuscripts by reflection FTIR, Anal. Bioanal. Chem. 405 (2013) 2699–2711.

[14] N.C. Scherrer, Z. Stefan, D. Francoise, F. Annette, K. Renate, Synthetic organic pigments of the 20th and 21st century relevant to artist's paints: Raman spectra reference collection, Spectrochim. Acta Part A: Mol. Biomol. Spectrosc. 73 (3) (2009) 505–524.

[15] M. Picollo, C. Cucci, A. Casini, L. Stefani, Hyper-spectral imaging technique in the cultural heritage field: New possible scenarios, Sensors 20 (10) (2020) 2843.

[16] U. Siripatrawan, Y. Makino, Hyperspectral imaging coupled with machine learning for classification of anthracnose infection on mango fruit, Spectrochim. Acta Part A: Mol. Biomol. Spectrosc. 309 (2024) 123825.

[17] E. Catelli, L.L. Randeberg, B.K. Alsberg, K.F. Gebremariam, S. Bracci, An explorative chemometric approach applied to hyperspectral images for the study of illuminated manuscripts, Spectrochim. Acta Part A: Mol. Biomol. Spectrosc. 177 (2017) 69–78.

[18] L. Pronti, M. Romani, G. Verona-Rinati, O. Tarquini, F. Colao, M. Colapietro, A. Pifferi, M. Cestelli-Guidi, M. Marinelli, Post-processing of VIS, NIR, and SWIR multispectral images of paintings. New discovery on the the drunkenness of Noah, painted by Andrea Sacchi, stored at Palazzo Chigi (Ariccia, Rome), Heritage 2 (3) (2019) 2275–2286.

[19] P. Ricciardi, J.K. Delaney, M. Facini, J.G. Zeibel, M. Picollo, S. Lomax, M. Loew, Near infrared reflectance imaging spectroscopy to map paint binders in situ on illuminated manuscripts, Angew. Chem. Int. Ed. 51 (23) (2012) 5607–5610.

[20] M. Kubik, Hyperspectral imaging: a new technique for the non-invasive study of artworks, in: Physical Techniques in the Study of Art, Archaeology and Cultural Heritage, vol. 2, Elsevier, 2007, pp. 199–259.

[21] C. Cucci, J.K. Delaney, M. Picollo, Reflectance hyperspectral imaging for investigation of works of art: old master paintings and illuminated manuscripts, Acc. Chem. Res. 49 (10) (2016) 2070–2079.

[22] C. Balas, G. Epitropou, A. Tsapras, N. Hadjinicolaou, Hyperspectral imaging and spectral classification for pigment identification and mapping in paintings by el greco and his workshop, Multimedia Tools Appl. 77 (2018) 9737–9751.

[23] E.M. Valero, M.A. Martínez-Domingo, A.B. López-Baldomero, A. López-Montes, D. Abad-Muñoz, J.L. Vílchez-Quero, Unmixing and pigment identification using visible and short-wavelength infrared: Reflectance vs logarithm reflectance hyperspaces, J. Cult. Herit. 64 (2023) 290–300.

[24] F. Grillini, J.-B. Thomas, S. George, Comparison of imaging models for spectral unmixing in oil painting, Sensors 21 (7) (2021) 2471.

[25] N. Rohani, E. Pouyet, M. Walton, O. Cossairt, A.K. Katsaggelos, Nonlinear unmixing of hyperspectral datasets for the study of painted works of art, Angew. Chem. 130 (34) (2018) 11076–11080.

[26] M. Vermeulen, K. Smith, K. Eremin, G. Rayner, M. Walton, Application of uniform manifold approximation and projection (UMAP) in spectral imaging of artworks, Spectrochim. Acta Part A: Mol. Biomol. Spectrosc. 252 (2021) 119547.

[27] T. Kleynhans, D.W. Messinger, J.K. Delaney, Towards automatic classification of diffuse reflectance image cubes from paintings collected with hyperspectral cameras, Microchem. J. 157 (2020) 104934.

[28] B. Grabowski, W. Masarczyk, P. Głomb, A. Mendys, Automatic pigment identification from hyperspectral data, J. Cult. Herit. 31 (2018) 1–12.

[29] R.T. Matenda, D. Rip, J.A.F. Pierna, V. Baeten, P.J. Williams, Differentiation of listeria monocytogenes serotypes using near infrared hyperspectral imaging, Spectrochim. Acta Part A: Mol. Biomol. Spectrosc. 320 (2024) 124579.

[30] Y. Shao, S. Ji, Y. Shi, G. Xuan, H. Jia, X. Guan, L. Chen, Growth period determination and color coordinates visual analysis of tomato using hyperspectral imaging technology, Spectrochim. Acta Part A: Mol. Biomol. Spectrosc. (2024) 124538.

[31] J. Yang, L. Sun, W. Xing, G. Feng, H. Bai, J. Wang, Hyperspectral prediction of sugarbeet seed germination based on gauss kernel SVM, Spectrochim. Acta Part A: Mol. Biomol. Spectrosc. 253 (2021) 119585.

[32] G. Capobianco, L. Pronti, E. Gorga, M. Romani, M. Cestelli-Guidi, S. Serranti, G. Bonifazi, Methodological approach for the automatic discrimination of pictorial materials using fused hyperspectral imaging data from the visible to mid-infrared range coupled with machine learning methods, Spectrochim. Acta Part A: Mol. Biomol. Spectrosc. 304 (2024) 123412.

[33] D.J. Mandal, M. Pedersen, S. George, H. Deborah, C. Boust, An experiment-based comparative analysis of pigment classification algorithms using hyperspectral imaging, J. Imaging Sci. Technol. 67 (3) (2023) 030403–1–030403–18.

[34] A. Polak, T. Kelman, P. Murray, S. Marshall, D.J. Stothard, N. Eastaugh, F. Eastaugh, Hyperspectral imaging combined with data classification techniques as an aid for artwork authentication, J. Cult. Herit. 26 (2017) 1–11.

[35] A. Chen, R. Jesus, M. Vilarigues, Identification of pure painting pigment using machine learning algorithms, in: Artificial Intelligence in Music, Sound, Art and Design: 10th International Conference, EvoMUSART 2021, Held As Part of EvoStar 2021, Virtual Event, April 7–9, 2021, Proceedings 10, Springer, 2021, pp. 52–64.

[36] M. Romani, G. Capobianco, L. Pronti, F. Colao, C. Seccaroni, A. Puiu, A. Felici, G. Verona-Rinati, M. Cestelli-Guidi, A. Tognacci, et al., Analytical chemistry approach in cultural heritage: the case of Vincenzo Pasqualoni's wall paintings in S. Nicola in Carcere (Rome), Microchem. J. 156 (2020) 104920.

[37] T. Kleynhans, C.M. Schmidt Patterson, K.A. Dooley, D.W. Messinger, J.K. Delaney, An alternative approach to mapping pigments in paintings with hyperspectral reflectance image cubes using artificial intelligence, Herit. Sci. 8 (2020) 1–16.
[38] Z. Khan, F. Shafait, A. Mian, Hyperspectral imaging for ink mismatch detection, in: 2013 12th International Conference on Document Analysis and Recognition, IEEE, 2013, pp. 877–881.
[39] Z. Khan, F. Shafait, A. Mian, Automatic ink mismatch detection for forensic document analysis, Pattern Recognit. 48 (11) (2015) 3615–3626.
[40] C.S. Silva, M.F. Pimentel, R.S. Honorato, C. Pasquini, J.M. Prats-Montalbán, A. Ferrer, Near infrared hyperspectral imaging for forensic analysis of document forgery, Analyst 139 (20) (2014) 5176–5184.
[41] A.U. Islam, M.J. Khan, M. Asad, H.A. Khan, K. Khurshid, iVision HHID: Handwritten hyperspectral images dataset for benchmarking hyperspectral imaging-based document forensic analysis, Data Brief 41 (2022) 107964.
[42] A. Morales, M.A. Ferrer, M. Diaz-Cabrera, C. Carmona, G.L. Thomas, The use of hyperspectral analysis for ink identification in handwritten documents, in: 2014 International Carnahan Conference on Security Technology, ICCST, IEEE, 2014, pp. 1–5.
[43] A.B. López-Baldomero, M. Martínez-Domingo, E.M. Valero, R. Fernández-Gualda, A. López-Montes, R. Blanc-García, T. Espejo, Selection of optimal spectral metrics for classification of inks in historical documents using hyperspectral imaging data, in: Optics for Arts, Architecture, and Archaeology (O3A) IX, vol. 12620, SPIE, 2023, pp. 99–111.
[44] A.B. López-Baldomero, E. Valero, A. Reichert, F. Moronta-Montero, M. Martínez-Domingo, A. López-Montes, Hyperspectral database of synthetic historical inks, Arch. Conf. 21 (2024) 11–16.
[45] R.J. Díaz, R. Cordoba, H. Grigoryan, M. Vieira, M. Melo, P. Nabais, V. Otero, N. Teixeira, S. Fani, H. Al-Abbady, The making of black inks in an arabic treatise by al-Qalalūsī dated from the 13th c.: reproduction and characterisation of iron-gall ink recipes, Herit. Sci. 11 (2023) 1–14.
[46] T.E. Arias, I.L. Stoytcheva, D.C. García, A.D. Benito, A.J. de Haro, Caracterización material y proceso de conservación de la Colección de documentos árabes manuscritos del Archivo Histórico Provincial de Granada, Al-Qantara 32 (2) (2011) 519–532.
[47] T. Espejo, I. Lazarova, M. Cano, La colección de manuscritos árabes del Archivo Histórico Provincial de Granada. Primeros apuntes sobre su caracterización, in: VIII Congreso Nacional de Historia del Papel en España: Actas, 2008, pp. 33–44.
[48] M.L. de Guevara, et al., Pleitos de Hidalguía. Extracto de sus expedientes que se conservan en el Archivo de la Real Chancillería de Granada correspondiente a la primera parte del reinado de felipe II (1556-1570): en cuatro volúmenes, Hidalgos: La Rev. de la Real Asociacid́e Hidalgos de EspaÑa 1 (565) (2021) 94–95.
[49] A. Duran, A. López-Montes, J. Castaing, T. Espejo, Analysis of a royal 15th century illuminated parchment using a portable XRF–XRD system and micro-invasive techniques, J. Archaeol. Sci. 45 (2014) 52–58.
[50] R. Inc, Resonon PikaL, 2023, https://resonon.com/Pika-L, [Online; Accessed 28 November 2023].
[51] R. Inc, Resonon PikaNIR, 2023, https://resonon.com/Pika-IR, [Online; Accessed 28 November 2023].
[52] H. Bay, T. Tuytelaars, L. Van Gool, Surf: Speeded up robust features, in: Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9, Springer, 2006, pp. 404–417.
[53] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[54] T.-C. Lee, R.L. Kashyap, C.-N. Chu, Building skeleton models via 3-D medial surface axis thinning algorithms, CVGIP, Graph. Models Image Process. 56 (6) (1994) 462–478.
[55] K. Ntirogiannis, B. Gatos, I. Pratikakis, An objective evaluation methodology for document image binarization techniques, in: 2008 the Eighth IAPR International Workshop on Document Analysis Systems, IEEE, 2008, pp. 217–224.
[56] G. Bonifazi, G. Capobianco, P. Cucuzza, S. Serranti, Hyperspectral imaging coupled with data fusion for plastic packaging waste recycling, in: SPIE Future Sensing Technologies 2023, 12327, SPIE, 2023, pp. 104–116.
[57] A. Smolinska, J. Engel, E. Szymanska, L. Buydens, L. Blanchet, General framing of low-, mid-, and high-level data fusion with examples in the life sciences, in: Data Handling in Science and Technology, vol. 31, Elsevier, 2019, pp. 51–79.
[58] J. Chen, C. Fu, T. Pan, Modeling method and miniaturized wavelength strategy for near-infrared spectroscopic discriminant analysis of soy sauce brand identification, Spectrochim. Acta Part A: Mol. Biomol. Spectrosc. 277 (2022) 121291.
[59] R. Ullah, I. Rehan, S. Khan, Utilizing machine learning algorithms for precise discrimination of glycosuria in fluorescence spectroscopic data, Spectrochim. Acta Part A: Mol. Biomol. Spectrosc. (2024) 124582.
[60] P. Xanthopoulos, P.M. Pardalos, T.B. Trafalis, P. Xanthopoulos, P.M. Pardalos, T.B. Trafalis, Linear discriminant analysis, Robust Data Min. (2013) 27–33.
[61] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 801–818.
[62] M. Iman, H.R. Arabnia, K. Rasheed, A review of deep transfer learning and recent advancements, Technologies 11 (2) (2023) 40.
[63] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, Microsoft coco: Common objects in context, in: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13, Springer, 2014, pp. 740–755.
[64] I. The MathWorks, Fitcsvm, 2024, https://es.mathworks.com/help/stats/fitcsvm.html, [Online; Accessed 30 July 2024].
[65] M. Grandini, E. Bagli, G. Visani, Metrics for multi-class classification: an overview, 2020, arXiv preprint arXiv:2008.05756.
[66] D. Bradley, G. Roth, Adaptive thresholding using the integral image, J. Graph. Tools 12 (2007) 13–21.
[67] F. Moronta-Montero, R. Gualda, A.B. López-Baldomero, M. Buzzelli, M. Martínez-Domingo, E. Valero, Evaluation of binarization methods for hyperspectral samples of 16th and 17th century family trees, Arch. Conf. 21 (2024) 94–100.
[68] F. Grillini, L. Aksas, P.-J. Lapray, A. Foulonneau, J.-B. Thomas, S. George, L. Bigué, Relationship between reflectance and degree of polarization in the VNIR-SWIR: A case study on art paintings with polarimetric reflectance imaging spectroscopy, Plos One 19 (5) (2024) e0303018.
[69] V. Corregidor, R. Viegas, L.M. Ferreira, L.C. Alves, Study of iron gall inks, ingredients and paper composition using non-destructive techniques, Heritage 2 (4) (2019) 2691–2703.
[70] M. Faries, Analytical capabilities of infrared reflectography: an art historian's perspective, Sci. Exam. Art: Mod. Tech. Conserv. Anal. (2005) 87–104.
[71] V. Chelladurai, D. Jayas, Near-infrared imaging and spectroscopy, in: Imaging with Electromagnetic Spectrum: Applications in Food and Agriculture, Springer, 2014, pp. 87–127.
[72] D. Mazzini, M. Buzzelli, D.P. Pauy, R. Schettini, A CNN architecture for efficient semantic segmentation of street scenes, in: 2018 IEEE 8th International Conference on Consumer Electronics-Berlin (ICCE-Berlin), IEEE, 2018, pp. 1–6.
[73] B. Havlınová, D. Babiaková, V. Brezová, M. Ďurovič, M. Novotná, F. Belányi, The stability of offset inks on paper upon ageing, Dye. Pigment. 54 (2) (2002) 173–188.