

<https://doi.org/10.1038/s40494-025-02029-7>

Spectral unmixing as a preprocessing step for SVM-based material identification in historical manuscripts



Ana Belén López-Baldero¹✉, Miguel Ángel Martínez-Domingo¹, Sony George²✉ & Eva María Valero¹

When performing material identification from hyperspectral images, a common challenge is the mixing of spectral signals at boundaries between materials. This study investigates spectral unmixing as a preprocessing step to improve machine learning-based classification of inks and writing supports in documents. Hyperspectral data of mock-ups and historical samples were acquired in the VNIR and SWIR ranges, including metallo-gallate, carbon-containing, and non-carbon-containing inks (sepia or mixtures with iron gall) applied to paper and parchment. A subtractive mixing model with automatic endmember extraction was used to generate presence maps and exclude pixels below a concentration threshold. Three support vector machine classifiers were trained using (1) unprocessed reflectance spectra, (2) reconstructed spectra from unmixing, and (3) pure unmixed spectra. Reconstructed spectra provided the best overall performance and classification maps, while unmixed spectra outperformed in ink identification, particularly bleed-through detection. Unmixing also revealed areas of lower classification confidence, offering potential for broader hyperspectral applications.

Understanding the material composition of manuscripts and historical documents remains fundamental for their study, conservation, and authentication¹. Inks and writing support analysis allows researchers to detect alterations, address degradation processes effectively², establish a plausible date, and understand the historical context³. This makes ink and writing support analysis a key tool for codicologists and historians who explore both the content and material composition of manuscripts. Historically, different types of inks have been used across cultures and periods, revealing much about sociocultural and technological shifts in document production⁴.

The use of hyperspectral imaging (HSI) to obtain compositional information has gained prominence in recent years. HSI combines spectroscopy and imaging technologies, yielding spatial information across a wide spectral range. It typically ranges from the ultraviolet to the short-wave infrared (SWIR), in a non-invasive and relatively fast data acquisition. The reflectance spectrum captured by HSI is characteristic of the material under study, enabling both material identification and spatial mapping. Initially, the targeted fields were remote sensing⁵, food industry⁶, and medicine⁷ among others⁸. In addition, HSI has been successfully applied in various cultural heritage contexts, including ink and pigment identification in historical documents^{9–12} and artworks^{13–16}.

Reflectance spectra have great potential for material classification even without any additional pre-processing steps, since they contain much of the necessary information for characterizing the constituent materials. Nevertheless, especially for ink identification, spectral similarity in the visible range, the use of different recipes for the same ink type, and the degradation state of the materials can complicate the process. In particular, ink thickness, ink concentration or stroke width variations across different parts of text introduce further complexity. For instance, at the edges of strokes, in thin traces, or in degraded areas where the ink layer is partially lost or the support has been affected by corrosion, the interaction between ink and writing support is stronger, either because the ink contributes less or because it produces changes in the spectra of the writing support in these areas. This is more prominent in the SWIR region, where radiation can penetrate certain inks, such as metallo-gallate types, making them distinguishable from other ink types and the writing support itself^{9,17}. As a result, the spectral response of the ink may vary depending on the type and condition of the writing support, since modern and aged materials have different reflectance properties¹⁸.

In pixel-wise classification, assigning a single material label to mixed pixels, where both ink and writing support contribute to the spectrum in different ways¹⁹, is often inaccurate. This leads to classification errors and biased performance evaluation. Previous studies have shown that such areas

¹Department of Optics, Faculty of Sciences, University of Granada, Granada, Spain. ²Department of Computer Science, Norwegian University of Science and Technology, Gjøvik, Norway. ✉e-mail: anabelenlb@ugr.es; sony.george@ntnu.no

are frequently misclassified, for example, as pencil marks²⁰ or as different types of ink²¹. These limitations reduce the accuracy of classifiers trained and tested directly on unprocessed spectral data.

The difficulty in detecting boundaries between materials is not specific to cultural heritage applications but is also reported in other domains of HSI. Sifnaïos et al.²² estimated that 97.45% of classification errors occur at the borders of objects and background. In remote sensing, several methods have been proposed to detect edges or boundaries^{23,24}, most of them based on neural networks and deep learning approaches. While these methods can model complex patterns, they often require large annotated datasets and can be affected by overfitting, underfitting, class imbalance, high computational cost, and limited interpretability.

In close-range HSI, the application of linear spectral unmixing (LSU) was proposed as the first step for detecting the boundaries of different objects²⁵. The goal of spectral unmixing is to decompose the measured spectrum into constituent pure component spectra (endmembers) and their proportions (concentrations), with results highly dependent on the selected input endmembers. Unlike typical supervised classification tools such as SAM (Spectral Angle Mapper), LSU generates abundance maps rather than discrete class labels. Spectral unmixing offers advantages over deep learning methods for regression modeling, including physical interpretability, computational efficiency, and ease of implementation^{26,27}. This approach has been successfully applied in the analysis of pigment mixtures in paintings^{15,16,19,28–30}, but its application to documents of historical interest remains underexplored.

This study addresses a limitation in the hyperspectral analysis of the written parts in historical documents: classification errors that occur when ink and writing support materials simultaneously contribute to the reflectance spectrum of a pixel. To mitigate this issue, we developed an automatic preprocessing step based on spectral unmixing, applied prior to classification and model evaluation, to separate the spectral contributions of ink and

writing support during both training and testing. Support Vector Machine (SVM) was selected as the classification algorithm, based on prior evidence identifying it as the most effective traditional machine learning model for historical ink classification⁹. A total of five classes were defined to allow preliminary material identification in most common cases for historical documents: pure metallo-gallate inks (MGP), carbon-containing inks (CC), non-carbon-containing inks (NCC), parchment, and paper. The proposed methodology, with two variants, improves classification accuracy and robustness, and besides it also identifies pixels affected by spectral mixing. This additional information allows researchers to recognize areas where classification confidence may be lower. Although the approach is demonstrated in the context of ink analysis in historical documents, the fundamental challenge of mixed pixel classification is common across HSI applications, suggesting broader applicability of this preprocessing strategy for improving material classification reliability in contexts where the presence of mixed spectra compromises classification approaches.

Methods

Hyperspectral imaging capture and data preparation

The training and test samples used in this study are identical to those presented in López-Balderero et al.⁹, enabling direct comparison with previous results within certain limits imposed by methodological differences between the approach in⁹ and the one described in this study. Mock-up samples were created following traditional historical recipes^{31,32} (see refs. 18,20 for details). They consist in metallo-gallate inks, sepia, and carbon-based inks, along with their mixtures, applied to different writing support materials (parchment and paper manufactured from either cotton, linen, hemp or a mixture of cotton and linen) (see Fig. 1 (a) and Figures 7 and 8 in ref.¹⁸). Historical documents were obtained from two archives, the Provincial Historical Archive and the Archive of the Royal Chancellery, in Granada, Spain. All documents contained different ink types applied to

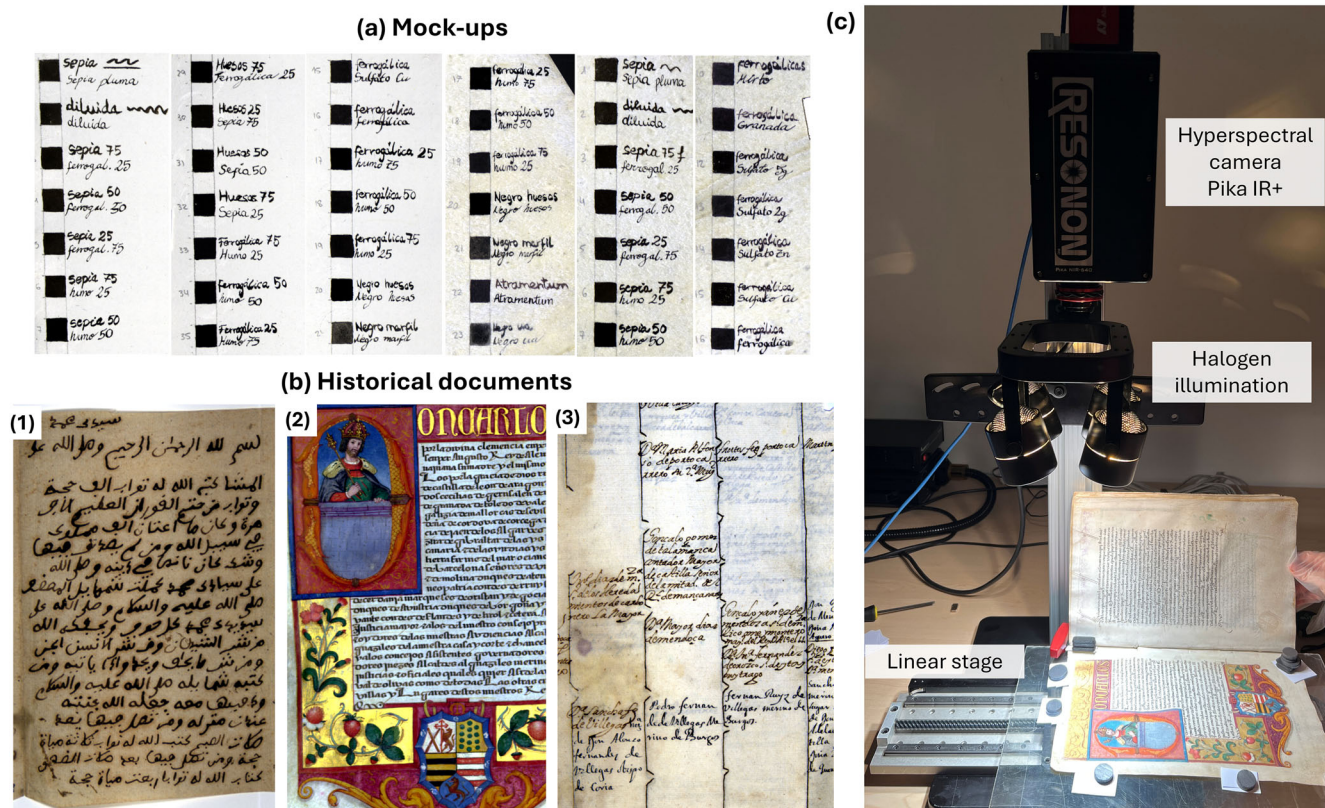


Fig. 1 | Examples of samples from different subsets. **a** Mock-ups of historical inks on cotton paper and parchment; **b** historical documents: (1) manuscript from the Provincial Historical Archive of Granada, (2) illuminated manuscript from the

Archive of the Royal Chancellery of Granada, (3) family tree document from the same archive; **c** hyperspectral image acquisition of an illuminated manuscript using the Pika IR+ (SWIR) camera.

various writing supports, with materials previously characterized using complementary analytical techniques, including Scanning Electron Microscopy (SEM), Fourier Transform Infrared (FTIR) spectroscopy, X-ray fluorescence (XRF), and optical microscopy (OM)^{33–36}. The document sets comprised: (1) 15th-century Arabic manuscripts, (2) lawsuits of nobility from the 15th–17th centuries, and (3) a genealogical tree book from the 16th–17th centuries (see Fig. 1 (b)).

Hyperspectral imaging was performed using two line-scan cameras from Resonon Ltd. (Bozeman, Montana, USA): a Pika L covering the visible to near-infrared range (VNIR: 380–1080 nm) with 300 spectral channels and a sensor size of 900 pixels/line, and a Pika IR+ covering the short-wave infrared range (SWIR: 888–1732 nm) with 368 spectral channels and a sensor size of 640 pixels/line. Considering the working distances of the two cameras, this results in pixel sizes of 0.16 and 0.22 mm for the VNIR and SWIR cameras, respectively. An example of the setup is shown in Fig. 1 (c). The system included a linear translation stage to move the object and four halogen lamps for illumination. Due to low signal-to-noise ratios at the spectral extremes, the acquired spectra were cropped, resulting in 121 spectral bands in the VNIR range (400–1000 nm) and 161 bands in the SWIR range (900–1700 nm), both with 5 nm sampling intervals after linear interpolation for obtaining regularly spaced bands. Reflectance calibration was performed using a 90% reflectance reference tile made by Sphere Optics Zenith Lite.

Spatial registration between VNIR and SWIR full document pages was performed using the MATLAB Registration Estimator App (R2023a, The MathWorks, Inc., Natick, MA, USA) to ensure pixel-wise correspondence across spectral ranges. Each VNIR cube, with higher spatial resolution, was registered onto the SWIR cube to minimize spatial artifacts in the fused data. Hence, the final spatial resolution of the registered cubes was that of the original SWIR cubes. Small regions of interest were manually extracted from registered images, and semi-automatic ground truth annotation was performed to identify areas containing different materials, following the methodology detailed in López-Baldero et al.⁹.

For data fusion, a low-level approach was initially applied by concatenating VNIR (400–950 nm) and SWIR (955–1700 nm) ranges. However, differences in sensor characteristics, spectral bandwidths, signal-to-noise ratios, and slight registration misalignments due to movement of the sample between VNIR and SWIR captures, affecting the Bidirectional Reflectance Distribution Function (BRDF)³⁷ created a spectral discontinuity in the 950–955 nm transition region. To address this issue, a logistic splicing correction method³⁸ was applied to achieve smooth spectral concatenation between the two ranges.

Unmixing preprocessing

As mentioned in the Introduction, spectral unmixing methods involve decomposing a mixed pixel or mixed target spectrum, composed of different pure spectral signals, into its constituent spectra, known as endmembers (EMs), and determining their relative concentrations. Unmixing requires assuming a mixing model $\mathbf{Y}(\lambda) = f(\mathbf{E}(\lambda, q), \mathbf{C}(q))$ on a per-pixel basis³⁹, where $\mathbf{E}(\lambda, q) = [\rho_1(\lambda), \rho_2(\lambda), \dots, \rho_q(\lambda)]$ denotes the spectral library containing spectral reflectances for each of the q EMs, and $\mathbf{C}(q) = [\alpha_1, \alpha_2, \dots, \alpha_q]^T$ represents the concentration vector of individual abundances. In our case, these abundances refer to the mixing weights, that is, the relative optical contributions of each EM within the mixed spectrum of a pixel. The process consists of two main steps: EM extraction and concentration estimation.

The EM extraction methods aim to estimate the spectra of the main raw constituent materials present in a scene. In the context of historical documents, where ink is applied on paper or parchment supports, two distinct EMs are expected: one corresponding to the ink and one to the support material. Note that the ink may consist of a mixture of different ink types; however, in this work, our objective during the unmixing step is not to identify individual ink components, but to separate the spectral contributions of ink and writing support in order to assist the subsequent classification performed by the SVM model. Therefore, the first step of our unmixing preprocessing approach was to automatically extract these two EMs from each sample:

1. Ink EM (EM1): The ink reflectance spectrum was extracted using ground truth (GT) annotations combined with morphological erosion. The erosion was applied to select pixels from the central areas of ink strokes (white region in Fig. 2a.), removing edge pixels where spectral mixing occurs. A square structuring element of 5 pixels was selected for border removal, reduced to 3 pixels for samples with very thin strokes. The spectra of the selected pixels were then averaged to obtain the reflectance spectrum of EM1 (ρ_1), the ink EM. The resulting spectrum, along with its standard deviation, is shown below a. in Fig. 2. This approach ensured that only pixels with minimal writing support contribution were included in the ink EM.
2. Support EM (EM2): A clean area of the writing support was automatically identified by sliding a 10×10 square window over the GT and selecting the area that contained only background pixels and maximized the minimum distance to the nearest ink pixel (red square in Fig. 2b.). When a 10×10 square could not be accommodated due to spatial constraints, a 4×25 pixel rectangle was used instead. The spectra within this area were averaged to form EM2 (ρ_2), the writing support EM. The resulting spectrum, along with its standard deviation, is shown below b. in Fig. 2.

As described above, the extraction of EMs relies on the availability of GT annotations. These can be obtained using different binarization techniques, such as those evaluated in ref.⁴⁰, followed by the manual assignment of class labels to each material, as further explained in⁹. Once the EMs were extracted, the relative contribution of each to the observed mixed spectra of each pixel (i.e. concentration maps) was calculated. This was done using a spectral mixing model, which describes how the spectral signatures of different EMs combine to form the observed spectrum. In this case, a subtractive mixing model⁴¹ was selected, as it provided the best results in previous studies¹⁹:

$$\mathbf{Y} = \prod_{i=1}^q \rho_i^{\alpha_i} \quad (1)$$

Where $\mathbf{Y}$ is the spectral reflectance of the mixture, q is the number of candidate EMs, ρ_i is the spectral reflectance of the i th EM, and α_i its concentration. This model offers several advantages, including straightforward implementation, clear physical interpretability, computational efficiency, and mathematical tractability that facilitates optimization and analysis. Other mixing models can be used, including the Kubelka-Munk model. However, its assumptions regarding layer thickness, homogeneity, and optical properties are not compatible with the characteristics of historical documents studied and the data available in this work⁴². Moreover, its pixel-wise application is computationally demanding, making it impractical for large hyperspectral datasets¹⁵.

The concentration of each EM was calculated through an optimization process using the *fmincon* function with the interior-point algorithm⁴³ implemented in MATLAB®. The cost function combined two spectral comparison metrics: the complement of the Goodness-of-Fit coefficient (*cGFC*⁴⁴) and the *RMSE*¹⁹, as proposed in³⁰. The sum-to-one constraint ($\sum_{i=1}^q \alpha_i = 1$) and a lower bound of zero for the concentrations α_i were added to the optimization.

After generating the concentration maps for both EMs, threshold values were applied to create presence maps. Based on results from multiple preliminary tests, only pixels with an ink concentration greater than 50% or a writing support concentration greater than 90% were kept. This excludes pixels at the edges of ink strokes, where it is difficult to assign a clear label as either ink or support due to material mixing.

Machine learning models and train/test split

Support Vector Machine (SVM) classifiers were employed due to their superior performance for hyperspectral ink classification in previous work⁹. Three SVM models were trained using identical configurations: a Gaussian kernel function with a box constraint parameter of 10, as determined

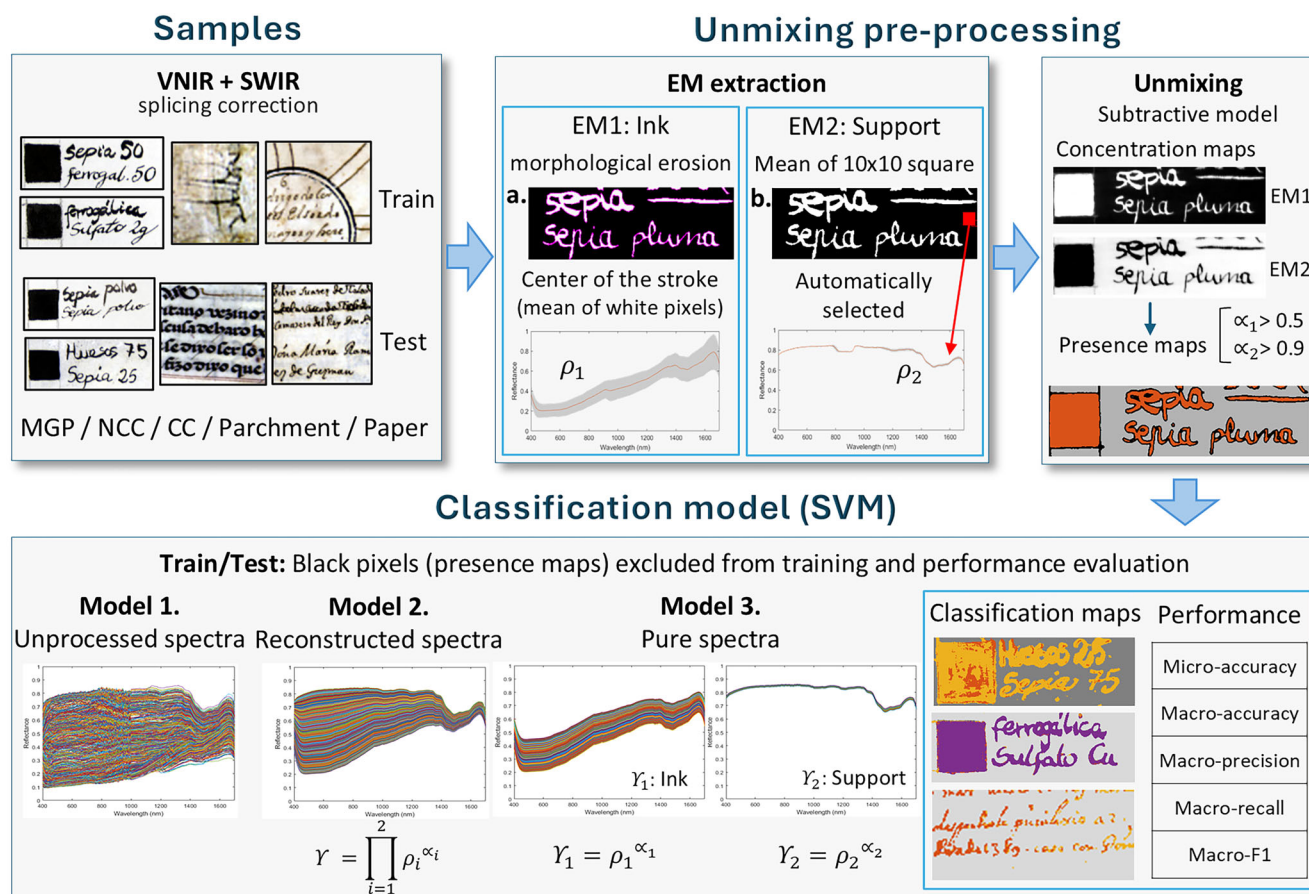


Fig. 2 | Workflow illustrating the steps involved in the process. From sample data capture and merging, unmixing pre-processing (EM extraction and presence maps), to training and testing the SVM classification models, with resulting classification maps and performance metrics.

through previous optimization. Five different classes were included: pure metallo-gallate inks (MGP), carbon-containing inks (CC), non-carbon-containing inks (NCC), parchment and paper, to develop classifiers versatile enough for most historical documents.

To ensure training data quality, pixels falling outside the concentration thresholds established in the unmixing preprocessing step (previous subsection) were excluded from the training set, thereby avoiding spectra with significant ink-support mixing that could introduce classification ambiguity. Three different models were then trained based on different spectral processing approaches (see Fig. 2 lower row):

- **Model 1: Unprocessed spectra**—This model used the unprocessed reflectance spectra, as in⁹, but only considering pixels within the previously computed presence maps for ink and writing support.
- **Model 2: Reconstructed spectra**—In this approach, the spectra within the presence maps were reconstructed using the subtractive mixing model after the preprocessing unmixing step. Each spectrum was calculated as the product of the EMs raised to their respective concentration coefficients.
- **Model 3: Pure spectra**—For each pixel within the presence maps, the spectral contributions of ink and writing support were separated, with each component obtained as its corresponding EM raised to its concentration coefficient. Unlike in Model 2, no product of the two EM contributions was calculated; instead, they were kept separate. The final spectrum retained either the ink or support component, depending on which had the highest concentration.

The same train-test split methodology used in ref.⁹ was applied for direct comparison, with the key difference that only pixels from the presence

maps were utilized for both training and testing. In addition to the original three ink classes (MGP, CC, NCC), two writing support material classes (paper and parchment) were incorporated. A total of 109 hyperspectral samples were used for training and 36 for testing, corresponding to a 75%–25% train-test split. The class-wise distribution of training and test samples is detailed in Table 1. Partitioning was performed at the document level, ensuring that different samples extracted from the same document fell into the same subset. This prevented bias in the training-test split by avoiding situations where test samples had training counterparts from the same document, thereby ensuring robust performance evaluation. Once the spectra were prepared according to each model's preprocessing approach, they were randomly shuffled before training to prevent learning order bias and enhance model generalization. The dataset presents class imbalance, as shown in Table 1. However, preliminary experiments indicated that usual class balancing techniques did not improve classification performance⁹.

Performance evaluation

Model evaluation was performed on the test set following the same preprocessing pipeline applied during training. For each test sample, the reflectance spectra underwent automatic unmixing preprocessing to extract EMs and calculate concentration and presence maps. The preprocessed spectra were then fed into each of the three trained models: unprocessed spectra (Model 1), reconstructed spectra (Model 2), and pure spectra (Model 3). The classification results were used to generate pixel-wise classification maps for visual inspection.

To ensure fair and meaningful evaluation, unmixing was applied to the test samples and presence maps with the same concentration thresholds as

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); and the two writing support materials: parchment and paper.

Class	Train		Test		Total	
	Samples	Pixels	Samples	Pixels	Samples	Pixels
MGP	49	200592	14	61528	63	262120
CC	45	299938	17	182833	62	482771
NCC	28	161899	7	19854	35	181753
Parchment	62	753220	12	229818	74	983038
Paper	47	761781	24	416683	71	1178464
Total	109	2177430	36	910716	145	3088146

**Fig. 3 | Classification maps obtained using SVM models, comparing different spectral processing approaches (Models 1, 2, and 3). Ground Truth (GT) images are shown in column 4, where black pixels indicate uncertain areas. False RGB**

images [605, 535, 430] nm are included in column 5. Purple: metallo-gallate ink (MGP); yellow: carbon-containing ink (CC); orange: non-carbon-containing ink (NCC); dark gray: parchment; light gray: cotton-linen paper.

for testing were obtained (50% for ink and 90% for writing support). Only pixels in the presence maps were used for performance metric calculations.

This evaluation strategy addresses a fundamental challenge in mixed-pixel classification: pixels outside the presence maps contain significant spectral mixing of both ink and support materials, making single-label assignment inherently ambiguous and potentially misleading. By excluding these pixels from evaluation, the performance metrics reflect the models' ability to classify clearly identifiable material regions, providing a more realistic assessment of classification performance. This exclusion does not compromise the identification of either ink or writing support, as the materials present in the excluded pixels are assumed to be the same as those in the classified regions. In rare cases, such as the presence of thin strokes made with a different material (e.g., a different ink), this assumption may not hold. However, such cases can be detected through inspection of the presence maps generated during the unmixing step, and the

concentration thresholds can be locally adapted to include thin strokes in a given area of the sample.

Model performance was assessed using two complementary approaches: micro- and macro-averaged performance metrics, as explained⁹. The micro-average approach treats all individual pixels equally, regardless of their class membership, without considering differences in the number of instances per class. In micro-averaging, accuracy, precision, recall, and F1-score are mathematically equivalent because they are all calculated as the ratio of correctly classified pixels to the total number of pixels across all classes. The macro-average approach gives each class equal weight in the final average, ensuring that performance is balanced across all material types or classes. Macro-average is computed as the arithmetic mean of the metrics for individual classes⁴⁵, making it a better measure of performance for imbalanced datasets. In this study, micro-accuracy, macro-accuracy, macro-recall, macro-precision, and macro-F1-score were calculated to provide a comprehensive and more complete performance evaluation.

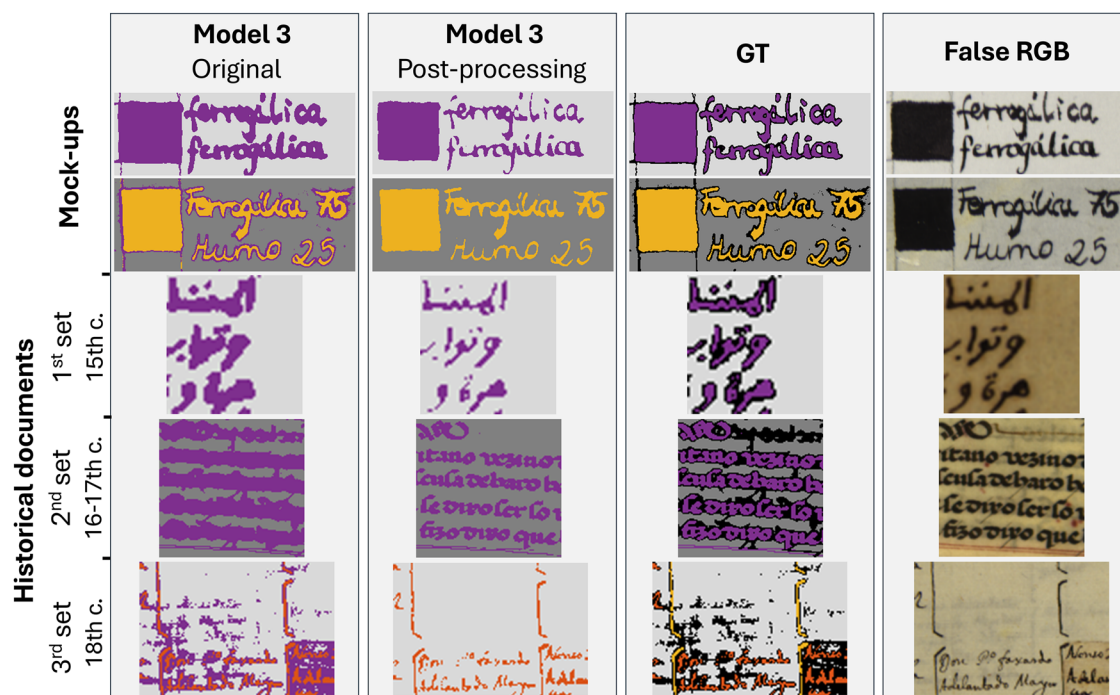


Fig. 4 | Classification maps obtained with SVM Model 3, before (column 1) and after (column 2) applying unmixing-based classification for uncertain pixels. Ground Truth (GT) images are shown in column 3, where black pixels indicate uncertain areas. False RGB images [605, 535, 430] nm are included in column 4.

Purple: metallo-gallate ink (MGP); yellow: carbon-containing ink (CC); orange: non-carbon-containing ink (NCC); dark gray: parchment; light gray: cotton-linen paper.

The full methodology workflow described in this section is illustrated in Fig. 2.

Results

Figure 3 presents examples of classification maps obtained using the three SVM models with different spectral processing approaches, along with the corresponding GT and false RGB images. The classification maps reveal that Model 2 generally produces more uniform classification of both ink-containing and support areas across all studied samples compared to the other models. Model 1 exhibits several classification errors: it tends to misclassify some carbon-containing (CC) mock-ups as non-carbon-containing (NCC) ink (first and third rows), and vice versa in one of the historical samples (7th row), while portions of the sepia pure sample (second row) are incorrectly classified as metallo-gallate ink (MGP). These issues are resolved in both Models 2 and 3. However, Model 3 presents its own challenges. It tends to classify stroke edges as MGP, making the strokes appear thicker than they actually are (particularly evident in rows 4 and 6). Additionally, Model 3 shows difficulties with writing support classification in historical samples, where portions of the support material are misclassified as MGP (see the last two rows in Fig. 3). This misclassification occurs particularly in areas with low support concentration according to unmixing (represented as black pixels in the GT images). This behavior is understandable, as Model 3 is trained with pure spectra, and the spectrum of MGP ink is the most similar to that of the writing support material, given that this ink becomes transparent in the infrared region. Notably, in rows 6 and 8, Model 3 detects bleed-through in both samples, that is, text or ink from the document's reverse side that was not originally included in the GT. Model 3 shows the highest sensitivity to this phenomenon, which is logical given its training with highly pure spectra (separated ink and writing support), enabling it to detect any spectral variations present in the document. This sensitivity represents a potential advantage, as it could be further explored to distinguish bleed-through text from front-side annotations, offering an additional application of the model. However, if bleed-through detection is not desired, the effect could be mitigated by adjusting the

classification thresholds or implementing adaptive rather than fixed thresholds (e.g., 50% for ink and 90% for writing support) depending on sample type or location within the sample. Alternatively, binarization could be applied to distinguish foreground from background, thereby better delimiting ink and writing support regions.

One of the key contributions of this study lies precisely in identifying and excluding areas where classification would be unreliable (represented as black areas in the GT images). For Model 3, these regions are typically located along stroke edges and are often misclassified as MGP. To refine the classification maps in these areas, we implemented a post-processing approach based on the unmixing results previously obtained. Specifically, we used the concentration maps to reassign the classification of pixels within uncertain regions, labeling each pixel as ink or writing support depending on which component showed the highest concentration. Once the material type was identified, the pixel was assigned the specific material (one of three inks or two writing supports) that was most frequent in its local neighborhood.

Figure 4 presents the results for selected samples after applying this post-processing method. The bleed-through information disappears in rows 4 and 5, while stroke edges appear better delineated. The last column includes false-color RGB images using VNIR bands at [605, 535, 430] nm as red, green, and blue channels, respectively, to better visualize bleed-through effects. This approach effectively resolves the edge issues seen in Model 3 without affecting classification in central stroke regions, and therefore does not alter the performance metrics reported in Table 2. However, it sacrifices bleed-through detection, and very thin strokes may remain difficult to identify, as shown in row 1.

Table 2 presents the mean performance metrics for the SVM models in the five-class classification task, distinguishing between metallo-gallate pure (MGP), carbon-containing ink (CC), non-carbon-containing ink (NCC), parchment, and paper. The results demonstrate the influence of different spectral preprocessing approaches on classification performance, with metrics computed only for pixels within the presence maps as previously described. It can be observed that all macro-metrics for Model 2 exceed

Table 2 | Performance comparison of SVM models using different spectral preprocessing approaches: unprocessed spectra (Model 1), reconstructed spectra (Model 2), and unmixing spectra (Model 3) for five-class classification (MGP, CC, NCC, parchment, paper).

Model	Micro-accuracy	Macro-accuracy	Macro-precision	Macro-recall	Macro-F1
Model 1	97.79	96.01	92.05	96.01	93.70
Model 2	98.84	98.49	96.79	98.49	97.58
Model 3	96.71	97.99	97.14	97.99	97.47

Table 3 | Performance comparison of SVM models using different spectral preprocessing approaches: unprocessed spectra (Model 1), reconstructed spectra (Model 2), and unmixing spectra (Model 3) for ink classification (MGP, CC, NCC).

Model	Micro-accuracy	Macro-accuracy	Macro-precision	Macro-recall	Macro-F1
Model 1	95.15	94.14	88.08	94.14	90.56
Model 2	97.64	97.84	95.53	97.84	96.59
Model 3	99.07	98.86	98.81	98.86	98.82

those of Model 1, while only the macro-precision of Model 3 surpasses Model 2, with other metrics remaining similar. Specifically, there is an approximately 4% difference in Macro-F1 when comparing Model 1 to Model 2, suggesting that unmixing preprocessing can significantly enhance classification performance in machine learning models. Micro-accuracy is the only metric where Model 1 shows a slightly higher value than Model 3. This can be explained by the fact that micro-accuracy does not account for class imbalance, meaning that classes with more representation in the test set, such as writing support classes, have a greater influence on the final metric. In this case, Model 3 performed worse when classifying the most represented classes. An analysis of the classification maps in Fig. 3 suggests that this discrepancy may be due to the misclassification of the support material in some historical samples (see Fig. 3 last two rows), as discussed in the classification map analysis.

For comparison with previous work⁹, paper and parchment pixels were excluded from the test results, and performance metrics were recalculated (see Table 3). However, the presence of writing support classes may still influence the predictions, as ink pixels could be misclassified as support. This accounts for the slightly inferior results obtained for Model 1 compared to those in ref.⁹, with a 91.79% Macro-F1. When writing support classes are excluded from calculations, Model 2 clearly outperforms Model 1, and Model 3 surpasses Model 2 across all evaluated metrics. Specifically, Model 3 achieves more than 8% higher F1-score compared to Model 1, reaching 98.82%, and outperforms the SVM model with cleaning post-processing proposed in ref.⁹.

Model 1 was designed to closely resemble the one in ref.⁹, but with adjusted training and test data to make it more comparable to Models 2 and 3 after preprocessing. The differences between Model 1 in this study and the one used in ref.⁹ are as follows: (i) training was conducted using only pixels from areas with high ink or writing support concentration (presence maps); (ii) the labels parchment and paper were included, not just the ink labels (MGP, CC, NCC); (iii) performance metrics for the test evaluation were calculated using only areas with high ink or support concentration. Additionally, the results presented here do not include the post-processing cleaning step proposed in ref.⁹. After incorporating the cleaning post-processing (by applying a median filter), results for Models 1 and 2 improved across all performance metrics for the five-class problem, achieving 98.43% Macro-F1 with Model 2 and 97.38% with Model 1. However, Model 3 showed reduced performance across all studied metrics

because classification errors at stroke edges propagated into central stroke regions. Therefore, this type of cleaning is not recommended for Model 3. To specifically address classification errors in stroke edges, a more effective approach is that presented in Fig. 4, which incorporates ink or writing support concentration information derived from the unmixing step.

Discussion

This study explored spectral unmixing as a preprocessing technique to improve the machine learning-based classification of historical inks and writing supports in hyperspectral imaging. We specifically assessed how spectral unmixing mitigates classification errors caused by mixed spectral contributions from ink and support materials. Three SVM models were developed and compared: Model 1 using unprocessed spectra, Model 2 using reconstructed spectra after unmixing, and Model 3 using pure unmixed spectra. The classification task distinguished between metallo-gallate pure ink (MGP), carbon-containing ink (CC), non-carbon-containing ink (NCC), parchment, and paper across both mock-up samples and historical documents.

Our results demonstrate that unmixing preprocessing improves classification performance over conventional unprocessed spectral approaches. Model 2 achieved the best overall five-class classification performance (Macro-F1: 97.58%) compared to Model 1 (93.70%), while Model 3 demonstrated superior ink-specific classification with an 8% improvement in Macro-F1 over Model 1, surpassing previous results⁹. Classification maps revealed that Model 2 provides the most uniform classification across both ink and writing support regions, while Model 3 excels in identifying the central ink stroke regions and detecting bleed-through.

By decomposing complex spectra into pure component signatures, spectral unmixing effectively addressed the challenge of mixed spectra, enabling more precise identification of historical materials. This not only improves technical performance but also adds value for end users in fields such as conservation science, restoration, and historical research, where reliable ink classification can help preservation strategies and contribute to understanding document provenance and writing practices. Although demonstrated in the context of historical documents, this preprocessing strategy has broad potential to enhance material classification in different fields affected by mixed spectra.

The main limitations include reduced classification accuracy at the edges of ink strokes, requiring post-processing for cleaner classification maps. In this study, a post-processing method based on unmixing information was implemented, which improved the results for Model 3. Regarding computational efficiency, the training of each SVM model required ~3 h, prediction about 8 min, while the unmixing-based preprocessing step added around 3 min per sample.

Future research could explore the adoption of adaptive thresholds to optimize unmixing parameters for individual samples, the automation of the structuring element size used to select the ink EM, or the implementation of a voting-based classification strategy, whereby the predominant material is inferred from the majority label assigned to its EMs. Additionally, integrating spectral unmixing into binarization pipelines holds significant potential, especially considering the ability of Model 2 to effectively separate ink and writing support regions.

Data availability

The datasets analyzed in this study are available in <https://doi.org/10.6084/m9.figshare.28319165>. The datasets generated during the current study are available from the corresponding author upon reasonable request.

Received: 27 June 2025; Accepted: 2 September 2025;

Published online: 06 October 2025

References

- González-García, S., López-Montes, A. & Espejo-Arias, T. 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*, 121–126 (CRC Press, London, 2014).
2. Melo, M. J. et al. Iron-gall inks: a review of their degradation mechanisms and conservation treatments. *Herit. Sci.* **10**, 145 (2022).
 3. Omayio, E. O., Indu, S. & Panda, J. Historical manuscript dating: traditional and current trends. *Multimed. Tools Appl.* **81**, 31573–31602 (2022).
 4. Nehring, G., Bonnerot, O., Gerhardt, M., Krutzsch, M. & Rabin, I. Looking for the missing link in the evolution of black inks. *Archaeol. Anthropol. Sci.* **13**, 1–10 (2021).
 5. Cloutis, E. A. Review article hyperspectral geological remote sensing: evaluation of analytical techniques. *Int. J. Remote Sens.* **17**, 2215–2242 (1996).
 6. Muñoz-Postigo, J. et al. Band selection pipeline for maturity stage classification in bell peppers: from full spectrum to simulated camera data. *J. Food Eng.* **365**, 111824 (2024).
 7. Freeman, J. et al. Multispectral and hyperspectral imaging: applications for medical and surgical diagnostics. In *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 'Magnificent Milestones and Emerging Opportunities in Medical Engineering'* (Cat. No. 97CH36136), vol. 2, 700–701 (IEEE, 1997).
 8. Rodrigues, E. M. & Hemmer, E. Trends in hyperspectral imaging: from environmental and health sensing to structure-property and nano-bio interaction studies. *Anal. Bioanal. Chem.* **414**, 4269–4279 (2022).
 9. López-Bal-domero, A. B., Buzzelli, M., Moronta-Montero, F., Martínez-Domingo, M. Á. & Valero, E. M. Ink classification in historical documents using hyperspectral imaging and machine learning methods. *Spectrochim. Acta Part A Mol. Biomolecular Spectrosc.* **335**, 125916 (2025).
 10. Kokla, V. Contemporary approach in the study of a Byzantine-illuminated manuscript on parchment. *Eur. Phys. J.* **136**, 786 (2021).
 11. Melessanaki, K., Papadakis, V., Balas, C. & Anglos, D. Laser induced breakdown spectroscopy and hyper-spectral imaging analysis of pigments on an illuminated manuscript. *Spectrochim. Acta Part B At. Spectrosc.* **56**, 2337–2346 (2001).
 12. George, S. & Hardeberg, J. Y. Ink classification and visualisation of historical manuscripts: Application of hyperspectral imaging. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, 1131–1135 (IEEE, 2015).
 13. Daniel, F. et al. Comparison between non-invasive methods used on paintings by goya and his contemporaries: hyperspectral imaging vs. point-by-point spectroscopic analysis. *Anal. Bioanal. Chem.* **409**, 4047–4056 (2017).
 14. Delaney, J. K., Dooley, K. A., Van Loon, A. & Vandivere, A. Mapping the pigment distribution of Vermeer's Girl with a Pearl Earring. *Herit. Sci.* **8**, 4 (2020).
 15. Rohani, N., Pouyet, E., Walton, M., Cossairt, O. & Katsaggelos, A. K. Nonlinear unmixing of hyperspectral datasets for the study of painted works of art. *Angew. Chem.* **130**, 11076–11080 (2018).
 16. Radpour, R., Gates, G. A., Kakoulli, I. & Delaney, J. K. Identification and mapping of ancient pigments in a Roman Egyptian funerary portrait by application of reflectance and luminescence imaging spectroscopy. *Herit. Sci.* **10**, 8 (2022).
 17. Corregidor, V., Viegas, R., Ferreira, L. M. & Alves, L. C. Study of iron gall inks, ingredients and paper composition using non-destructive techniques. *Heritage* **2**, 2691–2703 (2019).
 18. López-Bal-domero, A. B. et al. Hyperspectral dataset of historical documents and mock-ups from 400 to 1700 nm (HYPERDOC). *Sci. Data* **12**, 1248 (2025).
 19. Grillini, F., Thomas, J.-B. & George, S. Comparison of imaging models for spectral unmixing in oil painting. *Sensors* **21**, 2471 (2021).
 20. López-Bal-domero, A. B. et al. Hyperspectral database of synthetic historical inks. In *Archiving Conference*, vol. 21, 11–16 (Society for Imaging Science and Technology, 2024).
 21. López-Bal-domero, A. B. et al. 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, 99–111 (SPIE, 2023).
 22. Sifnaios, S. et al. A deep learning approach for pixel-level material classification via hyperspectral imaging. *arXiv preprint arXiv:2409.13498* (2024).
 23. Chen, C., Guo, B., Wu, X. & Shen, H. An edge detection method for hyperspectral image classification based on mean shift. In *2014 7th International Congress on Image and Signal Processing*, 553–557 (IEEE, 2014).
 24. Sima, H.-F., Gao, F., Liu, L.-L., Sun, J.-D. & Tang, C.-S. Edge fine classification optimization network for hyperspectral image classification. *J. Netw. Intell.* **8**, 518–533 (2023).
 25. Al-Khafaji, S. L., Zhou, J., Bai, X., Qian, Y. & Liew, A. W.-C. Spectral-spatial boundary detection in hyperspectral images. *IEEE Trans. Image Process.* **31**, 499–512 (2021).
 26. Nie, M. et al. Novel method for hyperspectral unmixing: fuzzy c-means unmixing. *Sens. Rev.* **36**, 184–192 (2016).
 27. Zhang, G. et al. Spectral variability augmented sparse unmixing of hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **60**, 1–13 (2022).
 28. Deborah, H., Ulfarsson, M. O. & Sigurdsson, J. Fully constrained least squares linear spectral unmixing of The Scream (Verso, 1893). In *2021 11th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 1–5 (IEEE, 2021).
 29. Kleynhans, T., Messinger, D. W. & Delaney, J. K. Towards automatic classification of diffuse reflectance image cubes from paintings collected with hyperspectral cameras. *Microchem. J.* **157**, 104934 (2020).
 30. Valero, E. M. et al. Unmixing and pigment identification using visible and short-wavelength infrared: reflectance vs logarithm reflectance hyperspaces. *J. Cult. Herit.* **64**, 290–300 (2023).
 31. Díaz, R. J. et al. The making of black inks in an Arabic treatise by al-Qalālūsī dated from the 13th c.: reproduction and characterisation of iron-gall ink recipes. *Herit. Sci.* **11**, 7 (2023).
 32. Contreras Zamorano, G. M. *La tinta de escritura en los manuscritos de archivo valencianos, 1250-1600. Análisis, identificación de componentes y valoración de su estado de conservación*. Ph.D. thesis, Universitat de València (2015).
 33. Arias, T. E., Stoytcheva, I. L., García, D. C., Benito, A. D. & de Haro, A. J. Caracterización material y proceso de conservación de la colección de documentos árabes manuscritos del Archivo Histórico Provincial de Granada. *Al-Qantara* **32**, 519–532 (2011).
 34. Espejo, T., Lazarova, I. & Cano, M. 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*, 33–44 (2008).
 35. Duran, A., López-Montes, A., Castaing, J. & Espejo, T. Analysis of a royal 15th century illuminated parchment using a portable XRF–XRD system and micro-invasive techniques. *J. Archaeol. Sci.* **45**, 52–58 (2014).
 36. de Guevara, M. L. 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. Asociación de. Hidalgos de. España* **1**, 94–95 (2021).
 37. Grillini, F. et al. 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**, e0303018 (2024).
 38. Grillini, F., Thomas, J.-B. & George, S. Logistic splicing correction for VNIR–SWIR reflectance imaging spectroscopy. *Opt. Lett.* **48**, 403–406 (2023).

39. Bioucas-Dias, J. M. et al. Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **5**, 354–379 (2012).
40. Buzzelli, M. et al. Handwritten ink segmentation algorithms for hyperspectral images of historical documents. *Multimedia Tools and Applications* (2025).
41. Burns, S. A. Subtractive color mixture computation. *arXiv preprint arXiv:1710.06364* (2017).
42. Kubelka, P. New contributions to the optics of intensely light-scattering materials. Part I. *J. Opt. Soc. Am.* **38**, 448–457 (1948).
43. Potra, F. A. & Wright, S. J. Interior-point methods. *J. Comput. Appl. Math.* **124**, 281–302 (2000).
44. Romero, J., García-Beltrán, A. & Hernández-Andrés, J. Linear bases for representation of natural and artificial illuminants. *JOSA A* **14**, 1007–1014 (1997).
45. Grandini, M., Bagli, E. & Visani, G. Metrics for multi-class classification: an overview. *arXiv preprint arXiv:2008.05756* (2020).

Acknowledgements

We would like to acknowledge David Torres Ibáñez, Director of the Archive of the Royal Chancellery of Granada, and Eva Martín López, Director of the Provincial Historical Archive of Granada, for their assistance. This research was supported by Grant PID2021-124446NB-I00 funded by MICIU/AEI /10.13039/501100011033 and by ERDF, EU, by the Ministry of Universities (Spain) [grant number FPU2020-05532], and by MICIU through a short-term mobility grant [grant number EST24/00040].

Author contributions

A.B.L.B.: Conceptualization, Investigation, Methodology, Data curation, Software, Formal analysis, Writing—original draft, Writing—review and editing. M.A.M.D.: Conceptualization, Methodology, Data curation, Software, Supervision, Writing—review and editing. S.G.: Conceptualization, Methodology, Supervision, Writing—review and editing, Project administration. E.M.V.: Conceptualization, Methodology, Resources, Supervision, Funding acquisition, Writing—review and editing, Project administration.

Funding

Open access funding provided by NTNU Norwegian University of Science and Technology (incl St. Olavs Hospital - Trondheim University Hospital).

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Ana Belén López-Baldero.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025