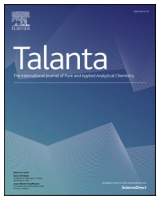
Talanta 305 (2026) 129678

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta



Development of comparative and machine learning–based methodologies for the identification of inks applicable in the field of cultural heritage and forensic science

Vanessa Pinna [a], Stefania Porcu [a,*], Gianluca Siotto [a], Enrica Tuveri [b,c], Pier Carlo Ricci [a], Edoardo Lodo [d], Pietro Coli [c], Roberto Cardia [c], Daniele Chiriu [a,**]

[a] Department of Physics, University of Cagliari, S.p. no. 8 Km 0700, 09042, Monserrato, CA, Italy
[b] Department of Chemical and Geological Science, University of Cagliari, S.P. No. 8 Km 0.700, 09042, Monserrato, Italy
[c] Scientific Investigation Department (RIS) of Cagliari, Via Ludovico Ariosto, 24, 09129, Cagliari, CA, Italy
[d] Active Label S.r.l Spin-off University of Cagliari, Via Africo 5, 09126, Cagliari, Italy

ARTICLE INFO

Keywords:
Raman spectroscopy
Ink analysis
Ink identification
Spectral comparison
Forensic applications
Cultural heritage
Spectral database

ABSTRACT

This study proposes the development of comparative and machine learning-based methodologies for the identification of inks and pigments, with potential applications in both cultural heritage diagnostics and forensic science. A preliminary selection of black inks from various pen brands was analyzed using Raman spectroscopy to define a framework for spectral comparison based on peak shifts and area ratios derived from curve fitting. The proposed method introduces a system based on spectral compatibility allowing the classification of inks based on their compositional similarity. In parallel, an automated analysis code was developed to enhance scalability and reproducibility. This system performs baseline removal, peak normalization, first-stage filtering of incompatible spectra, and refined deconvolution through pseudo-Voigt fitting, generating a numerical similarity score for each comparison. Results demonstrate that the approach allows quantitative estimation of ink compatibility and could be extended to broader datasets through the implementation of a spectral database.

1. Introduction

Inks have played a central role in human communication and cultural expression for millennia. The study of inks, both their composition and their degradation, offers insights into historical practices, technological transitions, and authorship [1–4]. In recent years, the scientific characterization of inks has gained increasing attention within the fields of cultural heritage and forensic science, motivated by the need of non-invasive methods to assess material authenticity, provenance, and condition without compromising the integrity of the artifacts [5–7].

Traditional inks, such as carbon-based inks, iron gall inks, and sepia, have been prepared using recipes that varied across time periods, geographical regions, and intended applications [8]. Carbon inks, made from soft or lamp black mixed with binders such as gum Arabic, are among the earliest used and remain stable over centuries [9,10]. Iron gall inks, widely used from the medieval period until the 19th century, consist of iron salts and tannic acids and are notorious for their corrosive effect on paper supports, leading to “ink burn” and the progressive loss of textual information. Later formulations introduced synthetic dyes and pigments, including aniline dyes and modern industrial colorants, further complicating the landscape of ink analysis due to overlapping spectral features and chemical variability [11].

Identifying and distinguishing between ink types is essential for multiple applications. In cultural heritage contexts, it supports dating, attribution, and the detection of retouching or forgeries. In forensic science, ink comparison can reveal document tampering, unauthorized additions, or substitution of pages [12]. However, despite its importance, ink identification remains a non-trivial task. The challenges arise from several factors: the chemical similarity of many inks (particularly black inks), the effects of aging and environmental exposure, the interaction with diverse substrates (e.g., cellulose-based or proteinaceous), and the stratigraphic complexity introduced by multilayered applications [13–15].

Over the past two decades, a range of spectroscopic and imaging

* Corresponding author.
** Corresponding author.
E-mail addresses: stefania.porcu@dsf.unica.it (S. Porcu), daniele.chiriu@unica.it (D. Chiriu).

https://doi.org/10.1016/j.talanta.2026.129678
Received 9 January 2026; Received in revised form 5 March 2026; Accepted 16 March 2026
Available online 19 March 2026

Table 1
Studied samples.

Reference Samples	Pen
Reference 1	PENTEL – “Feel-it! WOW”
Reference 2	BIC – “Cristal Original”
Reference 3	PENTEL – “Superb 0.7”
Reference 4	PENTEL – “Antibacterial +”
Reference 5	BIC – “Cristal Grip”
Reference 6	STAEDTLER – “Lumocolor Permanent”
Reference 7	BIC – “Orange Grip”
Reference 8	STAEDTLER – “Noris Stick 434”
Reference 9	COLOROSA – “Semigel 1 mm”
Reference 10	BIC – “Soft Feel Medium”

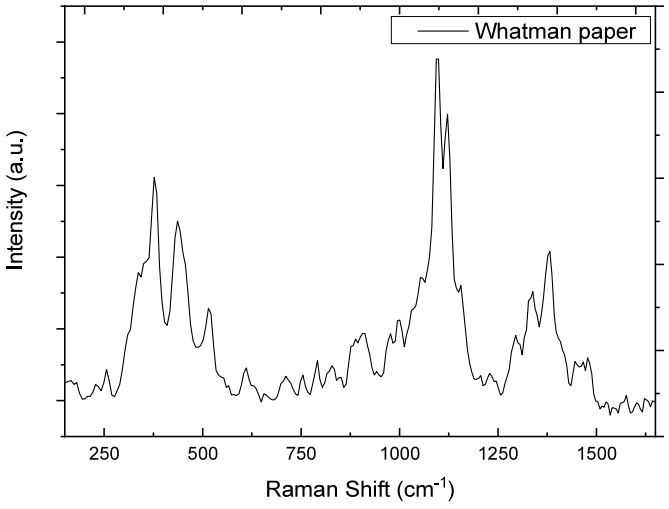


Fig. 1. Raman spectrum of Whatman paper.

Table 2
Assignations of Raman vibrations for Whatman paper.

Raman Shift (cm^{-1})	Assignments
376	Torsional and flexural vibrations of the pyran ring C–C–C
437-504	Bending vibration and expansion vibration of the C–C–O framework of the pyran ring
1093	Asymmetric and symmetric stretching vibrations of the glycosidic bond C–O–C
1316-1425	Bending vibrations of H–C–C, H–C–O, H–C–H, and C–O–H

techniques have been adopted and adapted to address these analytical challenges [7,13,14,16–22].

Laser-induced breakdown spectroscopy, though micro-destructive, has proven effective in ink analysis due to its ability to detect elemental components and reveal stratigraphic information when multiple ink layers are present [23–25]. For example, this technique has been used to determine the sequence of ink deposition in prints and to detect hidden or underlying inscriptions. Similarly, X-ray fluorescence spectroscopy, particularly in portable form, enables the non-invasive elemental analysis of inks, helping to differentiate formulations based on metal content, such as the iron and copper found in gall inks and phthalocyanine-based inks [26–28].

In parallel, hyperspectral imaging has introduced new capabilities for mapping the spatial distribution of inks across documents [8,29]. By capturing reflectance spectra at each pixel, hyperspectral imaging systems provide three-dimensional data cubes (two spatial and one spectral dimension) that can be processed using chemometric and machine learning techniques to classify inks based on their spectral response.

Vibrational spectroscopy, and particularly Raman spectroscopy, has

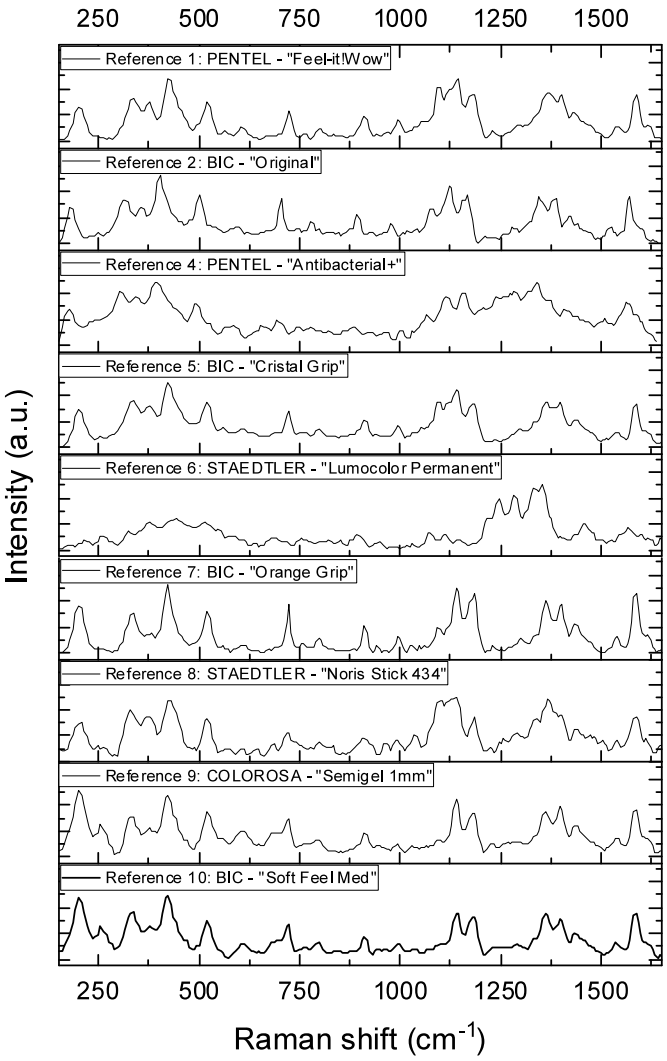


Fig. 2. Raman spectra of the analyzed samples.

emerged as one of the most powerful tools for ink identification [30–32].

Raman spectra provide molecular fingerprints based on vibrational transitions that can be used to discriminate between different ink compositions, even within the same chromatic family [33–35]. Its advantages include high molecular specificity, spatial resolution on the micrometer scale, and minimal sampling requirements, making it suitable for in situ applications on valuable or fragile artifacts.

Compared to elemental techniques such as XRF and LIBS, the Raman-based approach is more effective for discriminating against inks formulations, while offering higher chemical specificity than hyperspectral imaging, even though with longer acquisition times.

Despite these advances, the classification of ink remains a partially unsolved problem. Spectral overlaps, substrate interference, and aging-related changes introduce uncertainties that limit the reliability of direct spectral comparison.

In this context, the objective of this study is the identification and classification of modern black inks from various commercial pen brands. The proposed approach is based on Raman spectroscopy and employs curve fitting to extract key spectral features such as peak shifts and intensity ratios. These are then evaluated within a compatibility model that accounts for instrumental resolution and experimental uncertainty. The methodology emphasizes the standardization of spectral acquisition and the minimization of substrate influence, enabling more accurate and reproducible comparisons across samples. Additionally, the detection of outlier peaks is used to mark compositional anomalies and support the

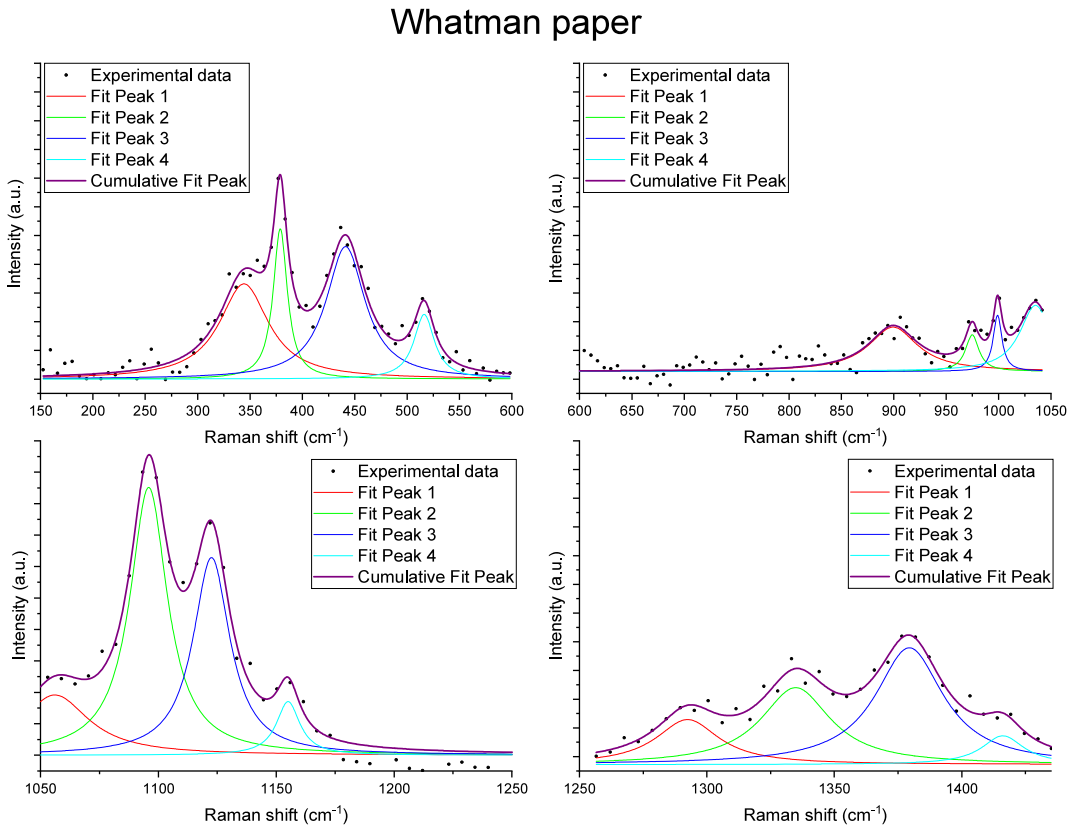


Fig. 3. Raman spectra deconvolution for Whatman paper divided in four sections.

identification of non-matching formulations.

In contrast to established chemometric approaches such as PCA, LDA, or SVM the proposed compatibility model acts as a statistical decision-making framework. By integrating the specific instrumental broadening and the variance observed in reference samples, it establishes a threshold for 'spectral identity', effectively quantifying the likelihood that two samples originate from the same chemical formulation.

Furthermore, an automated scoring algorithm was developed with the assistance of dedicated models designed to process large spectral datasets, exclude fewer probable observables, and refine the matching procedure to yield the most reliable assignment.

By enabling a classification of ink based on their spectral profiles, this approach contributes to the broader goals of non-invasive diagnostics in cultural heritage and forensic science. It also lays the basis for the creation of spectral databases that can support large-scale ink comparison, digital archiving, and decision-making in conservation and authentication practices.

2. Materials and methods

2.1. Ink samples preparation

Mock-up samples were prepared by coloring half of the surface of each Whatman filter paper disk using a different pen (Fig. S1 in Supporting Information). Whatman paper, made of pure cellulose, was chosen because of its homogeneous composition and absence of additives, which minimize background interference and ensure reliable and reproducible data acquisition [36,37]. This controlled substrate was selected to isolate the intrinsic Raman markers of the inks from external chemical variables, providing a fundamental baseline for the algorithm's validation. While this study focuses on pure cellulose, the methodology is developed to be potentially applicable to more complex archaeological or forensic supports by relying on invariant vibrational modes identified through automated baseline subtraction and peak deconvolution. All samples were obtained from ballpoint pens, except for Reference 6, which was derived from a marker pen (Table 1).

2.2. Characterization

Near infrared micro-Raman scattering measurements were carried out in back scattering geometry with the 1064 nm line of an Nd:YAG laser. Measurements were performed with a compact spectrometer B&WTEK (Newark, NJ, USA) i-Raman Ex integrated system with a spectral resolution of 9 cm^{-1}. The instrument is connected to a Raman Video Micro Sampling System equipped with a 20 × Olympus objective to select the area on the samples. For each sample, three different spots were selected along the ink stroke, and three spectra were acquired for each spot; these replicates were subsequently averaged to obtain a representative mean spectrum for each position, ensuring high signal-to-noise ratio and statistical repeatability. To evaluate the between-sample reproducibility, three independent mock-ups were analyzed for each pen model (same brand and type). The acquisition settings were kept constants for all measurements, with an integration time of 20 s, 2 accumulations per spectrum, and a laser power set at 30% (100% corresponds to 400 mW) to prevent thermal degradation of the substrates.

2.3. Data analysis

For the data analysis has been adopted a criteria based on the Raman shift of each spectral band and the relative area ratios, which are derived from the parameters resulting from curve fitting [38]. Structural variations or amorphization of a compound, resulting from external treatments, can be detected as shifts in peak positions and/or as broadening of certain bands [35,39]. An unknown sample has been selected, from which the peak positions and areas are extracted through fitting; these

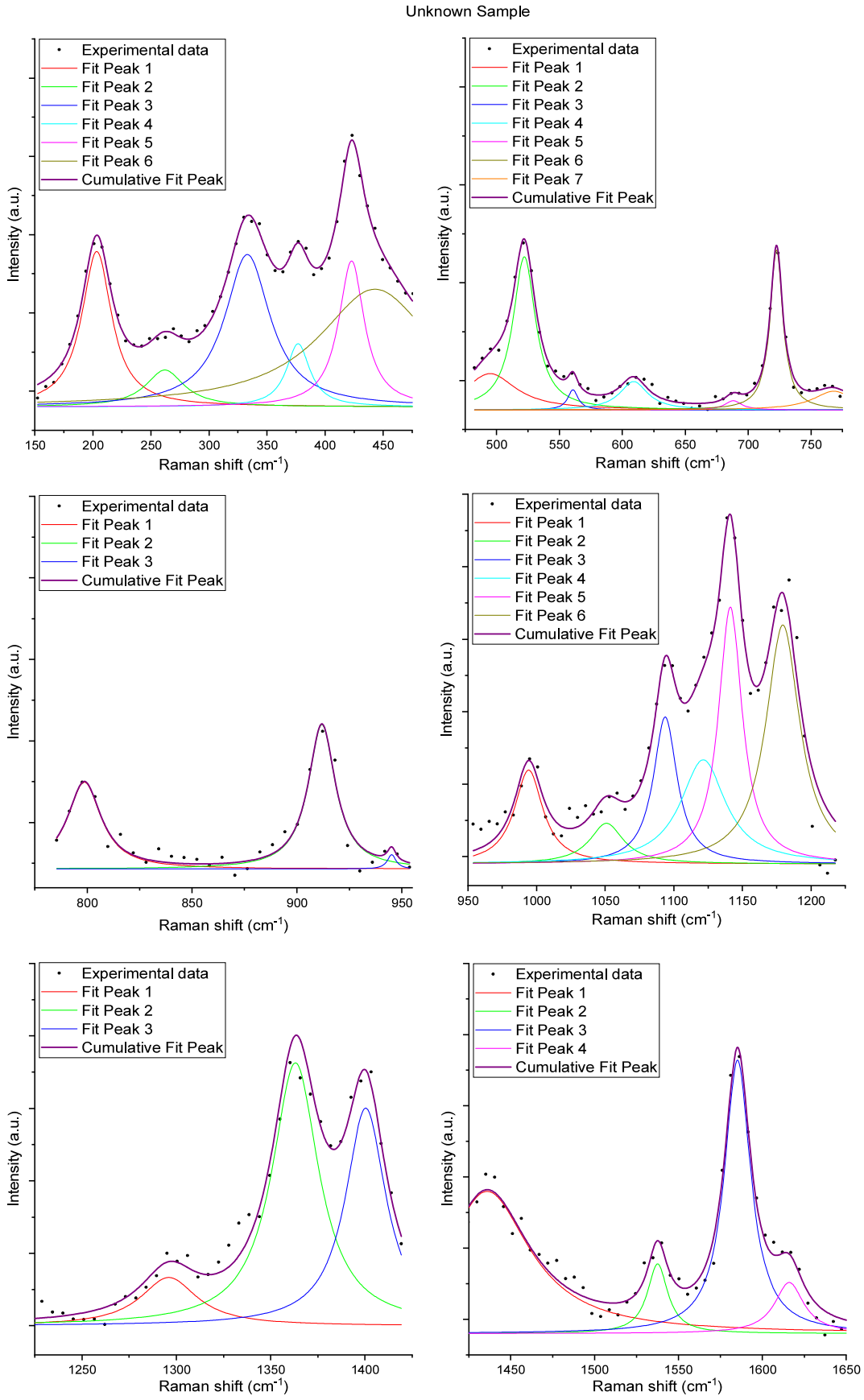


Fig. 4. Six-section deconvolution of the Raman spectrum of the unknown sample".

Table 3
Synopsis of peak positions in the Raman spectrum of the unknown samples. Indices of vibrational mode assignments: t, torsion; γ, out-of-plane deformation (with respect to the benzene ring); ν, stretching (s, symmetric; as, asymmetric); δ, bending.

Peak	x_c [cm^{-1}]	δx_c [cm^{-1}]	A [a. u.]	δA [a. u.]	Assignments
1	203	6	9,69	0,38	$\tau(CH_3)$
2	262	6	2,95	0,32	*δ (Ring)/Lattice mode*
3	423	6	7,88	0,57	$\delta(CNC)/\delta(CC_{center}C)$
4	495	6	2,91	0,44	δ (CNC)/δ (C–C–C)
5	561	6	0,32	0,08	$\gamma(CCC)/\delta(CNC)/\delta(CC_{center}C)$
6	609	6	1,23	0,12	$\delta(CCC)/\delta(CNC)/\nu_s(CC_{center}C)$
7	688	6	0,24	0,15	γ (CH)/δ (Ring)
8	723	6	3,12	0,17	ν(CN)
9	768	6	1,19	0,37	$\nu_s(CC_{center}C)/\nu(CN)$
10	799	6	1,63	0,32	δ (Ring) breathing
11	945	6	0,07	0,11	$\nu(C–C)/\rho(CH_3)$
12	1141	6	5,84	0,87	$\nu_s(CC_{center}C)/\nu(CN)$
13	1179	6	7,77	0,45	$\nu_s(CC_{center}C)/\delta(CCC)_{breathing}$
14	1363	6	8,55	0,69	$\nu_{as}(CC_{center}C)/\delta(CCC)_{ring}/\delta(CH)$
15	1400	6	6,19	0,62	$\delta(CH)/\delta_s(CH_3)/\delta(CCC)_{ring}$
16	1436	6	9,67	0,78	$\delta_{as}(CH_3)$
17	1538	6	1,13	0,23	$\nu(C_{ring}N)/\delta_s(CH_3)$
18	1585	6	5,73	0,18	$\nu(C_{ring}N)/\delta_s(CH_3)$
19	1616	6	1,18	0,18	$\nu(C_{ring}N)/\delta_s(CH_3)$

parameters serve as the baseline for comparison with the other reference samples. For the area ratio comparison, the area has been normalized to the most intense band and the area ratios calculated using the area of the peak corresponding to the most intense band of the reference sample. The area values obtained from fits with an estimated error above a certain threshold should be excluded from the analysis: values beyond this threshold are considered unreliable and are not used for subsequent calculations (Equation (1)).

$$\varepsilon = \frac{\delta A}{A} > 20\% \qquad \text{Equation 1}$$

To assess whether a peak in each ink sample is compatible with the corresponding peak in the unknown sample, the concept of discrepancy is applied. Discrepancy is a measure of the agreement between measurements of the same quantity; it is considered non-significant when the difference between the best estimates of the two values is less than the sum of their associated uncertainties, which, in this context, are primarily due to the spectrometer's resolution.

$$|x - x_{REF}| < (\delta x_2 + \delta x_1) \qquad \text{Equation 2}$$

To ensure the transferability of the method across different Raman systems, the compatibility threshold is not fixed but is dynamically linked to the specific instrument's spectral resolution through Equation (2). This approach allows to remain platform-independent, as the comparison logic scales with the specific acquisition settings (wavelength, resolution, and objective) rather than absolute intensity values.

Specifically, the discrepancy threshold for peak positions, Δx, is rigorously derived from the sum of the individual instrumental uncertainties (δx ≈ 6 cm^{-1} for the current setup), resulting in a global tolerance of 12 cm^{-1}. This ensures that compatibility is only claimed when differences fall within the combined confidence interval of the measurement. Regarding the relative area ratios, a fixed tolerance of 20% was established by propagating the uncertainties occurring from the baseline subtraction and the pseudo-Voigt curve-fitting process. The algorithm implements a filtering criterion that automatically excludes any vibrational mode where the relative error (ε) exceeds the 20% threshold, ensuring that the final compatibility score is based only on statistically significant features.

For the area ratios, parameters are considered compatible when:

$$\frac{|R_{REF} - R_{EXP}|}{R_{REF}} < 20\% \qquad \text{Equation 3}$$

where R_{REF} is the area ratio in the reference sample and R_{EXP} is the same ratio in the test sample.

The tolerance adopted for peak area ratios was derived from formal uncertainty propagation associated with the spectral preprocessing and fitting procedures. Peak areas were obtained through pseudo-Voigt fitting following polynomial baseline subtraction. The uncertainty of each fitted area was estimated from the covariance matrix of the non-linear least-squares fit combined with the residual variance of the baseline correction.

For a ratio between two peak areas $R = A_1/A_2$, the relative uncertainty was propagated according to standard error propagation:

$$\frac{\vartheta_R}{R} = \sqrt{\left(\frac{\vartheta A_1}{A_1}\right)^2 + \left(\frac{\vartheta A_2}{A_2}\right)^2} \qquad \text{Equation 4}$$

where A_1 and A_2 are the fitted peak areas and σ_{A1}, σ_{A2} their associated uncertainties. Across the analyzed spectra, the propagated relative uncertainty of area ratios typically ranged between 12% and 18%, depending on peak overlap and baseline variability. Based on this distribution, a 20% tolerance threshold was adopted as a conservative upper bound. Vibrational modes whose propagated relative error exceeded this value were automatically excluded from subsequent analysis steps. This filtering ensures that the compatibility scoring procedure relies exclusively on spectral features whose intensity ratios remain statistically significant within the estimated uncertainty bounds.

When any of the above conditions is satisfied, a unit score is assigned to the corresponding parameter; otherwise, no score is assigned. Additionally, if a test sample spectrum presents extra peaks not found in the unknown sample, indicating differences in ink composition, a negative unit score is assigned for each of those peaks.

Once scores have been assigned to all applicable criteria within the selected confidence interval, the compatibility percentage between inks is calculated as the algebraic sum of all scores divided by the total number of selected criteria in the spectrum of the unknown sample.

The "Compatibility Score" works as a statistically based accept/reject threshold. A score of 100% indicates a correct identification, while lower scores clearly discriminate against non-matching inks.

3. Discussion and results

The selection of a small group of inks serves as a starting point to develop a new methodology for the identification of inks and pigments, which could potentially be expanded to a larger number of pigments, with the goal of creating a comprehensive database.

Such a database would allow for a faster determination of the compatibility of an unknown sample, analyzed under experimental conditions consistent with those of the tabulated pigments, with any reference sample in the database.

To develop this methodology, different models of black ink pens, both from the same brand and from other well-known manufacturers, were considered. The objective is, on one hand, to investigate the possible existence of a common matrix among the ink constituents and, based on this, to determine the compatibility of the formulations used by different producers; on the other hand, to observe any differences in the way the compounds were thermally treated or mixed with other binders.

The results obtained using this method allow the discrimination of inks that are compatible within a certain percentage threshold.

Since the inks were deposited on paper, the contribution of the paper was excluded through a separate Raman analysis (Fig. 1), carried out following the same procedure applied to the ink samples. This approach enabled the identification and exclusion of Raman bands attributable to

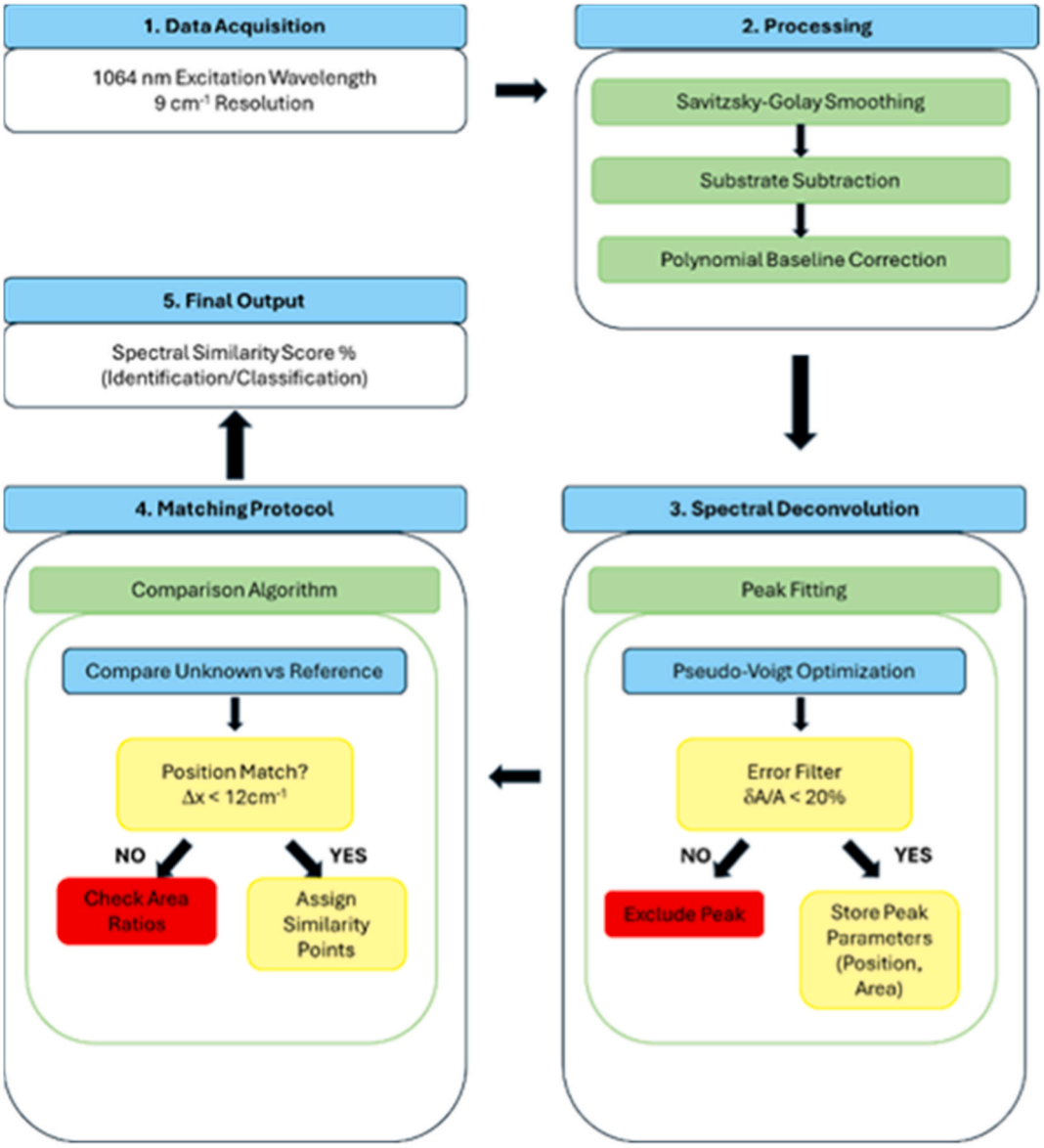


Fig. 5. Flowchart of the standardized analytical procedure developed for inks identification.

the paper from the ink spectra. The main Raman vibrational modes identified for the Whatman paper used as substrate are presented in Table 2 [40].

In some cases, it was not possible to obtain clearly distinguishable spectra. Specifically, Reference 3 showed localized heating at the point of analysis, even at 50% of the laser power, resulting in detector saturation. On the contrary, when using lower laser power, no distinguishable Raman bands were observed, even with prolonged acquisition times or an increased number of accumulations. Furthermore, tests using a near-infrared (NIR) excitation source at $\lambda = 785$ nm (i.e., higher excitation energy) also failed to obtain any detectable Raman signal for this sample.

The spectra obtained from the analysis of the ink samples are presented in Fig. 2. To validate the reproducibility of the method, each ink was analyzed using 15 replicates (3 spectra per spot across 5 different spots). The compatibility scores reported in the following analysis include the calculated 95-98% confidence intervals (CI). The results (reported below) demonstrate that the variance within replicates of the same ink is consistently below the 5% threshold, whereas the discrimination gap between different ink brands remains significantly higher, ensuring a clear-cut identification even when accounting for spectral noise and baseline variations.

The spectral region selected for analysis ranges from 150 cm^{-1} to 1700 cm^{-1}. The lower limit was chosen to exclude the tail of the excitation peak, while the region above 1700 cm^{-1} was omitted, as no Raman signals were observed in that range for any of the samples.

A baseline correction was applied to each ink spectrum to remove background counts caused by excitation light scattering.

The spectral signal was processed through deconvolution of the entire spectral profile to isolate the contribution of individual bands, which, when summed together, yield the best approximation of the original spectrum. The spectra of reference samples 4, 6, and 8 exhibited a highly noisy background, though distinguishable even if weak Raman signals. Therefore, prior to fitting the profiles, it was necessary to apply a smoothing function to the intensity data using the Savitzky-Golay method, to mitigate high-frequency noise and to facilitate the identification of the Raman features. To further prevent noise from being misinterpreted as an 'additional peak', a derivative-based filtering stage was employed. This ensures that only signals with a statistically significant signal-to-noise ratio are considered, making the manual discrepancy evaluation robust and independent of the specific smoothing parameters. The Savitzky-Golay smoothing parameters were explicitly set to a

Table 4
Criteria comparison between the unknown sample and reference samples 2,1 and 4.

Criteria	Type of criteria	Unknown Sample	Reference 2: BIC - "Original"			Reference 1: PENTEL - "Feel-it! Wow"			Reference 4: PENTEL - "Antibacterial +" (smooth 7 pt)				
												Additional peaks	
1	Peak 1	203	203	0	×	203	0	×	204	1	×	Peak X_1	1268
2	Peak 2	262	262	0	×	-			267	5	×		
3	Peak 3	423	423	0	×	423	0	×	419	4	×		
4	Peak 4	495	495	0	×	495	0	×	-				
5	Peak 5	561	561	0	×	562	1	×	552	9	×		
6	Peak 6	609	609	0	×	607	2	×	607	2	×		
7	Peak 7	688	688	0	×	685	3	×	684	4	×		
8	Peak 8	723	723	0	×	724	1	×	723	0	×		
9	Peak 9	768	768	0	×	764	4	×	774	6	×		
10	Peak 10	799	799	0	×	800	1	×	802	3	×		
11	Peak 11	945	945	0	×	945	0	×	948	3	×		
12	Peak 12	1141	1141	0	×	1143	2	×	1138	3	×		
13	Peak 13	1179	1179	0	×	1179	0	×	1184	5	×		
14	Peak 14	1363	1363	0	×	1366	3	×	1363	0	×		
15	Peak 15	1400	1400	0	×	1403	3	×	1393	7	×		
16	Peak 16	1436	1436	0	×	1433	3	×	1434	2	×		
17	Peak 17	1537	1537	0	×	1540	3	×	1530	7	×		
18	Peak 18	1585	1585	0	×	1587	2	×	1585	0	×		
19	Peak 19	1616	1616	0	×	1620	4	×	1610	6	×		
20	Ratio A2/A1	0,30	0,30	0	×	-			1,38	3,53			
21	Ratio A5/A1	0,81	0,81	0	×	0,51	0,37		0,60	0,26			
22	Ratio A7/A1	0,30	0,30	0	×	0,26	0,15	×	-				
23	Ratio A10/A1	0,13	0,13	0	×	0,25			0,23	0,82			
24	Ratio A12/A1	0,32	0,32	0	×	0,37	0,16	×	-				
25	Ratio A14/A1	0,17	0,17	0	×	-			0,19	0,10	×		
26	Ratio A21/A1	0,60	0,60	0	×	0,72	0,20	×	0,74	0,24			
27	Ratio A22/A1	0,80	0,80	0	×	0,96	0,19	×	1,30	0,62			
28	Ratio A24/A1	0,88	0,88	0	×	1,34	0,52		-				
29	Ratio A25/A1	0,64	0,64	0	×	0,70	0,10	×	0,11	0,83			
30	Ratio A26/A1	1,00	1,00	0	×	1,28	0,29		-				
31	Ratio A27/A1	0,12	0,12	0	×	0,12	0,00	×	-				
32	Ratio A28/A1	0,59	0,59	0	×	0,65	0,11	×	0,64	0,08	×		
33	Ratio A29/A1	0,12	0,12	0	×	0,08	0,30		-				
Total score			33			25			20			-1	
Matching			100%			76%			58%				

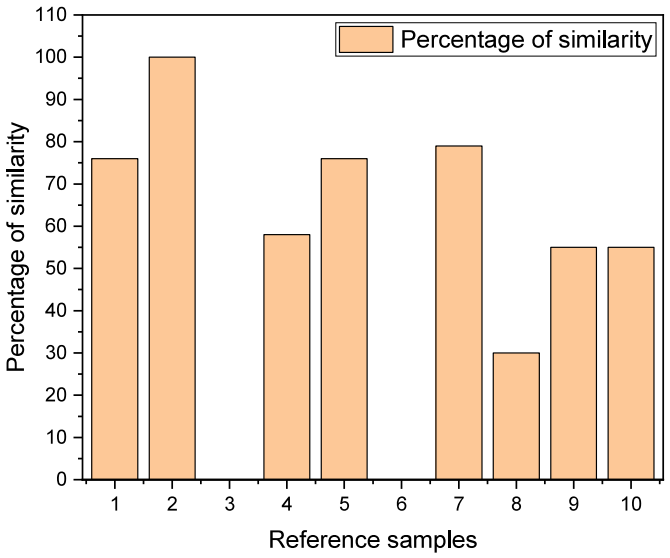


Fig. 6. Percentage of similarity between the unknown sample and the reference samples.

window size of 9 points and a polynomial order of 3. To confirm the robustness of our results, we performed a sensitivity analysis by varying the window size between 7 and 11 points and the polynomial order between 2 and 4. We observed that neither the detected peak positions nor the peak area ratios were significantly affected, and the resulting compatibility scores remained stable.

Given the large number of bands present, to improve the accuracy of the deconvolutions, each spectrum was divided into separate regions for individual analysis. Specifically, six sections were identified for the ink data, while the paper data were divided into four intervals. The band deconvolution was performed using Lorentzian profiles, as this function appropriately describes the line shape.

The analytical expression of the Lorentzian curve is given by:

$$y = y_0 + \frac{2A}{\pi}\frac{w}{4(x-x_c)^2 + w^2} \qquad \text{Equation 5}$$

where the parameters A, w, and x_c represent the area, the half-width at half-maximum (HWHM), and the peak center, respectively.

Fig. 3 shows the Raman spectra deconvolution for Whatman paper.

The most intense peak observed in the spectrum of Whatman paper, composed of pure cellulose, is located around 1096 cm^{-1}. The bands of the paper play a fundamental role in the data analysis process, and the positions of their centers are listed in Table 2.

It was observed that all samples, except for reference 6, displayed very similar bands at 423 cm^{-1}, associated with CNC bond bending, and at 1540, 1587, and 1620 cm^{-1}, associated with the stretching of benzene rings in triarylmethane compounds.

Additionally, a band at 203 cm^{-1} is observed, associated with the torsion of the methyl group (CH_3), a band at 723 cm^{-1} related to CN group stretching, and an out-of-plane deformation with respect to the benzene ring located around 561 cm^{-1} [41].

Fig. 4 shows the six-section deconvolution of the Raman spectrum of the unknown sample.

The peak positions, along with their respective areas obtained from the deconvolutions, are reported below (Table 3).

The same data processing pipeline, comprising Savitzky-Golay smoothing, polynomial baseline subtraction, and pseudo-Voigt deconvolution, was applied consistently to all acquired spectra. (see SI Figures S2-S10 in supporting Information and Fig. 5).

A comparison with the literature [41,42] clearly indicates that the main component of the inks (with the exception of reference 6 is crystal violet pigment (also known as gentian violet, $C_{25}H_{28}N_3Cl$, basic violet 3). The bands that do not match this compound, specifically those located at 262, 495, 688, 799, 945, and 1436 cm^{-1}, may correspond to a binder or solvent present in the ink formulation, depending on the manufacturer's recipe. These additives are generally used to enhance properties such as durability, consistency, fluidity, viscosity, and ink adhesion.

4. Application of methodology and discussion of results

Once the bands for the unknown sample, were identified through deconvolution procedures, the criteria for comparison were established based on the peak positions and by normalizing the areas relative to area A_1, which represents the largest peak area.

Since in the unknown sample, peaks 5, 7, 9, and 11 exhibit a percentage error $\varepsilon > 20\%$ for their respective areas, according to the rule established in Equation (1) these peaks must be excluded from further analysis as they are considered unreliable data; the same treatment is applied to other samples when necessary.

A total of 33 parameters were identified for matching with the other inks, consisting of 19 peaks and 14 area ratios.

The possible presence of additional peaks in the samples compared, relative to the reference, contributes to a negative score, resulting in a reduction in the score assigned based on the corresponding criteria.

In the specific case of this experiment, given that the spectral resolution of the spectrophotometer used is ~ 9 cm^{-1} and that a minimum of 3 pixels is required to define a peak, the estimated error on peak position is set to $\delta x = 6$ cm^{-1}.

According to Equation (2), in cases of non-significant discrepancy $|x - x_{REF}| < 12\ cm^{-1}$, the criterion value assigned to the peak is maintained.

Referring to Equation (3), the confidence interval for assigning the criterion regarding area ratios is defined as $\frac{|R_{REF} - R_{EXP}|}{R_{REF}} < 20\%$, where R_{REF} is the area ratio derived from the data of the unknown sample.

To validate the scalability and accuracy of the identification system, a leave-one-out cross-validation style evaluation was used. In this procedure, each ink spectrum in the dataset was alternately designated as an 'unknown sample' and removed from the reference library. This query sample was processed using the exact same acquisition and deconvolution pipeline as the references to ensure that any findings were inherent to the ink composition rather than data processing. The 'unknown' was then compared against all remaining entries to calculate its compatibility score and quantify the method's ability to correctly match replicates while discriminating against distinct formulations.

Table 4 reports the calculated compatibility percentages between the unknown sample and Reference 2, Reference 1, and Reference 4 inks according to the methodology described above.

Furthermore, detailed assessments of the compatibility of the unknown sample with the different inks are provided in Supporting Information, from Table S1 to Table S3 in Supporting Information.

The unknown sample shows 100% compatibility with Reference 2 and 79% compatibility with Reference 7. Ballpoint pens from the same manufacturer, such as Reference 2, BIC – "Cristal Original," and Reference 7, BIC – "Orange Grip," show a difference of compatibility of 21%. This implies that even a 21% difference, derived from the criteria used in the calculation, highlights differences between the two inks. It is also notable that the unknown sample is largely comparable to Reference 1, the PENTEL – "I feel-it! Wow" pen (76%), despite belonging to a different manufacturer. In this case, the distinction could be further refined by tightening the confidence interval for the area ratios, as expressed in Equation (3), from 20% to 10%. Doing so results in a compatibility percentage of 61% for Reference 1 (PENTEL – "I feel-it! Wow") and 73% for Reference 7 (BIC – "Orange Grip"), allowing a more precise assessment of which of the two inks is more like the chosen standard.

In contrast, the unknown sample is not comparable to the marker

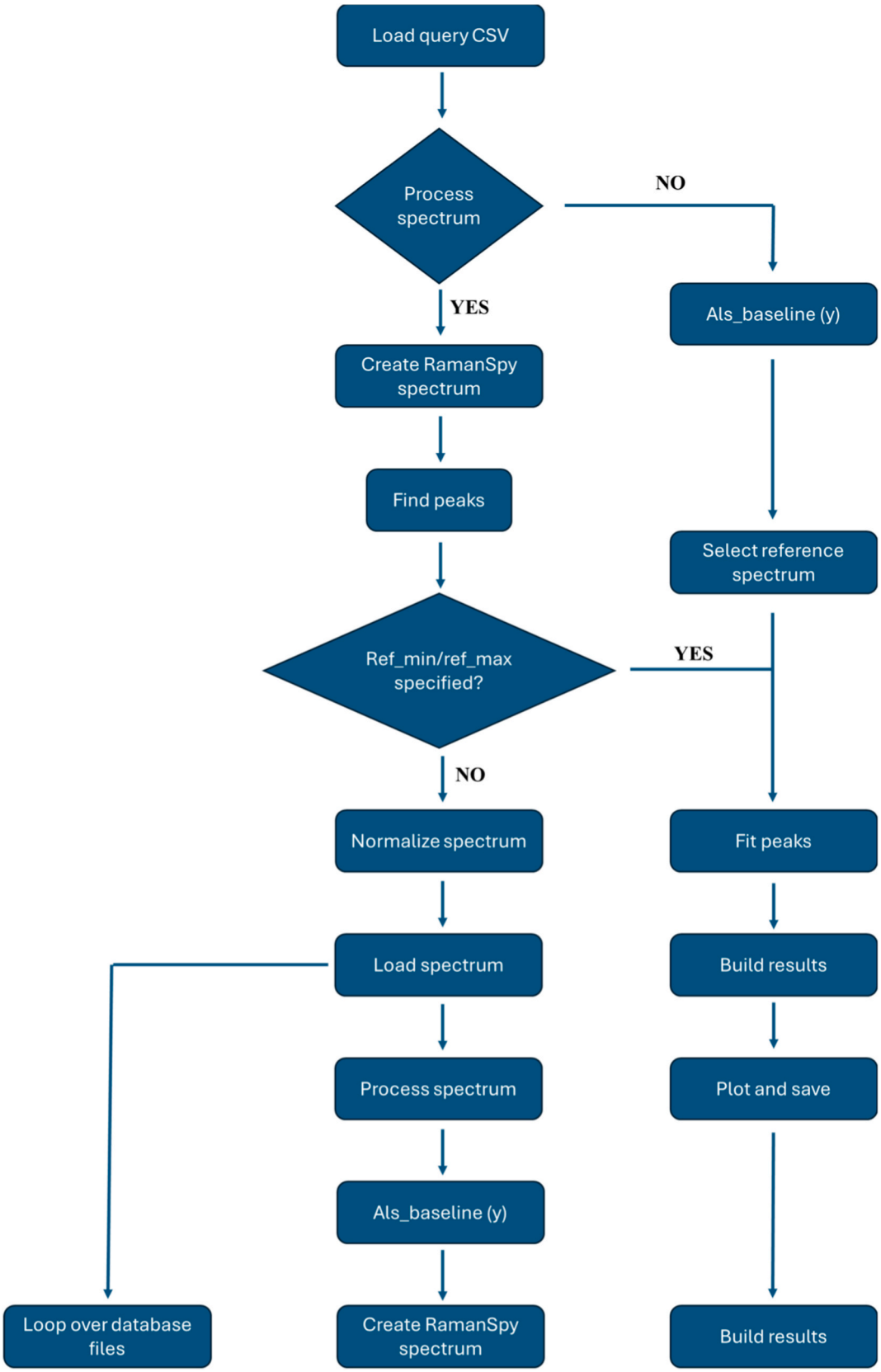


Fig. 7. Flow chart of the code.

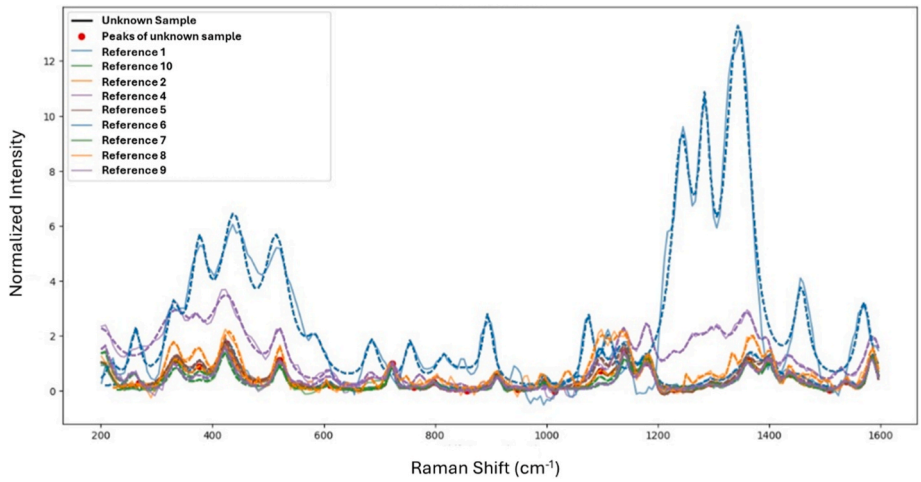


Fig. 8. Fit peaks using the automatic calculation algorithm.

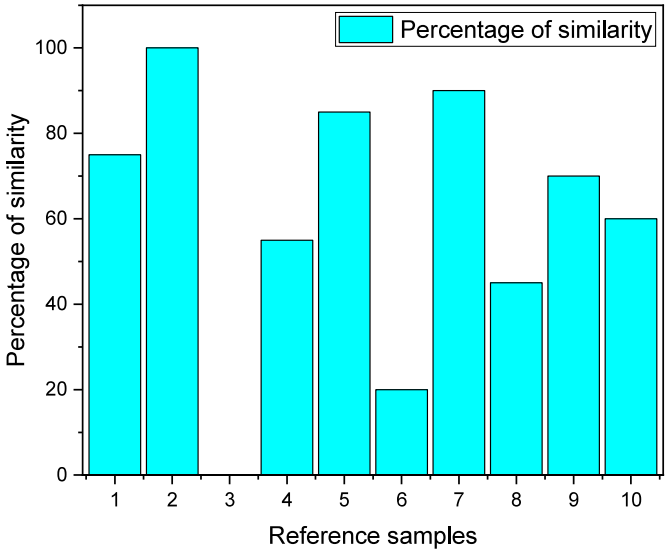


Fig. 9. Percentage of similarity using the automated analysis.

corresponding to Reference 6, STAEDTLER – "Lumocolor Permanent," since their compatibility is effectively zero. Not only does Reference 6 share no common area ratios with the unknown sample, but its spectrum also exhibits a high number of additional peaks, further reducing similarity.

Similarly, the compatibility of the unknown sample with the spectrum of Reference 8, STAEDTLER – "Noris Stick 434," is relatively low (30%) due to the appearance of additional peaks that contribute negatively to the score, indicating a high probability that the two ballpoint pens can be readily distinguished.

Consider the pens corresponding to Reference 9, COLOROSA – "Semigel 1 mm" (55%), and Reference 10, BIC – "Soft Feel Med." (55%): although the former is from a different manufacturer and the latter from the same manufacturer, they both show the same level of compatibility with the unknown sample. Nevertheless, a 55% match is sufficient to distinguish that their spectra does not closely resemble that of unknown sample.

An interesting point to note is that the spectrum of the unknown sample is clearly not confusable with that of Reference 3, PENTEL – "Superb 0.7," since its spectrum cannot be processed using the employed experimental setup because with an excitation at 1064 nm, the pigment contained in this sample produces a strong luminescence effect.

In general, the spectra subjected to intensity smoothing (References 4,6,8) display additional peaks and partially compromised area ratios that show little affinity with the reference sample. This result highlights that, since deconvolutions cannot be performed directly on the original data, there is a consequent reduction in the compatibility percentage with reference ink.

In Fig. 6, the histogram reports the similarity percentages between the reference inks examined and the unknown sample.

Finally, the results obtained quantitatively demonstrate the applicability of this technique in distinguishing compounds based on a calculated compatibility percentage. Assuming the analysis of an unknown pigmented compound of similar composition and relying on a database containing the comparison criteria for a large number of samples collected with a given experimental setup, it would be possible to determine its percentage compatibility with each entry and assign its chemical identity accordingly.

5. Automatized data analysis

Starting from the exposed approach, we implemented an automatic calculation algorithm designed specifically on the parameters defined in previous paragraphs. With the help of a Python code, we built a sequence of operations consisting in different steps to clean the row data, promote a first stage of filtering and then carrying out the final refinement. The code acts in the overall spectrum, allowing a complete vision of the entire process involving each compared peak. Fig. 7 shows the process flowchart of the code which represents each part of the logic sequence.

The first step refers to the baseline remotion from the spectra, activating a threshold on the noise variation and telling apart the baseline profile. The functions "ramanspy" and "scispy" of the code, acting as derivative of a selected range of points, ensures the result. The second step implies the normalization of the spectra intensities to a refence peak individuated as the common one among the spectra or as a reference for the user. The third step starts with the study of position matching related to the most intense peaks (a parameter of maximum intensity percentage can be set to proceed for gradual comparison). This phase corresponds to a first filtering procedure that excludes all the targets in which the peaks are out of the selected confidence. In the perspective of a large number of comparisons the first filtering reduces the range of analysis. Even if at the follow of this step the number of spectra to be analyzed is anyway big, a more stringent intensity percentage parameter can be selected to refine out. Also in this case, the functions "ramanspy" and "scipy" of the code ensures the result. When the position of the principal peaks corresponds to an acceptable limited number of spectra, the refinement can

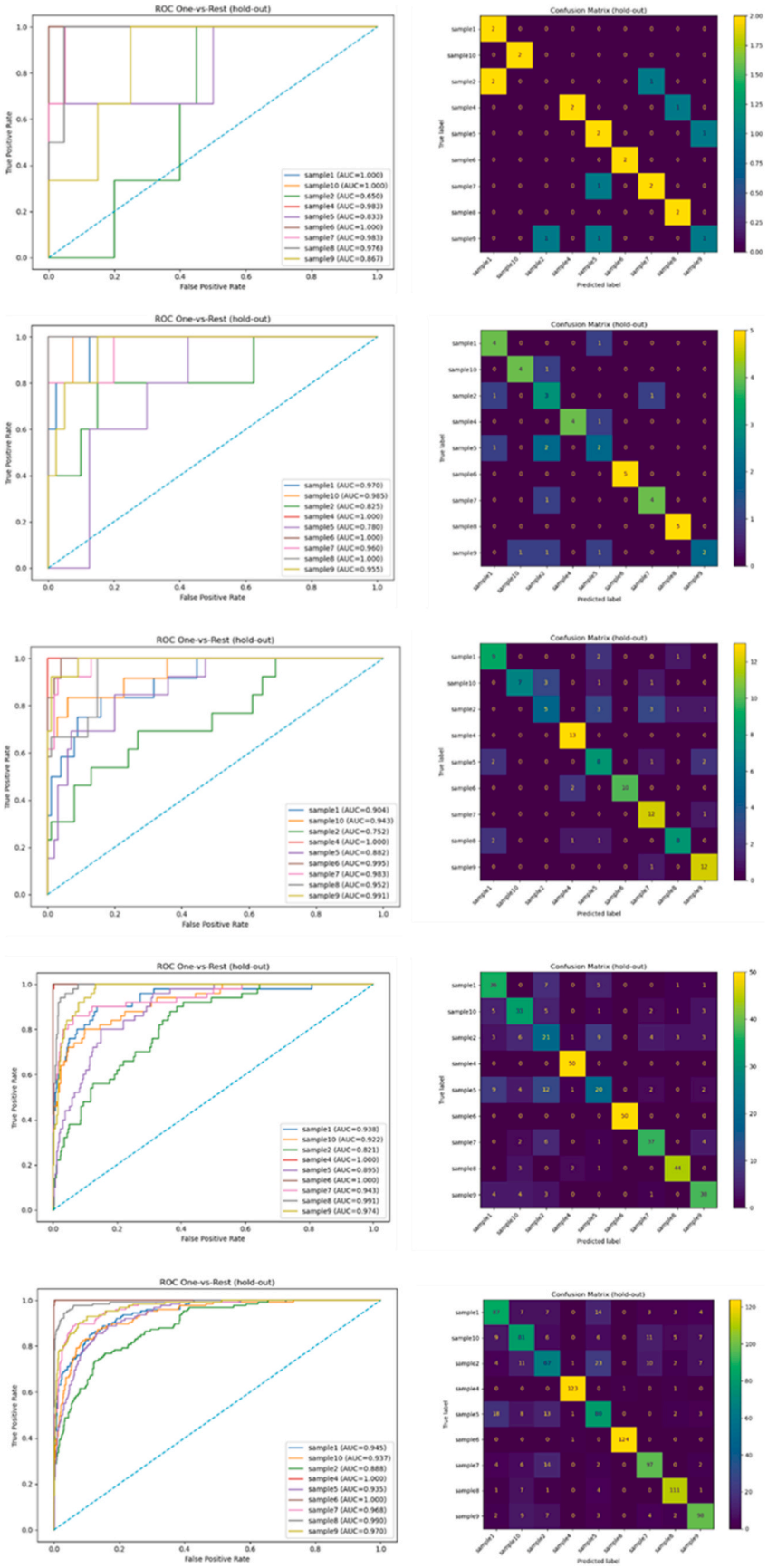


Fig. 10. ROC curves and confusion matrices for the classification of Raman spectra across ten ink types using an SVM-RBF model.

be activated. A deconvolution of the peaks with pseudo-Voigt profile expressed by equation (6) assures that all contribution of homogeneous and inhomogeneous broadening is considered into the analysis.

The normalized pseudo-Voigt formula is a linear combination of a normalized Gaussian and a normalized Lorentzian function. The specific formula used is

$$y = y_0 + A\left[\eta\frac{1}{1+\frac{(x-x_c)^2}{w_L^2}} + (1-\eta)e^{\frac{-\ln(2)(x-x_c)^2}{w_G^2}}\right] \quad \text{Equation 6}$$

where y_0 is the offset, x_c the center of the peak, A the area under the curve, η is the shape factor ($\eta = 1$ is Lorentzian and $\eta = 0$ is Gaussian), w_G is the Gaussian full-width at half-maximum (FWHM) and w_L is the Lorentzian full-width at half-maximum (FWHM).

This step corresponds to an effective refinement operation where the peaks position, peaks area and FWHM are optimized to well describe each spectrum. Following this procedure, a real scoring operation can be attributed to each result, and a complete parametric comparison can be obtained. Fig. 8 reports an example of results carried out using 8 cm^{-1} as confidence in the peak position.

With the help of this specific algorithm of analysis a neural net of AI can be trained to optimize the calculation time and compare a very large number of targets.

Noticeable differences occur in the similarity percentages obtained with the comparative method and those produced by the automated analysis (Fig. 9). These discrepancies result from the distinct ways in which the spectral data are processed: in the comparative approach, the spectrum is divided into six separate regions that are individually evaluated and subsequently averaged to produce a final similarity value, whereas in the automated system the spectrum can be analyzed in its entirety, allowing the algorithm to exploit all available spectral information at once.

To complement the rule-based chemometric workflow, we implemented a preliminary supervised Machine Learning (ML) framework for the classification of Raman spectra across ten ink types. Each spectrum was transformed into a fixed-length feature vector derived from the most prominent spectral peaks, specifically focusing on position, normalized area, height, and shape parameters, to ensure a physically interpretable data representation. The classification pipeline integrates feature standardization, dimensionality reduction via Principal Component Analysis (PCA), and a Support Vector Machine (SVM) with an RBF kernel, trained and assessed using a stratified hold-out validation scheme. Model performance was evaluated through one-vs-rest ROC curves and confusion matrices, providing objective quantitative metrics(Fig. 10).

Given the limited availability of experimental spectra, a data augmentation procedure was introduced to generate synthetic samples through the controlled addition of noise and statistical variability, allowing for progressive training with larger datasets. To ensure statistical robustness despite the limited initial samples, we employed a stratified hold-out scheme coupled with data augmentation, providing a reliable performance estimation comparable to a Stratified K-Fold Cross-Validation approach. Our results indicate a clear enhancement in classification performance as the training set size increases, with most classes achieving high AUC values and a progressive reduction in misclassifications, even among spectrally similar inks. While these findings demonstrate the feasibility of ML-based discrimination within our proposed pipeline, this study represents an initial step. The current sample size limits generalization, consequently, future work will focus on expanding the experimental dataset and refining augmentation strategies. This integration establishes a foundation for a scalable ML framework suited for large spectral databases.

To assess the forensic reliability of the compatibility scoring approach, an additional statistical validation was carried out based on replicate comparisons across the dataset. In particular, the distributions of compatibility scores obtained from intra-class comparisons (replicates belonging to the same ink formulation) and inter-class comparisons (comparisons between different inks) were analyzed. Intra-class comparisons were generated by comparing replicate spectra obtained from the same ink sample, following the same acquisition and processing pipeline described above. Inter-class comparisons were obtained by comparing spectra belonging to different ink formulations in the dataset. This procedure allowed the construction of two independent statistical populations representing the expected score distributions for matching and non-matching inks. The resulting distributions show a clear separation between the two groups. Intra-class comparisons systematically produce high compatibility scores, reflecting the reproducibility of the Raman spectral features extracted from replicate measurements. On the contrary inter-class comparisons generate significantly lower compatibility values due to the presence of peak position discrepancies, incompatible area ratios, and additional spectral features. Based on these distributions, a decision threshold was defined in order to classify whether two spectra should be considered compatible or not. The threshold value was selected at the point that maximizes the separation between intra-class and inter-class populations while minimizing classification errors. Compatibility scores equal to or above this threshold are therefore classified as matching inks, while scores below the threshold are classified as non-matching inks. Using this criterion, the performance of the compatibility scoring system was evaluated through a confusion matrix derived from all replicate comparisons in the dataset. The confusion matrix reports the number of: True Positives (TP: replicate spectra of the same ink correctly classified as compatible); False Positives (FP: spectra from different inks incorrectly classified as compatible); False Negatives (FN: replicate spectra of the same ink incorrectly classified as incompatible). From the confusion matrix (reported in Fig. 10) obtained with the highest numbers of samples (1125) we extracted the values of the TP (868), FN (257), FP (257) and the accuracy (77.16%). The results demonstrate that the compatibility scoring framework provides a reliable discrimination between inks belonging to different formulations while maintaining a high level of reproducibility for replicate measurements of the same ink. Overall, this validation confirms that the compatibility-based comparison strategy, when supported by standardized spectral preprocessing and peak deconvolution, provides a statistically robust tool for the discrimination of writing inks. The approach can therefore be extended to larger spectral libraries, where the combination of compatibility scoring and automated analysis algorithms may enable rapid screening and classification of large numbers of ink samples.

6. Conclusions

This study presents a comparative Raman-based methodology for the identification and classification of modern black ink, with potential applications in both cultural heritage diagnostics and forensic science. By combining peak-position analysis, normalized area ratios, and a scoring strategy grounded in instrumental uncertainty, the proposed approach enables a quantitative assessment of spectral compatibility between inks. The method is effective in discriminating against formulations even within the same manufacturer highlighting compositional differences through the detection of additional bands or non-matching peak ratios.

The results demonstrate that Raman spectroscopy, when supported by systematic deconvolution procedures, can distinguish between inks that share a common pigment, such as crystal violet, while also identifying differences related to binders, additives, or manufacturing processes. The methodology further highlights the importance of standardized data acquisition, baseline removal, and the careful exclusion of unreliable observables, all of which are essential for achieving reproducible and comparable compatibility scores.

In this context, it must be highlighted that certain ink formulations may not be addressable under the specific excitation and instrument

configuration chosen, as demonstrated by the sample that yielded detector saturation. Such cases are categorized as 'unclassifiable' within the standardized framework; since the methodology requires identical experimental conditions for a valid comparison, a divergent optical response (luminescence or thermal instability) is itself a diagnostic indicator of a different chemical nature or the presence of specific fluorophores, inherently excluding compatibility with the reference group.

While the present study utilizes a standardized substrate to establish a high-resolution reference framework, future research will be directed towards testing the impact of environmental degradation (such as photodegradation and oxidation) and the presence of paper fillers or coatings. These subsequent investigations will explore how UV exposure and thermal stress influence the spectral shifts and area ratios, further expanding the robustness of the database for real-world forensic and cultural heritage applications.

The implementation of an automated algorithm provides an additional and complementary pathway for large-scale, high-quantity spectral comparison. Unlike the comparative method, where spectra are divided into six sections and averaged, the automated system evaluates the entire spectral profile, thereby exploiting a greater amount of information and yielding similarity values that may differ from those obtained manually. These discrepancies reflect intrinsic differences in data handling and illustrate the potential of automated approaches for refining ink classification, especially in the context of extensive spectral databases.

Overall, the findings confirm that the presented methodology is suitable for the discrimination of ink formulations and for the identification of unknown samples based on a quantitative compatibility index. With further extension to a broader dataset and integration into a structured spectral database, this approach could work as a powerful tool for authentication, provenance studies, and forensic investigations. Moreover, the automated analysis developed in this work provides a foundation for future implementation of machine learning or neural-network models capable of performing rapid comparisons across thousands of spectra, enhancing the speed, reliability, and scalability of ink identification in real-world applications.

CRediT authorship contribution statement

Vanessa Pinna: Methodology, Investigation, Formal analysis, Data curation. **Stefania Porcu:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Formal analysis, Conceptualization. **Gianluca Siotto:** Methodology, Formal analysis. **Enrica Tuveri:** Visualization, Methodology, Investigation. **Pier Carlo Ricci:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Edoardo Lodo:** Formal analysis. **Pietro Coli:** Visualization, Methodology. **Roberto Cardia:** Visualization, Methodology. **Daniele Chiriu:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Founding

This research was funded by Fondazione di Sardegna, project FDS2022 “New diagnostic techniques for ancient books restauration and conservation” CUP F73C23001560007.

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

All the authors acknowledge the Scientific Investigation Department (RIS) of Cagliari.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.talanta.2026.129678.

Data availability

The code used for the automatized data analysis is available at https://github.com/Lodo808/Spectrum-identifier.git

References

[1] T. Christiansen, M. Cotte, W. de Nolf, E. Mouro, J. Reyes-Herrera, S. de Meyer, F. Vanmeert, N. Salvadó, V. Gonzalez, P.E. Lindelof, K. Mortensen, K. Ryholt, K. Janssens, S. Larsen, Insights into the composition of ancient Egyptian red and black inks on papyri achieved by synchrotron-based microanalyses, Proc. Natl. Acad. Sci. USA 117 (2020) 27825–27835, https://doi.org/10.1073/pnas.2004534117.

[2] N. Teixeira, P. Nabais, V. Otero, R.J. Díaz Hidalgo, M. Ferretti, M. Licchelli, M. J. Melo, A study on the degradation of iron gall inks and to preserve them using green approaches, Heritage 8 (2025) 1–21, https://doi.org/10.3390/heritage8070261.

[3] T. Ghigo, A Systematic Scientific Study of Coptic Inks from the Late Roman Period to the Middle Ages Doctoral Dissertation, 2020.

[4] N. Ferrer, M. Carme Sistach, Analysis of sediments on iron gall inks in manuscripts, Restaurator 34 (2013) 175–193, https://doi.org/10.1515/res-2013-0010.

[5] R. Cieśla, Forensic examination of inks used as inscription on historical documents, Int. J. Conserv. Sci. 14 (2023) 481–496, https://doi.org/10.36868/IJCS.2023.02.07.

[6] C. Neumann, R. Ramotowski, T. Genessay, Forensic examination of ink by high-performance thin layer chromatography-the United States secret service digital ink library, J. Chromatogr. A 1218 (2011) 2793–2811, https://doi.org/10.1016/j.chroma.2010.12.070.

[7] A. Koochakzaei, T. Ghaffari, Identification of traditional black Persian inks by spectroscopic and spectral imaging techniques: presenting a flowchart method, Vib. Spectrosc. 127 (2023) 103545, https://doi.org/10.1016/j.vibspec.2023.103545.

[8] A.B. López-Baldomero, M. Buzzelli, F. Moronta-Montero, M.Á. Martínez-Domingo, E.M. Valero, Ink classification in historical documents using hyperspectral imaging and machine learning methods, Spectrochim. Acta Mol. Biomol. Spectrosc. 335 (2025) 125916, https://doi.org/10.1016/j.saa.2025.125916.

[9] A. Gambaro, R. Ganzerla, M. Fantin, E. Cappelletto, R. Piazza, W.R.L. Cairns, Study of 19th century inks from archives in the Palazzo Ducale (Venice, Italy) using various analytical techniques, Microchem. J. 91 (2009) 202–208, https://doi.org/10.1016/j.microc.2008.11.002.

[10] T. Ghigo, I. Rabin, P. Buzi, Black Egyptian inks in late antiquity: new insights on their manufacture and use, Archaeol. Anthropol. Sci. 12 (2020) 1–14, https://doi.org/10.1007/s12520-019-00977-3.

[11] D. Tamburini, F. Sabatini, S. Berbers, M.R. van Bommel, I. Degano, An introduction and recent advances in the analytical study of early synthetic dyes and organic pigments in cultural heritage, Heritage 7 (2024) 1969–2010, https://doi.org/10.3390/heritage7040094.

[12] Z. Khan, F. Shafait, A. Mian, Automatic ink mismatch detection for forensic document analysis, Pattern Recogn. 48 (2015) 3615–3626, https://doi.org/10.1016/j.patcog.2015.04.008.

[13] M. Suzuki, N. Akiba, K. Kurosawa, Y. Akao, Y. Higashikawa, Differentiation of black writing ink on paper using luminescence lifetime by time-resolved luminescence spectroscopy, Forensic Sci. Int. 279 (2017) 268–274, https://doi.org/10.1016/j.forsciint.2017.09.003.

[14] V.N. Aginsky, Natural aging of ink and ink fading are different physical and chemical processes: a document dating case report, Journal of the American Society of Questioned Document Examiners 26 (2025) 1–25, https://doi.org/10.69525/jasqde.292.

[15] Q. Sun, Y. Luo, P. Xiang, X. Yang, M. Shen, Analysis of PEG oligomers in black gel inks: discrimination and ink dating, Forensic Sci. Int. 277 (2017) 1–9, https://doi.org/10.1016/j.forsciint.2017.04.022.

[16] L. Monico, K. Janssens, M. Cotte, S.M. Webb, F. Vanmeert, V. Gonzalez, G. Van der Snickt, K. Keune, B.G. Brunetti, G. Falkenberg, J. Garrevoet, A. van Loon, M. Vermeulen, M. Maguregui, S. De Meyer, E.A. Clerici, F.T.H. Broers, L. Cartechini, N. De Keyser, F. d'Acapito, F. Meirer, A. Romani, F. Rosi, S.C. Marri, D. Comelli, N. Deleu, I. Fazlic, M. Ghirardello, C. Holé, S. Pérez-Diez, M. Thoury, C. Miliani, Advanced X-ray techniques to study the alteration of pigments in paintings, La Rivista Del Nuovo Cimento 48 (2025) 315–434, https://doi.org/10.1007/s40766-025-00070-7.

[17] A.B. López-Baldomero, M.Á. Martínez-Domingo, S. George, E.M. Valero, Spectral unmixing as a preprocessing step for SVM-based material identification in historical manuscripts, Npj Heritage Science 13 (2025) 1–9, https://doi.org/10.1038/s40494-025-02029-7.

[18] P. Piersigilli, R. Citroni, F. Mangini, F. Frezza, Electromagnetic techniques applied to cultural heritage diagnosis: state of the art and future prospective: a

comprehensive review, Appl. Sci. (Switzerland) 15 (2025) 1–51, https://doi.org/10.3390/app15126402.
[19] M. Botticelli, V. Risdonne, T. Visser, C. Young, M.J. Smith, J.M. Charsley, M. Rutkauskas, Y. Altmann, D.T. Reid, Reflecting the past, imag(in)ing the past: macro-reflection imaging of painting materials by fast MIR hyperspectral analysis, Eur. Phys. J. Plus 138 (2023), https://doi.org/10.1140/epjp/s13360-023-03958-7.
[20] V.Ó.A.G. da Silva, M. Talhavini, I.C.F. Peixoto, J.J. Zacca, A.O. Maldaner, J.W. B. Braga, Non-destructive identification of different types and brands of blue pen inks in cursive handwriting by visible spectroscopy and PLS-DA for forensic analysis, Microchem. J. 116 (2014) 235–243, https://doi.org/10.1016/j.microc.2014.05.013.
[21] L. Cséfalvayová, M. Strlič, H. Karjalainen, Quantitative NIR chemical imaging in heritage science, Anal. Chem. 83 (2011) 5101–5106, https://doi.org/10.1021/ac200986p.
[22] A.B. López-Baldomero, M. Buzzelli, F. Moronta-Montero, M.Á. Martínez-Domingo, E.M. Valero, Ink classification in historical documents using hyperspectral imaging and machine learning methods, Spectrochim. Acta Mol. Biomol. Spectrosc. 335 (2025), https://doi.org/10.1016/j.saa.2025.125916.
[23] F. Cicconi, V. Lazic, A. Palucci, A.C.A. Assis, F.S. Romolo, Forensic analysis of commercial inks by laser-induced breakdown spectroscopy (LIBS), Sensors (Switzerland) 20 (2020) 1–20, https://doi.org/10.3390/s20133744.
[24] M. Oujja, A. Vila, E. Rebollar, J.F. García, M. Castillejo, Identification of inks and structural characterization of contemporary artistic prints by laser-induced breakdown spectroscopy, Spectrochim. Acta Part B At. Spectrosc. 60 (2005) 1140–1148, https://doi.org/10.1016/j.sab.2005.05.021.
[25] Y.W. Hui, N.A. Mahat, D. Ismail, R.K.R. Ibrahim, Laser-induced breakdown spectroscopy (LIBS) for printing ink analysis coupled with principle component analysis (PCA), AIP Conf. Proc. 2155 (2019), https://doi.org/10.1063/1.5125514.
[26] H. de la Codre, M. Radepont, J.P. Echard, O. Belhadj, S. Vaiedelich, V. Rouchon, The use of XRF imaging for the chemical discrimination of iron-gall ink inscriptions: a case study in Stradivari's workshop, X Ray Spectrom. 50 (2021) 244–252, https://doi.org/10.1002/xrs.3160.
[27] C. Ma, H. Dou, Z. Zhao, X. Qiu, H. Li, X. Wang, Review of in-situ non- and micro-destructive techniques for pigment analysis in architectural heritage, Npj Heritage Science 13 (2025), https://doi.org/10.1038/s40494-025-01675-1.
[28] C. Luízar Obregón, M.A. Zamalloa Jara, F.L. Rojas Arizapana, Y.J. Chura Huayllani, J.F. Gonzales Bellido, J. Olivera Olivera, XRF elemental analysis of inks in South American manuscripts from 1779 to 1825, Herit. Sci. 9 (2021) 1–15, https://doi.org/10.1186/s40494-021-00619-9.
[29] M.F. Cheng, A. Mukundan, R. Karmakar, M.A.E. Valappil, J. Jouhar, H.C. Wang, Modern trends and recent applications of hyperspectral imaging: a review, Technologies 13 (2025), https://doi.org/10.3390/technologies13050170.
[30] Y.J. Lee, G.W. Kim, C.W. Jeong, H.T. Kim, M.J. Choi, S.B. Park, T.J. Lee, H.J. Kim, Raman spectroscopy and one-dimensional convolutional neural networks for the forensic identification of red stamp inks in questioned documents, J. Raman Spectrosc. (2025) 1–14, https://doi.org/10.1002/jrs.70018.
[31] P. Buzzini, E. Suzuki, Forensic applications of Raman spectroscopy for the in situ analyses of pigments and dyes in ink and paint evidence, J. Raman Spectrosc. 47 (2016) 16–27, https://doi.org/10.1002/jrs.4818.
[32] D. Chiriu, P.C. Ricci, G. Cappellini, C.M. Carbonaro, Ancient and modern paper: study on ageing and degradation process by means of portable NIR μ-Raman spectroscopy, Microchem. J. 138 (2018) 26–34, https://doi.org/10.1016/j.microc.2017.12.024.
[33] N.M. Grechukha, K.O. Gorshkova, M.S. Panov, I.I. Tumkin, E.O. Kirillova, V. V. Lukianov, N.P. Kirillova, V.A. Kochemirovsky, Analysis of the aging processes of writing ink: raman spectroscopy versus gas chromatography aspects, Appl. Sci. (Switzerland) 7 (2017), https://doi.org/10.3390/app7100991.
[34] R.R. Souza, L.M. Duarte, G.B. Ferreira, Raman spectroscopy for forensic pen ink investigation: a review, Anal. Methods 17 (2025) 2515–2528, https://doi.org/10.1039/d4ay01951b.
[35] P. Colomban, A. Slodczyk, Raman intensity: an important tool to study the structure and phase transitions of amorphous/crystalline materials, Opt. Mater. 31 (2009) 1759–1763, https://doi.org/10.1016/j.optmat.2008.12.030.
[36] U.P. Agarwal, Chapter 9 an Overview of Raman Spectroscopy as Applied to Lignocellulosic Materials, 1986.
[37] L. Polavarapu, A. La Porta, S.M. Novikov, M. Coronado-Puchau, L.M. Liz-Marzán, Pen-on-paper approach toward the design of universal surface enhanced Raman scattering substrates, Small 10 (2014) 3065–3071, https://doi.org/10.1002/smll.201400438.
[38] F.A. Pisu, S. Porcu, R. Carboni, V. Mameli, C. Cannas, S. Naitza, R.T. Melis, M. Mussi, D. Chiriu, Innovative method for provenance studies in cultural heritage: a new algorithm based on observables from high-resolution Raman spectra of red ochre, Spectrochim. Acta Mol. Biomol. Spectrosc. 329 (2025) 125581, https://doi.org/10.1016/j.saa.2024.125581.
[39] S. Saber-Samandari, K. Alamara, S. Saber-Samandari, K.A. Gross, Micro-Raman spectroscopy shows how the coating process affects the characteristics of hydroxylapatite, Acta Biomater. 9 (2013) 9538–9546, https://doi.org/10.1016/j.actbio.2013.08.021.
[40] T.C. David, F.A. Pisu, S. Porcu, C.M. Carbonaro, J. Kodric, D. Chiriu, Effects of saline solutions on paper-based cultural heritage: non-invasive techniques for studying flooded ancient books, Heritage 8 (2025), https://doi.org/10.3390/heritage8020040.
[41] D. Saviello, M. Trabace, A. Alyami, A. Mirabile, P. Baglioni, R. Giorgi, D. Iacopino, Raman spectroscopy and surface enhanced Raman scattering (SERS) for the analysis of blue and black writing inks: identification of dye content and degradation processes, Front. Chem. 7 (2019) 1–9, https://doi.org/10.3389/fchem.2019.00727.
[42] M.V. Cañamares, C. Chenal, R.L. Birke, J.R. Lombardi, DFT, SERS, and single-molecule SERS of crystal violet, J. Phys. Chem. C 112 (2008) 20295–20300, https://doi.org/10.1021/jp807807j.