Journal of Cultural Heritage 75 (2025) 139–146

Contents lists available at ScienceDirect

Journal of Cultural Heritage

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

Original article

An artificial intelligence-based semiquantitative method based on visible spectroscopy and imaging to analyse inorganic red pigments in wall paintings

Roberto Sáez-Hernández[a,*], Jordi Cruz[b,*], Manel Alcalà-Bernàrdez[c], Ángel Morales-Rubio[a], M. Luisa Cervera[a]

[a] Department of Analytical Chemistry, University of Valencia, C/ Dr. Moliner, 50, Burjassot, 46100, Spain
[b] Department of Fundamental Sciences, EUSS School of Engineering, Pg. Sant Joan Bosco, 74, Barcelona, 08017, Spain
[c] Applied Chemometrics Group & ENG4BIO, Department of Chemistry, Universitat Autònoma de Barcelona, Barcelona, 08193, Spain

ARTICLE INFO

Article history:
Received 25 March 2025
Accepted 14 July 2025
Available online 26 July 2025

Keywords:
Artificial neural networks
Colorimetry
Machine learning
Mural paintings
Pigments
Prediction

ABSTRACT

Artificial Intelligence (AI) and Machine Learning (ML) are revolutionizing data analysis by introducing innovative and enhanced methods for data processing. In this article, a chemometric semiquantitative model based on visible spectroscopy and digital image colorimetry was applied to estimate the metal content in inorganic pigments. The model utilized Support Vector Machines (SVM) and Artificial Neural Networks (ANN) regression methods to correlate spectral and colorimetric data with elemental composition. Replicas were prepared and painted with three red inorganic pigments (cinnabar, hematite and minium), and they were analysed using portable X-ray fluorescence, visible reflectance spectroscopy, and digital imaging. Cross-reference between elemental and colorimetric information was performed using Support Vector Regression and Artificial Neural Networks, and the models were validated through Venetian-blinds cross-validation. In the calibration step, Root Mean Square Errors (RMSE) for Fe, Pb, and Hg were 0.03, 3.5, and 3.0 %, respectively, with correlation values (R^2) of 0.99, 0.90, and 0.94. For the prediction set, RMSE was 3.0, 2.6, and 2.3 %, for Fe, Pb, and Hg, respectively, with R^2 of 0.83, 0.92 and 0.81. This article demonstrates that innovative data treatment models, coupled with non-invasive and portable techniques, allow us to estimate the content of elements in inorganic pigments in Cultural Heritage samples.

1. Introduction

1.1. Analytical investigations of mural paintings

The analysis of wall painting composition is a well-established area within chemical and spectroscopic applications to Cultural Heritage. Numerous studies have focused on developing and applying analytical solutions to specific cases. Non-invasive and non-destructive methods are often preferred due to the historical and economic value of cultural samples. Non-invasive and non-destructive methods are often preferred, given that cultural samples are of great historical and economic value. Another desired property of these techniques is portability, as investigations of historical wall paintings are restricted to in situ methodologies. In this context, different techniques arise as the most commonly employed ones.

In terms of elemental composition, X-ray fluorescence (XRF) is widely used as it combines both desired properties: it provides information on the elemental composition of the sample without causing damage to the sample. Additionally, a portable version of this instrumentation allows for in situ investigations [1]. This technique provides a semiquantitative result in a fast way, accounting for many different elements at the same time. There are some drawbacks to consider, most of them deriving from the fact that it is a portable and handheld instrument. One of them is the lack of sensitivity for some elements, which does not allow for quantifying some light elements in the periodic table [2]. However, XRF has been extensively applied to investigate the presence and abun-

Abbreviations: ANN, artificial neural network; ML, machine learning; PCA, principal component analysis; RMSE, root mean square error; RMSECV, root mean square error of cross-validation; RMSEP, root mean square error of prediction; SNV, standard normal variate; SVM, support vector machine; XRF, X-ray fluorescence.

* Corresponding authors.
E-mail addresses: roberto.saez@uv.es (R. Sáez-Hernández), jcruz@euss.cat (J. Cruz).

https://doi.org/10.1016/j.culher.2025.07.018

dance of pigments in historical artifacts, especially for those with high atomic number, like mercury, iron, or lead [3].

The irregular shape of the sample can also become a problem when working with XRF in the field of Cultural Heritage [3], despite this being rarely the case in wall painting investigations. It is also interesting to note that portable XRF investigations are most usually restricted to a point-wise methodology, meaning that quantifications are carried out for one spot at a time. In the last years, XRF mapping and imaging instruments have been developed, allowing for obtaining a spatial distribution of the elemental composition with high spatial resolution [4] or wider and bigger samples [5–7]. These instruments are, however, relatively unaffordable, and some more low-cost solutions have been proposed [8,9].

This technique has been extensively implemented in Cultural Heritage in numerous case studies. For example, this technique can be applied to investigate metallic artefacts [10,11], ancient written manuscripts [12], or remains like amphorae [13]. When it comes to the investigation of pigments, XRF has demonstrated to be appropriate to investigate the nature of preparatory layers, addressing the presence or absence of certain compounds like gypsum in such layers [14], as well as the presence (or absence) of certain elements characteristic of inorganic red pigments [14,15], like Fe or Pb. Green and blue pigments have been also analyzed by XRF, identifying Cu in some frescoes remains [15–17], as well as some other hues like orange [18]. Interestingly, this technique has also proven useful to identify painted areas that have undergone color transformation using elemental information about Pb and As in different spots in the Pompeii archaeological site [19].

Some other analytical techniques widely employed when investigating wall painting samples are Raman, FTIR, or VIS-NIR spectroscopies. These techniques not only provide information about inorganic compounds, but these also allow to analyze organic components. Thus, FTIR has been used to investigate binders that might have been used to adhere the pigment to the wall [20], or degradation products [21,22]. Raman spectroscopy has also been applied to characterize mural paintings, both in terms of mortar preparation, pigment utilization, and restoration steps [23,24]. Visible reflectance spectroscopy is particularly useful for artworks, as it allows for spectroscopic description of surfaces in the visible or near-infrared range. It has been employed to confirm the presence of cinnabar and iron-based pigments in Roman mural paintings [25,26], or green pigments in Pompeii mural paintings [27].

1.2. Machine learning and artificial intelligence in the field of cultural heritage

All the aforementioned traditional techniques have been extensively applied to the field of analytical investigations in Cultural Heritage. However, given the impressive amount of data these techniques can generate, their interpretation can sometimes fall short without proper data treatment techniques.

Chemometrics, Machine Learning (ML), and Artificial Intelligence (AI) are distinct but related fields. Chemometrics applies statistical and mathematical methods to chemical data, aiding in the interpretation of complex chemical systems and improving measurement accuracy. Machine learning, a subset of AI, focuses on developing algorithms that learn from data to make predictions or decisions across various sectors [28]. ML techniques, including supervised, unsupervised, and reinforcement learning, identify data patterns to enhance model performance over time. Artificial intelligence is a broader field aiming to create systems that can perform tasks requiring human intelligence, such as reasoning, learning, and problem-solving [29]. AI encompasses ML, natural language processing, and robotics, with the goal of replicating human cognitive abilities.

While chemometrics specializes in chemical data analysis, ML is domain-agnostic, and AI seeks to build intelligent systems. Both ML and chemometrics utilize various algorithms, broadly categorized into supervised and unsupervised learning. The distinction lies in the presence of prior knowledge in the training data. Unsupervised learning operates without prior knowledge of an object's class or dependent variable value. In contrast, supervised learning aims to create a model that assigns an object to a group or predicts a variable's value. Examples of supervised learning techniques include logistic regression, decision trees, Support Vector Machines (SVM), and Artificial Neural Networks (ANN) [30].

SVM work by finding the optimal hyperplane that best separates the data points of different classes in an N-dimensional space. The goal is to maximize the margin between the classes, which helps in making accurate predictions on new data. There are two SVM classes, the linear ones used when the data is linearly separable, meaning it can be separated by a straight line, and the nonlinear ones when the data is not linearly separable. It uses kernel functions to transform the data into a higher-dimensional space where it becomes separable [31].

ANN is a computational model inspired by how biological neural networks work in the human brain. It consists of interconnected units called neurons, which process information in a manner similar to the brain's neurons. The structure consists of neurons, which are basic units that receive input, process it, and pass it on to other neurons. Neurons are organized into layers: the input layer, which receives the initial data, and the intermediate hidden layers, which process inputs received from the input layer, sends them to the output layer, which produces the final output [32].

The application of ML in Cultural Heritage has spanned from its ability to scan and analyze documents for literature reviewing purposes [33], scanning the conservation state of buildings [34], to pigment identification using XRF and pattern recognition [35]. Raman spectra have been thoroughly analyzed using ML and ANNs in order to create a database containing information about different pigments and binders in Roman and Renaissance frescoes [36]. Similarly, visible reflectance spectra obtained from hyperspectral images were analyzed through ANNs to distinguish pigments in paintings [37].

It can be observed that ML tools have started to gain significance in the field of analytical investigations in Cultural Heritage, given their potential to analyze big datasets and, at the same time, model non-linear relationships among variables. A very recent review has stated "various trends have already emerged in their implementation, the most notable being their assistance in processing and interpreting spectroscopic data for the identification of historic pigments" [38], demonstrating how the application of ML to chemical data for historical pigments research is a growing topic. However, all the information provided so far is limited to qualitative investigations, i.e., to the classification and/or identification of pigments. For example, a model to classify painted spots in black or white has been reported [39] using visible reflectance spectroscopy. Addressing a richer color palette, hyperspectral images in the visible and near infrared regions were used to train convolutional neural networks. These networks were then applied to qualitatively identify pigments in different painted surfaces [40]. A diverse color palette was also created to classify historical inorganic pigments. In this case, RGB images were input to SVM and convolutional neural networks, with satisfactory classification results overall [41].

In this context, this article intends to develop a ML approach based on SVM and ANN to implement a semiquantitative approach to estimate the pigment abundance in mural painting

Table 1
Summary of the different types of samples present in the dataset.

Mixtures	Elements	Proportion	Coding
Cinnabar – hematite	Hg – Fe	0.2 g Hg + 0.6 g Fe	Mix A
		0.6 g Hg + 0.2 g Fe	Mix B
Cinnabar – minium	Hg – Pb	0.2 g Hg + 0.6 g Pb	Mix C
		0.6 g Hg + 0.2 g Pb	Mix D
Hematite – minium	Fe – Pb	0.2 g Fe + 0.6 g Pb	Mix E
		0.6 g Fe + 0.2 g Pb	Mix F

replicas from colorimetric information. To that end, digital imaging and visible reflectance spectroscopy were investigated regarding their predictive potential, expanding the current knowledge present in the Cultural Heritage field regarding quantitative ML applications.

2. Research aim

This study aims to develop and optimize a non-invasive methodology for semiquantitative estimation of inorganic red pigment concentrations in wall paintings. Specifically, it combines non-invasive visible reflectance spectroscopy and smartphone imaging with chemometric tools to predict the concentrations of lead, mercury, and iron in red pigments (minium, cinnabar, and hematite For this purpose, a heterogeneous set of laboratory mockups (in terms of elemental composition of the samples) was prepared, and different colored spots were analyzed, cross-referencing elemental information with spectral data using Support Vector Machines and Artificial Neural Networks.

3. Materials and methods

3.1. Mock-up's preparation

Mockups were prepared following Vitruvius' guidelines, as compiled by Piovesan et al. [42]. Sand and lime were acquired locally at a construction supply store, and sand was manually separated into fine sand and coarse sand by sieving.

A ceramic tile served as the porous surface, with layers of sand-lime mixture applied and allowed to dry between applications. First, a coarse sand-lime layer in a 3:1 proportion was applied, mechanically flattened, and left to dry overnight. Next, on top of that one, a fine sand-lime layer was applied, once again flattened and left to dry. Finally, before applying the pigments, a fine layer of water-suspended lime was applied, on top of which the pigments were painted on the surface.

Three different chemical species were selected for the pigments: cinnabar, minium, and hematite. These chemicals were acquired via Agaragar (https://agaragar.net/), and the Kremer references were as follows: hematite, ref. #48651, minium, ref. 42500, and cinnabar, ref. #10620.

These solid pigments were suspended in water to use them. In this regard, masses ranging from 0.2 to 0.8 g were weighted and suspended in 1 mL of water, shaken, and applied to the wall using a brush. Consequently, different suspensions were prepared with a different amount of each pigment. These values were obtained by trial and error, ensuring a proper suspension that allowed to work and obtain different colour hues and intensities. Additionally, binary mixtures of two different pigments were also prepared in a similar fashion. In total, 0.8 g of pigment was suspended in 1 mL of water. However, in this instance, a 1:3 proportion among the pigments was employed (i.e., 0.2 g of one pigment + 0.6 g of another one). In Table 1, the different types of mixtures are summarized.

In total, the dataset was composed of 102 samples, with the distribution shown in Table 2.

Table 2
Number of samples per pigment.

Sample	Number of samples
Cinnabar	22
Hematite	26
Minium	24
Mix A	5
Mix B	5
Mix C	5
Mix D	5
Mix E	5
Mix F	5
total	102

3.2. Elemental analysis

Elemental composition was measured using a portable XRF spectrometer (S1 Titan analyzer, Bruker). Using the internal calibration software "GeoChem Trace", the elemental profile of each spot was acquired in mass percentage. The instrument uses a Rh X-ray tube (50 kV), and a X-Flash silicon drift detector (with a resolution of 147 eV; FWHM: 5.9 eV), and the measuring spot is 3 mm. Each sample was measured in triplicates in different areas, and the result was averaged to represent the global concentration of each sample.

3.3. Colorimetric analysis

Color was measured using two approaches: smartphone-based imaging and spectrocolorimetry. The colorimetric information obtained through these techniques allowed to objectively determine the color of the samples, as these record superficial information of the wall.

3.3.1. Spectrocolorimetric analysis

The visible reflectance spectra of each sample were captured using a spectrocolorimeter (Konica Minolta CM-26d) in the region of 400–740 nm, with a 10 nm-step resolution. The illuminant, in this case, was a reference D65 from a pulsed xenon lamp, and the instrument was set in the Specular Component Excluded (SCE) mode with a 10° standard observer. Each sample was measured three times in three different regions, and the average spectrum was considered as representative.

3.3.2. Smartphone-based image acquisition

Each tile was photographed under laboratory lighting conditions, in order to acquire images that could be representative of a real-scenario investigation. In this sense, no strict lighting sources was applied. As prior research has stated, one needs to consider some experimental factors when working with smartphones as colour capturing devices. On the one hand, there is inter-device variability, and on the other hand, there is external lighting influence [43].

To tackle these issues, a reference color chart (ColorChecker, 24 colors, 6 gray-scale ones) was included in each photograph. The real CIELAB descriptors of this ColorChecker were obtained using the spectrocolorimeter. By comparing the real CIELAB with the values obtained from the image, linear correction models were created. These models allowed to convert raw CIELAB descriptors (influenced by lighting conditions and device characteristics) into independent CIELAB values. Subsequently, these models were applied to the raw CIELAB values of the samples, converting them into corrected CIELAB. This approach has previously been demonstrated to correctly deal with the color characterization of digital cameras [44–46].

Table 3
Descriptive statistics of the elemental composition of the samples as measured using XRF. Results are expressed in mass percentage.

		Types of samples								
		Cinnabar (n = 22)	*Hematite (n = 26)*	*Minium (n = 24)*	*Mix_A (n = 5)*	*Mix_B (n = 5)*	*Mix_C (n = 5)*	*Mix_D (n = 5)*	*Mix_E (n = 5)*	*Mix_F (n = 5)*
Hg	**Median**	**15**	**0.005**	**0.026**	**8.1**	**13**	**5.0**	**13**	**0.040**	**0.029**
	Min	*2.03*	*0*	*0*	*7.5*	*7.9*	*3.2*	*10.1*	*0.03*	*0.01*
	Max	*60.74*	*0.02*	*0.06*	*9.4*	*16.4*	*6.0*	*16.3*	*0.05*	*0.04*
Fe	**Median**	**0.103**	**8**	**0.126**	**17.9**	**4.0**	**0.130**	**0.098**	**4.0**	**14**
	Min	*0.08*	*0.3*	*0.11*	*17.2*	*3.08*	*0.11*	*0.09*	*2.7*	*11.1*
	Max	*0.13*	*21.9*	*0.18*	*18.7*	*4.32*	*0.15*	*0.10*	*5.3*	*19.3*
Pb	**Median**	**0.1**	**0.00**	**8**	**0.031**	**0.052**	**26**	**6**	**21**	**5.1**
	Min	*0.018*	*0*	*0.18*	*0.03*	*0.03*	*15.0*	*3.0*	*12.0*	*3.4*
	Max	*1.9*	*0.15*	*46.6*	*0.04*	*0.06*	*29.4*	*8.7*	*31.4*	*7.4*

3.4. Data treatment: preprocessing, model creation and validation

Chemometric analyses were conducted using Solo, version 9.3 (Eigenvector Research, Inc., Wenatchee, WA, USA). Regarding the regression models using spectral data, various pre-processing methods such as Mean Centering, Standard Normal Variate, and Savityzky-Golay Derivatives were evaluated to develop the selected prediction models. For the models created using CIELAB descriptors, all models were performed using data autoscaling.

Quantitative models were created using 67 % of the samples (68) for calibration and 33 % (34) for validation, including samples from all different types. Sample selection was performed using the Kennard-Stone algorithm [47]. Data was column-autoscaled before applying the algorithms.

Several chemometric models were developed using Support Vector Machine (SVM) regression models using radial basis functions and Artificial Neural Network (ANN) regression models using the back-propagation neural network algorithm.

The constructed models were internally validated using the Venetian blinds (20 %) cross-validation method. Model parameters such as the R^2 coefficient, the root mean squared error of calibration (RMSEC), and the root mean squared error of cross-validation (RMSECV) were considered for selecting the best model. The prediction capacity of the models was evaluated using various statistical parameters, including the root mean squared error of prediction (RMSEP) for the validation set samples.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-1}} \quad (1)$$

Where yi is the reference value of the sample i, ŷi is the predicted value of the sample i, and n is the number of samples in the set (calibration, cross-validation or prediction).

The optimal model selection was achieved considering the highest R^2 between predicted and reference values, the lowest RMSE (Eq. (1)), and the lowest bias (Eq. (2)), for calibration, cross-validation and prediction.

$$\mathrm{Bias} = \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)}{n} \quad (2)$$

4. Results and discussion

4.1. Elemental composition of the samples

The elemental composition of the 102 samples is summarized in Table 3. Results demonstrate that the replicas' preparation allowed to accomplish the main objective of that part: to obtain heterogeneity in terms of elemental composition of the samples. Thus, these elemental composition values were then employed to build the predictive regression models that will be explained in the following sections.

4.2. Colorimetric description of the samples and quality of the camera characterization

The quality of smartphone-based color characterization was validated by comparing CIELAB values with those obtained from the spectrocolorimeter. To compare reference (spectrocolorimeter) CIELAB values with smartphone imaging CIELAB ones (obtained as described in 3.3.2), the CIEDE2000 values were computed. These formulae represent the reference approach to compare two sets of colors. Thus, the real CIELAB values of each sample were compared with the ones obtained by imaging, obtaining a CIEDE2000 value. The final result for the whole dataset for CIEDE2000 was (4.8, 5.9) units, expressed as the confidence interval at 95 % probability. This range is in good accordance with values previously presented in the literature addressing color difference in Cultural Heritage [48–50]. Additionally, it should be noticed that these CIEDE2000 values have been obtained through a very simple and straightforward camera characterization process, providing a very convenient balance between methodological complexity and the accuracy of the results.

In terms of chromatic diversity, Table 4 represents how the different elemental compositions (Table 3) reflect on the final color as registered by the images.

As regards to reflectance spectra of the samples, Fig. 1 shows the average spectra for each sample type (i.e., per kind of pigment, either pure or mixture).

4.3. Regression methods to estimate the elemental composition

Using the information shown in Sections 4.1 and 4.2, two different Machine Learning techniques were employed to estimate the elemental profile in the samples: Support Vector Machines (SVM) and Artificial Neural Networks (ANN). These models were created using the colorimetric information (either the spectra or the color descriptors) as predictors of the response variable (elemental content of all three elements: Fe, Pb, and Hg).

In each case, the quality of the models was assessed in terms of the error of prediction of the calibration, cross-validation and prediction steps, as described in Section 3.4.

4.3.1. Prediction using digital imaging color values

Three different models were trained using the digital color values as predictors and the elemental contents (in mass percentage), one for each element. Results are synthesized in Table 5.

The results demonstrate how the ML methodologies' ability to capture and represent non-linear relationships allows them to predict metallic content in the samples. The results showed outstanding correlation values (R2) for calibration, cross-validation, and prediction sets, with RMSEP below 4.0 % for the prediction sets. Additionally, no relevant bias is observed in any case.

Table 4
Descriptive statistics of the color parameters of the samples. Results of the confidence interval are expressed for 95 % probability.

		Types of samples								
		Cinnabar (n = 22)	*Hematite (n = 26)*	*Minium (n = 24)*	*Mix_A (n = 5)*	*Mix_B (n = 5)*	*Mix_C (n = 5)*	*Mix_D (n = 5)*	*Mix_E (n = 5)*	*Mix_F (n = 5)*
L*	Cl_{95}	(57, 66)	(40, 51)	(64, 71)	(33, 35)	(36, 38)	(58, 60)	(54, 57)	(43, 46)	(33, 35)
	min	48.9	32.7	55.2	32.4	35.5	57.3	52.8	42.5	33.5
	max	88.3	79.8	88.0	34.8	37.8	60.2	57.2	47.4	35.9
a*	Cl_{95}	(24, 33)	(20, 25)	(23, 32)	(19, 22)	(26, 27)	(45, 48)	(45, 47)	(34, 35)	(22.6, 23.4)
	min	4.5	0.6	5.8	17.7	25.9	44.2	45.0	33.4	22.4
	max	40.8	47.0	28.1	22.1	27.5	47.6	47.6	35.7	23.5
b*	Cl_{95}	(18, 25)	(18, 21)	(36, 46)	(17, 20)	(23, 24)	(49, 54)	(42, 45)	(34, 36)	(21, 22)
	min	4.4	8.9	5.7	15.5	22.7	46.9	41.2	33.9	21.4
	max	31.3	26.1	55.3	19.9	24.5	52.8	44.5	35.7	21.9

Table 5
Prediction parameters of the ML models using color descriptors.

Element	Parameter	Algorithm	Step		
			Calibration	Cross-validation	Prediction
Hg	***RMSE***	SVM	4.54	6.57	4.03
	R^2		0.85	0.65	0.57
	Bias		-0.73	-0.55	0.13
Fe	***RMSE***	SVM	0.80	1.39	2.56
	R^2		0.98	0.93	0.89
	Bias		0.04	0.001	-0.20
Pb	***RMSE***	SVM	2.61	3.58	3.72
	R^2		0.93	0.84	0.88
	Bias		-0.33	0.01	0.25

*RMSE and Bias are expressed as mass percentage.

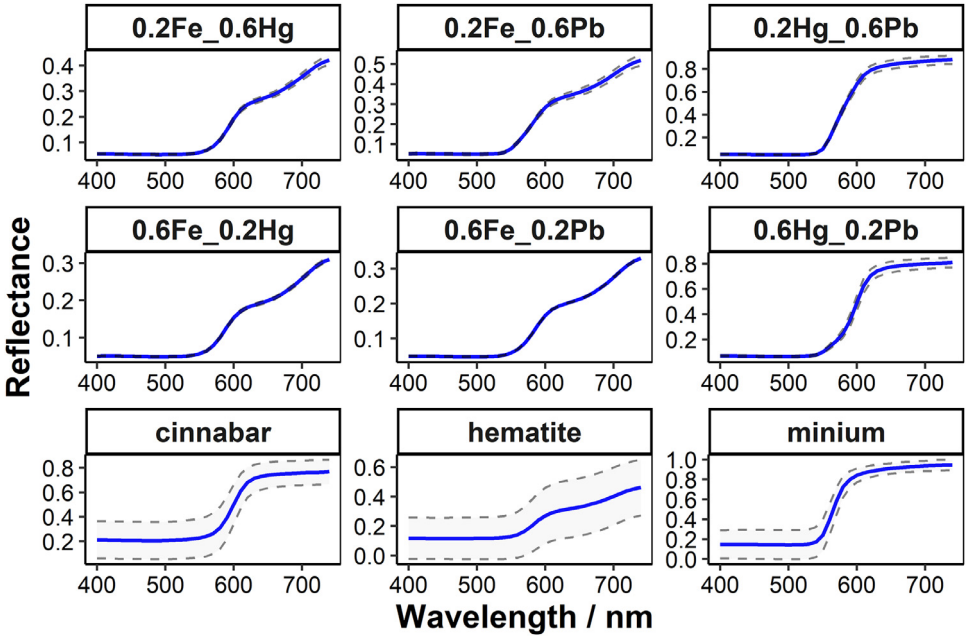

Fig. 1. Reflectance spectra for each kind of pigment. In blue, the average spectra. In light grey, the standard deviation.

When it comes to the prediction using spectral data, results are summarized in Table 6.

In terms of prediction error, it can be observed that Fe is predicted with very good performance using both approaches: imaging and the spectral fingerprint. Regarding the prediction of Hg, results are improved when spectral data is used in comparison to digital imaging. This improvement is also observed for Pb, with increased correlation values (R^2) and lower bias and RMSE for the prediction set.

The improvement observed when applying the spectral data compared to the imaging procedure is most probably because digital imaging devices, like smartphone cameras, work by condensing spectral information into three unique channels [51,52]. In this regard, all the spectral features that are originally present in the spectra shown in Fig. 1 are mathematically condensed into three channels by integration in specific zones of the visible range. Thus, the total amount of information is significantly reduced from 34 variables (range from 400 to 740 nm, in 10 nm steps) to only three variables (the original RGB coordinates that are subsequently transformed).

A Principal Component Analysis (PCA) of the spectral data (SNV, mean centered), shows some deeper insights into these results.

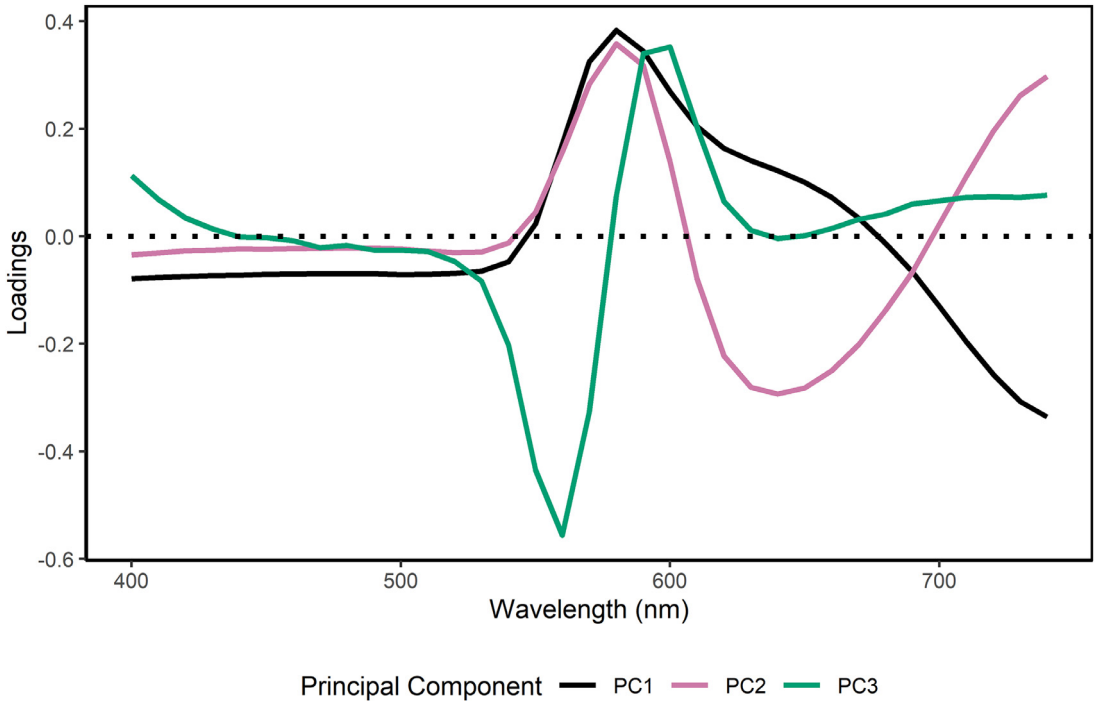

Fig. 2. Loadings for the three main principal components of the spectral dataset.

Table 6
Prediction parameters of the ML models using spectral data.

Element	Parameter	Algorithm	Step		
			Calibration	Cross-validation	Prediction
Hg	***RMSE***	SVM	2.95	4.97	2.30
	R^2		0.94	0.80	0.81
	Bias		-0.11	-0.28	0.80
Fe	***RMSE***	SVM	0.03	0.95	2.99
	R^2		0.99	0.94	0.83
	Bias		0.005	0.04	0.17
Pb	***RMSE***	ANN	3.50	5.90	2.63
	R^2		0.88	0.66	0.92
	Bias		-0.09	-0.89	-1.02

*RMSE and Bias are expressed as mass percentage.

First, the three first principal components reach a total of 99.2 % variance explained (72.1 % for PC1, 98.0 % accumulated variance explained if PC2 is considered). This result demonstrates how most of the significant information available in the visible range can be condensed into three channels with a minimal loss of information (in this case, 0.8 % of the total variance of the dataset). Second, the loadings of the three main principal components allow to obtain Fig. 2.

Most of the information is comprised from approximately 530 nm onwards, as the spectra shown in Fig. 1 do not present distinctive features before that point.

4.4. Potential applications and model usability

In spectral terms, the predictive capability of the developed models is constrained to the spectral variability encompassed within the training dataset. This implies that only visible reflectance spectra falling within the modeled domain can be reliably predicted in both spectrally and by using CIELAB data from mobile phone images, where predictions are valid only within the colorimetric range captured during model development.

Nevertheless, due to the robust design of experiments implemented in this study, the models demonstrate the ability to semi-quantitatively predict the presence of red pigments, specifically cinnabar, hematite, and minium, in wall painting samples. These results highlight the potential of the methodology for practical applications in cultural heritage studies.

The performed models could also be employed by other researchers, provided that the red pigments under investigation correspond to those included in the training set. To ensure compatibility of the mobile images, it is essential to normalize them prior to the CIELAB transformation step. This preprocessing step mitigates variability introduced by differing lighting conditions and device-specific color responses, thereby aligning the input data with the conditions under which the models were developed.

Regarding the use of data from other colorimeters, it is something more critical, because of instrumental differences, but these differences can be overcome by doing data transformation. When these conditions are met, the models offer a practical and accessible tool for the semiquantitative analysis of red pigments in mural paintings using mobile imaging technologies.

5. Conclusions

This study successfully demonstrates the potential of combining non-invasive visible reflectance spectroscopy with chemometric tools to estimate inorganic pigment concentrations in wall paintings. SVM and ANN provided reliable semiquantitative predictions of lead, mercury, and iron concentrations in red pigments.

The non-invasive nature of this methodology ensures the preservation of cultural heritage artifacts while providing essential compositional data. SVM and ANN proved effective in correlating spectral data with elemental concentrations, highlighting their utility in the field of art conservation. This approach can be applied to in-situ analysis of historical wall paintings, offering a valuable tool for conservators and researchers.

Overall, the study provides a robust framework for the non-invasive analysis of inorganic pigments, paving the way for further

advancements in preserving and understanding cultural heritage. This methodology not only enhances the accuracy of pigment analysis but also ensures the integrity and longevity of precious artworks.

Declaration of generative AI

During the preparation of this work the authors used Copilot in order to improve readability and language. After using this tool/service, the authors reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Acknowledgments

Authors thank the Valencian Regional Government (Conselleria de Educación, Cultura, Universidades y Empleo) for the funding (CIPROM/2023/37).

References

[1] K. Janssens, G. Van der Snickt, F. Vanmeert, S. Legrand, G. Nuyts, M. Alfeld, L. Monico, W. Anaf, W. De Nolf, M. Vermeulen, J. Verbeeck, K. De Wael, Non-invasive and non-destructive examination of artistic pigments, paints, and paintings by means of X-ray methods, in: R. Mazzeo (Ed.), Analytical Chemistry for Cultural Heritage, Springer International Publishing, Cham, 2017, pp. 77–128, doi:10.1007/978-3-319-52804-5_3.

[2] L. Bonizzoni, ED-XRF analysis for cultural heritage: is quantitative evaluation always essential? J. Phys. Conf. Ser. 630 (2015) 012001, doi:10.1088/1742-6596/630/1/012001.

[3] A. Galli, L. Bonizzoni, True versus forged in the cultural heritage materials: the role of PXRF analysis, X-Ray Spectrom. 43 (2014) 22–28, doi:10.1002/xrs.2461.

[4] K. Keune, J. Mass, F. Meirer, C. Pottasch, A. van Loon, A. Hull, J. Church, E. Pouyet, M. Cotte, A. Mehta, Tracking the transformation and transport of arsenic sulfide pigments in paints: synchrotron-based X-ray micro-analyses, J. Anal. At. Spectrom. 30 (2015) 813–827, doi:10.1039/C4JA00424H.

[5] M. Alfeld, MA-XRF for historical paintings: state of the art and perspective, Microsc. Microanal. 26 (2020) 72–75, doi:10.1017/S1431927620013288.

[6] M. Alfeld, J.V. Pedroso, M. van E. Hommes, G.V. der Snickt, G. Tauber, J. Blaas, M. Haschke, K. Erler, J. Dik, K. Janssens, A mobile instrument for in situ scanning macro-XRF investigation of historical paintings, J. Anal. At. Spectrom. 28 (2013) 760–767, doi:10.1039/C3JA30341A.

[7] G. Sciutto, T. Frizzi, E. Catelli, N. Aresi, S. Prati, R. Alberti, R. Mazzeo, From macro to micro: an advanced macro X-ray fluorescence (MA-XRF) imaging approach for the study of painted surfaces, Microchem. J. 137 (2018) 277–284, doi:10.1016/j.microc.2017.11.003.

[8] P.H.O.V. Campos, C.R. Appoloni, M.A. Rizzutto, A.R. Leite, R.F. Assis, H.C. Santos, T.F. Silva, C.L. Rodrigues, M.H. Tabacniks, N. Added, A low-cost portable system for elemental mapping by XRF aiming in situ analyses, Appl. Radiat. Isot. 152 (2019) 78–85, doi:10.1016/j.apradiso.2019.06.018.

[9] P. Colomban, M. Gerken, M. Gironda, V. Mesqui, On-site micro-XRF mapping of enameled porcelain paintings and sculpture. First demonstration, J. Eur. Ceram. Soc. 45 (2025) 116849, doi:10.1016/j.jeurceramsoc.2024.116849.

[10] S. Mattiello, O. De Pascale, V. Palleschi, G. Fiorentino, G.S. Senesi, Application of handheld/portable spectroscopic tools to the identification, inner stratigraphy and mapping of archaeological metal artefacts, J. Phys. Photonics 6 (2024) 035005, doi:10.1088/2515-7647/ad46a7.

[11] P. Silveira, T. Falcade, Applications of energy dispersive X-ray fluorescence technique in metallic cultural heritage studies, J. Cult. Herit. 57 (2022) 243–255, doi:10.1016/j.culher.2022.09.008.

[12] F. Pozzi, A. Rizzo, E. Basso, E.M. Angelin, S.F. de Sá, C. Cucci, M. Picollo, Portable spectroscopy for cultural heritage, in: Portable Spectroscopy and Spectrometry, John Wiley & Sons, Ltd, 2021, pp. 499–522, doi:10.1002/9781119636489.ch43.

[13] M. Ramacciotti, G. Gallello, D. Navarro-Martos, A. Doménech-Carbó, C. Roldán, E. Hernández, S. Garrigues, A. Pastor, An innovative multi-analytical approach based on spectroscopic and electrochemical techniques to study a complex Roman amphorae collection, Appl. Clay Sci. 198 (2020) 105857, doi:10.1016/J.CLAY.2020.105857.

[14] S. Almaviva, R. Fantoni, F. Colao, A. Puiu, F. Bisconti, V. Fiocchi Nicolai, M. Romani, S. Cascioli, S. Bellagamba, LIF/Raman/XRF non-invasive microanalysis of frescoes from St. Alexander catacombs in Rome, Spectrochim. Acta. A Mol. Biomol. Spectrosc. 201 (2018) 207–215, doi:10.1016/j.saa.2018.04.062.

[15] V. Crupi, B. Fazio, G. Fiocco, G. Galli, M.F. La Russa, M. Licchelli, D. Majolino, M. Malagodi, M. Ricca, S.A. Ruffolo, V. Venuti, Multi-analytical study of Roman frescoes from Villa dei Quintili (Rome, Italy), J. Archaeol. Sci. Rep. 21 (2018) 422–432, doi:10.1016/j.jasrep.2018.08.028.

[16] A. Paradisi, A. Sodo, D. Artioli, A. Botti, D. Cavezzali, A. Giovagnoli, C. Polidoro, M.A. Ricci, Domus Aurea, the 'Sala Delle Maschere': chemical and spectroscopic investigations on the Fresco paintings, Archaeometry 54 (2012) 1060–1075, doi:10.1111/j.1475-4754.2012.00678.x.

[17] S. Pagès-Camagna, S. Colinart, The Egyptian green pigment: its manufacturing process and links to Egyptian blue, Archaeometry 45 (2003) 637–658, doi:10.1046/j.1475-4754.2003.00134.x.

[18] O. Hahn, D. Oltrogge, H. Bevers, Coloured prints of the 16th century: non-destructive analyses on coloured engravings from Albrecht DÜrer and contemporary artists, Archaeometry 46 (2004) 273–282, doi:10.1111/j.1475-4754.2004.00157.x.

[19] I. Marcaida, M. Maguregui, S. Fdez-Ortiz de Vallejuelo, H. Morillas, N. Prieto-Taboada, M. Veneranda, K. Castro, J.M. Madariaga, In situ X-ray fluorescence-based method to differentiate among red ochre pigments and yellow ochre pigments thermally transformed to red pigments of wall paintings from Pompeii, Anal. Bioanal. Chem. 409 (2017) 3853–3860, doi:10.1007/s00216-017-0329-3.

[20] N. Vornicu, C. Bibire, E. Murariu, D. Ivanov, Analysis of mural paintings using in situ non-invasive XRF, FTIR spectroscopy and optical microscopy, X-Ray Spectrom. 42 (2013) 380–387, doi:10.1002/xrs.2459.

[21] A. Nevin, J.L. Melia, I. Osticioli, G. Gautier, M.P. Colombini, The identification of copper oxalates in a 16th century Cypriot exterior wall painting using micro FTIR, micro Raman spectroscopy and gas chromatography-mass spectrometry, J. Cult. Herit. 9 (2008) 154–161, doi:10.1016/j.culher.2007.10.002.

[22] B. Salvadori, V. Errico, M. Mauro, E. Melnik, L. Dei, Evaluation of gypsum and calcium oxalates in deteriorated mural paintings by quantitative FTIR spectroscopy, Spectrosc. Lett. 36 (2003) 501–513, doi:10.1081/SL-120026615.

[23] M. Veneranda, M. Irazola, A. Pitarch, M. Olivares, A. Iturregui, K. Castro, J.M. Madariaga, In-situ and laboratory Raman analysis in the field of cultural heritage: the case of a mural painting, J. Raman Spectrosc. 45 (2014) 228–237, doi:10.1002/jrs.4448.

[24] M. Maguregui, U. Knuutinen, I. Martínez-Arkarazo, A. Giakoumaki, K. Castro, J.M. Madariaga, Field Raman analysis to diagnose the conservation state of excavated walls and wall paintings in the archaeological site of Pompeii (Italy), J. Raman Spectrosc. 43 (2012) 1747–1753, doi:10.1002/jrs.4109.

[25] A. Dal Fovo, A. Mazzinghi, S. Omarini, E. Pampaloni, C. Ruberto, J. Striova, R. Fontana, Non-invasive mapping methods for pigments analysis of Roman mural paintings, J. Cult. Herit. 43 (2020) 311–318, doi:10.1016/j.culher.2019.12.002.

[26] V. Guglielmi, M. Andreoli, V. Comite, A. Baroni, P. Fermo, The combined use of SEM-EDX, Raman, ATR-FTIR and visible reflectance techniques for the characterisation of Roman wall painting pigments from Monte d'Oro area (Rome): an insight into red, yellow and pink shades, Environ. Sci. Pollut. Res. 29 (2022) 29419–29437, doi:10.1007/s11356-021-15085-w.

[27] C. Cucci, M. Picollo, L. Chiarantini, G. Uda, L. Fiori, B. De Nigris, M. Osanna, Remote-sensing hyperspectral imaging for applications in archaeological areas: non-invasive investigations on wall paintings and on mural inscriptions in the Pompeii site, Microchem. J. 158 (2020) 105082, doi:10.1016/j.microc.2020.105082.

[28] S. Wold, K. Esbensen, P. Geladi, Machine learning: trends, perspectives, and prospects, Chemom. Intell. Lab. Syst. 2 (1987) 37–52, doi:10.1016/0169-7439(87)80084-9.

[29] S.J. Russell, P. Norvig, Artificial Intelligence: A Modern Approach, Third edition, Global edition, Pearson, Boston Columbus Indianapolis, 2016.

[30] M. Fiorucci, M. Khoroshiltseva, M. Pontil, A. Traviglia, A. Del Bue, S. James, Machine learning for cultural heritage: a survey, Pattern Recognit. Lett. 133 (2020) 102–108, doi:10.1016/j.patrec.2020.02.017.

[31] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning: With Applications in R, Springer US, New York, NY, 2021, doi:10.1007/978-1-0716-1418-1.

[32] A. Krogh, What are artificial neural networks? Nat. Biotechnol. 26 (2008) 195–197, doi:10.1038/nbt1386.

[33] A. Harth, The study of pigments in cultural heritage: a review using machine learning, Heritage 7 (2024) 3664–3695, doi:10.3390/heritage7070174.

[34] M. Mishra, Machine learning techniques for structural health monitoring of heritage buildings: a state-of-the-art review and case studies, J. Cult. Herit. 47 (2021) 227–245, doi:10.1016/j.culher.2020.09.005.

[35] M. Vermeulen, A. McGeachy, B. Xu, H. Chopp, A. Katsaggelos, R. Meyers, M. Alfeld, M. Walton, XRFast a new software package for processing of MA-XRF datasets using machine learning, J. Anal. At. Spectrom. 37 (2022) 2130–2143, doi:10.1039/D2JA00114D.

[36] S. Di Frischia, P. Giammatteo, F. Angelini, V. Spizzichino, E. De Santis, L. Pomante, Enhanced data augmentation using GANs for Raman spectra classification, in: Proceedings of the IEEE International Conference on Big Data, 2020, pp. 2891–2898, doi:10.1109/BigData50022.2020.9377977.

[37] T. Kleynhans, C.M. Schmidt Patterson, K.A. Dooley, D.W. Messinger, J.K. Delaney, An alternative approach to mapping pigments in paintings with hyperspectral reflectance image cubes using artificial intelligence, Herit. Sci. 8 (2020) 84, doi:10.1186/s40494-020-00427-7.

[38] A. Towarek, L. Halicz, S. Matwin, B. Wagner, Machine learning in analytical chemistry for cultural heritage: a comprehensive review, J. Cult. Herit. 70 (2024) 64–70, doi:10.1016/j.culher.2024.08.014.

[39] A. Chen, R. Jesus, M. Vilarigues, Identification and visualization of pure and mixed paint pigments in heritage artwork using machine learning algorithms, SN Comput. Sci. 4 (2022) 115, doi:10.1007/s42979-022-01529-8.

[40] T. Kleynhans, C.M. Schmidt Patterson, K.A. Dooley, D.W. Messinger, J.K. Delaney, An alternative approach to mapping pigments in paintings with hyperspectral reflectance image cubes using artificial intelligence, Herit. Sci. 8 (2020) 84, doi:10.1186/s40494-020-00427-7.

[41] B. Bilici Genc, E. Bostanci, B. Eskici, H. Erten, B. Caglar Eryurt, K. Acici, D. Ketenoglu, T. Asuroglu, Development of a new non-destructive analysis method in cultural heritage with artificial intelligence, Electronics 13 (2024) 4039 . (Basel), doi:10.3390/electronics13204039.
[42] R. Piovesan, C. Mazzoli, L. Maritan, P. Cornale, Fresco and lime-paint: an experimental study and objective criteria for distinguishing between these painting techniques, Archaeometry 54 (2012) 723–736, doi:10.1111/j.1475-4754.2011.00647.x.
[43] R. Sáez-Hernández, M.J. Luque, A.R. Mauri-Aucejo, Á. Morales-Rubio, M.L. Cervera, A two-way analytical investigation of ancient gold coins: elemental and colorimetric description of precious materials, Heliyon 10 (2024), doi:10.1016/j.heliyon.2024.e34623.
[44] A. Molada-Tebar, Á. Marqués-Mateu, J.L. Lerma, S. Westland, Dominant color extraction with K-means for camera characterization in cultural heritage documentation, Remote Sens. 12 (2020) 520, doi:10.3390/rs12030520.
[45] A. Molada-Tebar, Á. Marqués-Mateu, J.L. Lerma, Correct use of color for cultural heritage documentation, in: Proceedings of the ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences IV-2-W6, 2019, pp. 107–113, doi:10.5194/isprs-annals-IV-2-W6-107-2019.
[46] A. Molada-Tebar, J.L. Lerma, Á. Marqués-Mateu, Camera characterization for improving color archaeological documentation, Color Res. Appl. 43 (2018) 47–57, doi:10.1002/col.22152.
[47] R.W. Kennard, L.A. Stone, Computer aided design of experiments, Technometrics 11 (1969) 137–148, doi:10.1080/00401706.1969.10490666.
[48] M.Á. Martínez-Domingo, A.B. López-Baldomero, M. Tejada-Casado, M. Melgosa, F.J. Collado-Montero, Colorimetric evaluation of a reintegration via spectral imaging—Case study: Nasrid tiling panel from the Alhambra of Granada (Spain), Sensors 24 (2024) 3872, doi:10.3390/s24123872.
[49] M.Á. Martínez-Domingo, M. Melgosa, K. Okajima, V.J. Medina, F.J. Collado-Montero, in: Spectral image processing for museum lighting using CIE LED illuminants, 19, 2019, p. 5400, doi:10.3390/s19245400.
[50] M.A. Barbero-Álvarez, S. Brenner, Colour fidelity-respecting workflow for digital cultural heritage preservation and documentation, 2024 accessed November 23, 2024, doi:10.2312/gch.20241255.
[51] M. Tkalcic, J.F. Tasic, Colour spaces: perceptual, historical and applicational background, in: Proceedings of the IEEE Region 8 EUROCON 2003. Computer as a Tool, 1, 2003, pp. 304–308, doi:10.1109/EURCON.2003.1248032.
[52] J. Jiang, D. Liu, J. Gu, S. Süsstrunk, What is the space of spectral sensitivity functions for digital color cameras? in: Proceedings of the IEEE Workshop on Applications of Computer Vision WACV, 2013, pp. 168–179, doi:10.1109/WACV.2013.6475015.