Animal Species Identification in Historical Parchments by Continuous Wavelet Transform–Convolutional Neural Network Classifier Applied to Ultraviolet–Visible–Near-Infrared Spectroscopic Data

Roy, Nicolas; Pierard, Henry; BOUHY, Julie; Mayer, Alexandre; Deparis, Olivier; Gravis, David

Published in:
Intelligent Computing

DOI:
10.34133/icomputing.0101

Publication date:
2024

Document Version
Publisher's PDF, also known as Version of record

Intelligent Computing
A SCIENCE PARTNER JOURNAL

RESEARCH ARTICLE

Animal Species Identification in Historical Parchments by Continuous Wavelet Transform–Convolutional Neural Network Classifier Applied to Ultraviolet–Visible–Near-Infrared Spectroscopic Data

Nicolas Roy[1,2†], Henry Pièrard[1,2], Julie Bouhy[1], Alexandre Mayer[1,2], Olivier Deparis[1,3], and David Gravis[1,3*†]

[1]Namur Institute of Structured Matter (NISM), University of Namur, Namur, Belgium. [2]Namur Institute for Complex Systems (naXys), University of Namur, Namur, Belgium. [3]Heritages, Transmissions, Inheritances Institute (PaTHs), University of Namur, Namur, Belgium.

*Address correspondence to: david.gravis@unamur.be (D.G.)
†These authors contributed equally to this work.

Citation: Roy N, Pièrard H, Bouhy J, Mayer A, Deparis O, Gravis D. Animal Species Identification in Historical Parchments by Continuous Wavelet Transform–Convolutional Neural Network Classifier Applied to Ultraviolet–Visible–Near-Infrared Spectroscopic Data *Intell. Comput*. 2024;3:Article 0101. https://doi.org/10.34133/icomputing.0101

Submitted 26 March 2024
Revised 26 June 2024
Accepted 1 July 2024
Published 17 October 2024

Identification of animal species in medieval parchment manuscripts is highly relevant in cultural heritage studies. Usually, species identification is performed with slightly invasive methods. In this study, we propose a contactless methodology based on reflectance spectrophotometry (ultraviolet–visible–near-infrared) and a machine learning approach for data analysis. Spectra were recorded from both historical and modern parchments crafted from calf, goat, and sheep skins. First, a continuous wavelet transform was performed on the spectral data as a preprocessing step. Then, a semisupervised neural network with a 2-component architecture was applied to the preprocessed data. The network architecture chosen was CWT-CNN (continuous wavelet transform–convolutional neural network), which, in this case, is composed of a convolutional autoencoder and a single-layer dense network classifier. Species classification on holdout historical parchments was attained with a mean accuracy of 79%. The analysis of Shapley additive explanations values highlighted the main spectral ranges responsible for species discrimination. Our study shows that the animal species signature is encoded in a wide band-convoluted wavelength range rather than in specific narrow bands, implying a complex phenotype expression that influences the light scattering by the material. Indeed, the overall skin composition, in both micro- and macroscopic physicochemical properties, is relevant for animal identification in parchment manuscripts.

Introduction

Data-oriented approaches are emerging in the field of cultural heritage as a powerful tool for the accurate inference of some properties of interest based on data generated by experimental analysis techniques. For instance, such approaches associated to spectroscopic data have been successfully employed to characterize historical papers [1], photographic silver prints [2], color photographs [3], Islamic papers [4], Chinese papers [5], and paper books [6]. Many of these studies used data preprocessing techniques and predicted the physical characteristics or production dates of the heritage objects through advanced data analysis models. Lately, studies in cultural heritage considered machine learning models such as random forest (RF), k-nearest neighbors (kNN) [6], and support vector machines (SVMs) [7] but also convolutional neural networks (CNNs) [8]. Data-oriented methodologies have also been recently employed in the field of zooarchaeology [9]. Zooarchaeology, a branch of science at the crossroads of zoology and archaeology, focuses on the relation between humans and animals in prehistory and history [10]. In this vast field, some studies aim to identify species of long-dead animals through various analysis techniques [11,12]. More generally, zooarchaeology deals with the identification of the materials used to craft tools and objects of animal origin; for instance, it examines European medieval parchment manuscripts made from the processed skins of young domestic animals (mainly goats, sheep, and calves), sometimes stillborn or dead young [13]. Since the 2000s, the materiality of writing has become a field of study on its own [14–16]. In particular, interest in the animal origin of parchments has grown continuously in the past decades [14,17–21]. At first, studies on the animal origin of parchments aimed to complement historical studies on parchment manuscripts. Indeed, codicology—the discipline that studies codex manuscripts—is increasingly concerned with the production, economy, and materiality of medieval parchments [20]. Materials science studies on parchment are useful not only to historians,

but also to curators and restorers, who seek to adapt parchment conservation conditions and improve restoration practices [22–25]. In this context, knowing the animal species of preserved parchments in museums and archive collections is critical [11,19,26]. For this purpose, electrostatic zooarchaeology by mass spectrometry (eZooMS) has become a reference technique [27]. This technique discriminates between parchment animal species (goat, calf, and sheep) by analyzing the species-specific variations of collagen peptide sequences through their mass fingerprints [27]. Although the species is readily obtainable by eZooMS, the technique requires microsampling of collagen from the parchment, i.e., the parchment must be gently rubbed to rip collagen off the surface [27]. The microsamples then require aqueous sample preparation [27] before analysis by mass spectrometry.

As one deals with historical parchments from collections, experimental characterization techniques must be as minimally invasive as possible. A panel of techniques are available and adapted to meet this requirement [18,22,24,28–30]. Among them, spectrophotometry methods such as ultraviolet–visible–near-infrared (UV–visible–NIR), Raman, and Fourier transform infrared spectroscopy techniques are excellent candidates in this prospect, as they enable strictly noninvasive measurements (i.e., measurements without contact). Moreover, they do not require sample preparation (use of chemicals) and can be performed in an atmospheric environment. Furthermore, spectral photometric measurements can be easily implemented in libraries, museums, and restoration studios. This facilitates in situ data acquisition and sharing between research teams through the collaborative building of a database of spectra and ensures that data were acquired in the same conditions and normalized. These optical techniques are therefore preferable to eZooMS, which is microinvasive.

A previous study [18] showed that parchment species information could be extracted from UV–visible–NIR scattering and absorption spectral measurements. However, the signature of the animal species appeared to be nontrivially embedded in the spectrophotometric data; thus, retrieving the information required the application of principal component analysis. In fact, the UV–visible–NIR spectral response of organic materials often includes broad molecular overtone and combination bands, resulting in intricate spectra that complicate the attribution of distinct features to particular chemical components [31].

Unsurprisingly, the animal phenotype manifests itself in the skin structure, which enables its recognition. For example, experts can discriminate between calf, goat, and sheep skins by examining hair follicle patterns with the naked eye [13,32]. It is therefore reasonable to assume that the various scattering and absorption properties of parchment carry the indirect signature of the species phenotype. At the microscopic scale, the chemical composition of the parchment (which comprises collagen, fats, etc.) influences light absorption in the ultraviolet to infrared range, whereas, at the mesoscopic scale, the collagen fiber network influences light scattering.

In Ref. [18], the spectrophotometric characterization of parchment involved measurements of both diffuse and total (hemispherical) reflectance, transmittance, and absorptance. In the present study, we chose to measure reflectance spectra only, which was sufficient for species identification. The advantage of restricting measurements to reflectance is that such measurements do not limit parchment size (see Materials and Methods). The physical interpretation of spectral bands is, however, made more difficult when only reflectance is recorded (see discussion below). Unlike eZooMS, which is limited to collagen analysis, spectrophotometry can probe a range of parchment chemical components, as well as their physicochemical properties. The measurement wavelength range must be chosen according to the targeted components (collagen, fats, etc.). Electronic transitions causing optical absorption are located in the UV–visible–NIR range, whereas rotational and vibrational transitions take place in the mid- and far-IR ranges. The latter transitions are characteristic of not only local molecular structure but also aspects of the chemical environment of the molecules, such as the degree of crystallinity of the material. This explains why mid- and far-IR spectrophotometry is often used to quantify parchment degradation (aging) [30,33]. While it could also help identify the species through specific molecules (e.g., fats), we therefore preferred UV–visible–NIR to mid- and far-IR spectrophotometry, because the former spectral ranges are less sample-structure-specific and contain some molecular fingerprints that would strengthen the identification better than the latter.

Chemometric characterization of spectra is often supported by some type of systematic data analysis, for instance, principal component analysis (PCA), followed by kNN for classification [18], partial least squares (PLS) regression [6], support vector machine regression (SVR) [34], or a CNN [35,36]. Given the intricate embedding of information, the use of machine learning algorithms is a natural choice. For instance, in dating historical papers, a recent study applying RF and kNN algorithms to NIR spectroscopic data [6] reached a date resolution as low as 2 years.

In this study, the extraction of species information embedded in recorded spectra was achieved with a larger and more diverse dataset than in Ref. [18] as well as with a more advanced data analysis method. Specifically, we used an original machine learning (ML) method to classify the animal species (to assign labels) to parchments based on UV–visible–NIR spectroscopic data. In our method, a semisupervised convolutional autoencoder (CAE) and classifier operates on a continuous wavelet transform (CWT) representation of the UV–visible–NIR spectra. This network architecture takes advantage of the unsupervised feature-learning capabilities of an autoencoder (AE) to reduce the size of the classification network part that requires species-labeled spectra for training. The complexity of the network is further reduced as the first convolution layer is provided by CWT. Moreover, CWT improves the network performance and stability to hyperparameters while being an efficient interface for feature selection and spectrum filtering. This architecture follows the parsimony principle [37] to mitigate overfitting: it reduces the number of trainable parameters while exploiting both labeled and unlabeled data. Other tested ML methods show poor classification accuracy on our dataset.

Materials and Methods

The dataset: Definition, processing, and characterization of samples

Both modern and historical parchments processed from sheep, calf, and goat skins were used in this study. Skins from different breeds and individuals were selected to increase the diversity in each species. This sample diversity is important for the robustness of the machine learning model. We used $4 \times 4\,cm^2$ pieces from different regions of the skin to increase the number

of samples in the case of modern parchments. Similarly, the 2 sides of the parchment (the flesh and grain sides) were regarded as different samples since they exhibit different surface scattering properties. For historical parchments, species were labeled by mass spectroscopy peptide sequencing [38] and all 3 species were represented. In total, the sample set consisted of 55 (25) calf, 75 (19) sheep, and 68 (20) goat modern (historical) samples. A more detailed description of the sample set can be found in the Supplementary Materials.

For each sample, the total (hemispherical) reflectance was measured, which included both the diffuse and specular components of the reflected light at wavelengths ranging from 250 to 2,500 nm. A double-beam spectrophotometer (Perkin Elmer Lambda 750) was equipped with an integrating sphere, which was used to collect light scattered in all directions. The wavelength scanning step was fixed at 10 nm to enable quick acquisition (less than 3 min) while keeping resolution high enough. It is worth noting that these specifications can be reached by smaller (handheld) devices and/or devices equipped with fiber optics that could better fit conservators' requirements.

The dataset resulting from this collection of samples presents several challenges. As mentioned, the useful features for species identification are embedded in the bulk of the spectra. Consequently, these spectra cannot be accurately classified—i.e., assigned to a specific species—using straightforward methods such as kNN alone. This challenge is made even stronger given the covariate shift between historical and modern spectra. The influence of natural aging induces a shift in the feature (covariate) distribution between modern parchments and historical parchments that underwent aging in various conditions. This aging may cause broadening of spectral bands due to collagen subpopulations and the appearance of degradation components. Nonetheless, the utilization of modern parchment spectra stands as a pivotal element for the successful configuration of the classifier. This is primarily due to the scarcity of labeled historical spectra, which makes the integration of modern data essential in the training procedure. Ultimately, the limited size of the dataset imposes constraints on the application of deep learning techniques. It is crucial to exercise meticulous caution to prevent overfitting and ensure that the models remain generalizable and effective across unseen data.

CWT representation

As stated in Coppola et al. [6], a robust spectral preprocessing technique ought to eliminate unwanted sources of variability, such as noise, baseline shifts, and scattering. There are countless filtering methods: Savitzky–Golay (SG), spatial filters, finite impulse response, Fourier masking, and splines can reduce noise and/or baseline shifts. In our approach, we chose to work with the CWT. The CWT is widely employed in signal processing, notably in speech recognition, where it eases the visualization of spectral components across the timeline. When applied to spectral data $R(\lambda)$, the timeline is actually the wavelength of light. The CWT convolves the spectra with wavelets of decreasing scale σ in order to observe features of decreasing size. From observation of the baseline up to the sharpest features, CWT acts like a mathematical microscope. Similar preprocessing is not uncommon in the treatment of near-infrared spectra [35].

CWT is a powerful preprocessing step for CNNs as it provides a first convolution layer that does not require numerous trainable parameters. The CWT is expressed as the convolution of each spectra $R(\lambda)$ by a wavelet, i.e.,

$$W(\lambda,\sigma) = \frac{1}{\sqrt{\sigma}} \int_{-\infty}^{\infty} R(\lambda') \Phi\left(\frac{\lambda' - \lambda}{\sigma} \right) d\lambda'. \quad (1)$$

The wavelet, used for convolution, is the common Ricker wavelet,

$$\Phi(\lambda,\sigma) = \frac{2}{\sqrt{3\sqrt{\pi}}} \left(1 - \frac{\lambda^2}{\sigma^2} \right) e^{-\lambda^2/(2\sigma^2)}. \quad (2)$$

In practice, the CWT turns a spectrum $R_i(\lambda)$ into a scalogram $W_i(\lambda,\sigma)$. The discrete choice of wavelet scales σ_j and measured spectral wavelengths λ_k makes the scalogram a 2-dimensional (2D) picture, ready to be later processed by a 2D CNN.

As a filtering step, the considered wavelet scales σ range from σ_0 to $\sigma_0 + \Delta\sigma$, defining a window on the full scalogram (Fig. 1). This is indeed a feature scale filtering: $\Delta\sigma$ acts as the bandwidth and σ_0 controls the scales captured within the band. This windowed scalogram can be considered as both a filtering step and a feature selection that can be further refined by the CNN layer based on the training data.

Network architecture

In this study, we harnessed the unsupervised feature extraction of autoencoders to assist a classification network in a common semisupervised [39] training procedure. The autoencoder and classifier network were integrated in a single architecture, as shown in Fig. 1. With this combination, we aimed to improve the generalization capability of the classification network in a context where labels are difficult to obtain for a considerable part of the dataset (as discussed above).

An autoencoder is a neural network specifically designed and trained for the purpose of encoding and then reconstructing its input such that the difference between input and output is minimized [40]. A pivotal aspect of the autoencoder lies in its capability to reduce the dimensionality of the analyzed data during its encoding operation. This makes AEs a formidable tool for the unsupervised learning of compressed data representations. In order to achieve accurate reconstruction, the autoencoder learns to distinguish meaningful features of the input data in its encoder component (Fig. 1). This capability has been used in the past to accelerate the training of deeper networks [41]. It is noteworthy that autoencoders do not require any labeled data as they train solely on the task of data reconstruction.

The autoencoder is a 2D CNN operating on the CWT representation discussed above. The classification network is a single dense layer starting from latent representation indexed −1 in Fig. 1. The classification network is optimized through a negative log-likelihood (NLL) loss function while the reconstruction from the AE uses a mean squared error (MSE) loss function. The optimal reconstruction loss accounted for approximately 30% in the total loss when optimizing the network parameters, whereas the classification loss accounted for 70%. This was desired as the classification task must take precedence over the auxiliary reconstruction. The reconstruction was made even rougher due to the small central (zeroth order) latent

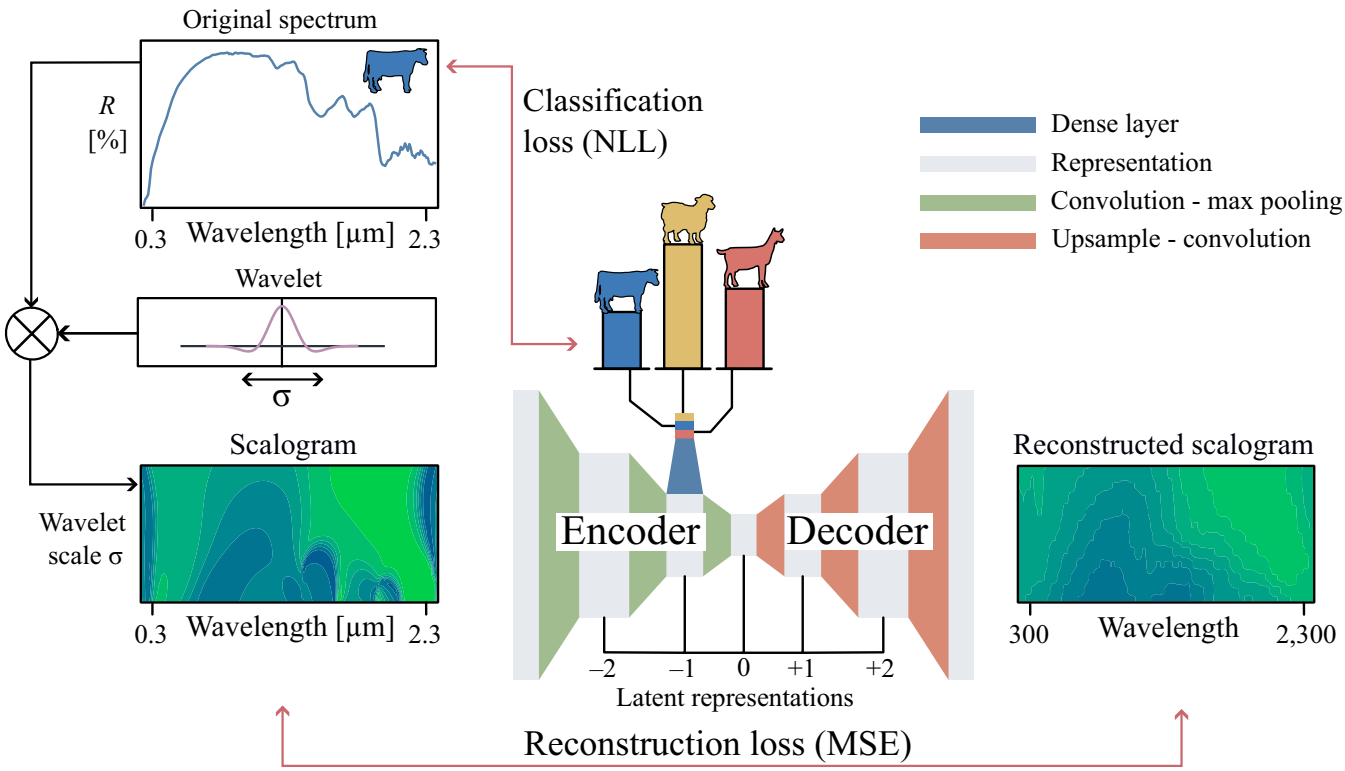

Fig. 1. Architecture of the semisupervised classifier. An illustrative calf reflectance spectrum is shown on the top left. It is turned into the scalogram below by convolution with the Ricker wavelet for scales from $\sigma_0 = 0.1$ μm with $\Delta\sigma = 0.32$ μm. The network is represented in the center with its input and output (reconstruction) on the left and right sides, respectively. The classification network is highlighted in blue. Losses associated to the classification and reconstruction are also illustrated.

representation. The small size of the latent space favored the extraction of synthetic patterns in the data.

During the training procedure, a rudimentary data augmentation procedure was used: spectral values were randomly scaled by a uniform random factor between 0.5 and 1.5. Because our spectra were normalized to the light source intensity, the network did not take into account the absolute reflectance values and only considered shifts in distributions between species. Gaussian noise ($\mu = 0$, $\sigma = 0.05$) was applied on the whole scalogram as a regularization measure. The training procedure included $\mathbb{L}_2$ regularization at a rather high scale of 0.01 in order to mitigate over-fitting.

Results

Performance metrics

To assess the performance of the classification models, we employed 2 distinct metrics. The first metric involved k-fold validation ($\mathcal{K}$), wherein one-fourth of the dataset was allocated for validation purposes. We repeatedly trained 4 models, each time altering the portion reserved for validation. As the dataset includes both modern and historical parchments, the training and validation sets contained both.

The second metric, which we refer to as historic validation ($\mathcal{H}$), aligns closely with our objective of identifying the animal species of historical parchments. In this approach, we trained the network using only the modern parchment spectra (holding out the historical parchment spectra) and subsequently validated its performance against all 64 historical parchment spectra. This metric was challenging, since there was a notable covariate shift between the spectra of historical and modern

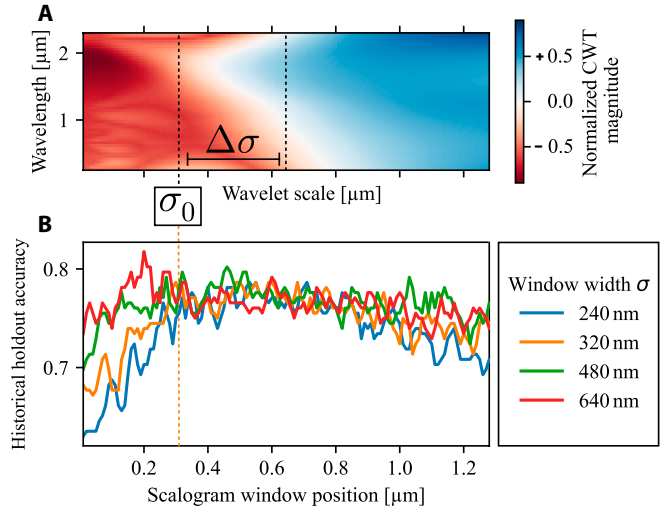

Fig. 2. Influence of scalogram feature selection on classification performance. (A) A normalized scalogram from sheep parchment is shown with wavelet widths ranging from 10 to 1,280 nm. (B) The classification accuracy is charted against the scalogram window offset (σ_0) for different window widths ($\Delta\sigma$).

parchments. Consequently, the network must exhibit a certain level of robustness to handle this shift. These dual metrics offered respectively less and more severe evaluations of the performance of our model.

Guidelines for feature selection

This section presents an empirical study of the importance of spectral features relative to their scale. Feature selection was

leveraged to gain a deeper understanding of the data at hand. The selection consisted in considering only a vertical band of the scalogram illustrated in Fig. 2A. The illustrative scalogram ranges from wavelet scales of 0.01 to 1.28 μm. The band is defined by a width ($\Delta\sigma$) and an offset (σ_0) on the wavelet scale axis. In Fig. 2B, the historical classification accuracy ($\mathcal{H}$) is charted against the window position σ_0 for different window widths $\Delta\sigma$. Four window widths were considered: $\Delta\sigma = 0.24$, 0.32, 0.48, and 0.64 μm. The historical classification accuracy is shown for each window size in Fig. 2B.

In Fig. 2, we observe that, for each window size $\Delta\sigma$, the accuracy tends to be higher for higher offsets (σ_0). The window sizes also have a strong impact, since larger windows lead to better accuracy with smaller offsets. The accuracy for the $\Delta\sigma = 0.24$,0.32, 0.48, and 0.64 μm windows rises for offsets ranging from $\sigma_0 = 0$ to ~0.3 μm. An illustrative window of $\Delta\sigma = 0.320$ μm is shown in Fig. 2A. The window offset ($\sigma_0 = 0.28$ μm) is placed in an area where accuracy rises with increasing σ_0. The increase of $\sigma_0 + \Delta\sigma$ to ~0.6 μm accumulates wavelet scales that are informative input to the network. After passing 0.6 μm, the accuracy starts to fall, indicating that the larger-scale components alone cannot provide enough information. This observation indicates that crucial species information lies in multiscale spectral features spanning from around $\sigma = 0.3$ to 0.6 μm. A well-performing range for σ_0 and various $\Delta\sigma$ is illustrated in Fig. 2.

Benchmark of CWT-CAE classifier

In the previous section, we revealed that the CWT-CAE approach can reach an accuracy from 75% to 79% (Fig. 2B) with permissive parameter selection. In this section, this result is compared to the results of other popular classification methods (Table). For all methods, the parameters were systematically tuned to achieve the reported optimal performance. Whenever the algorithm presented stochastic processes, multiple independent executions of the model-fitting procedure were recorded and the average performance was reported. Code for this benchmark is made available for reproducibility at https://github.com/Kaeryv/Manuspector. Reported methods include kNN and RF, already used in the cultural heritage field [6], as well as 1D and 2D CNNs [35,36]. Preprocessing methods include SG and CWT. PCA-kNN uses PCA before kNN for dimension reduction, as in Ref. [18].

Our method, CWT-CAE, is the best method on both validation metrics. A representative value of 79% classification accuracy was obtained by averaging the optimal accuracy obtained in 10 successful independent training processes of CWT-CAE. Among all model realizations (successful or failed), the best model achieved an accuracy of 83%, while the average accuracy (including failed models) was 75%. Failed models are outliers characterized by particularly low accuracy that can go well below 50%. The failure of these models is due to an unfortunate initialization of the weights in the training procedure that led to a network with constant output. One hypothesis to explain these failures is that the use of strong $\mathbb{L}_2$ weight regularization may sometimes trap the network in local optima where all the weights are zero. A similar evaluation approach was taken for a CAE operating directly on the spectra without CWT (SG-CAE1D), but the method proved to be so sensitive to its hyperparameters that we deemed it unreliable.

Considering the results presented in Table, the limitations of certain methods, such as kNN, become apparent. The disparity between the 4-fold accuracy (0.82) and historical accuracy (0.59) suggests a failure of the kNN model to generalize effectively to historical spectra. This outcome is expected, given that kNN relies on distances between samples, making it highly susceptible to covariate shift between historical and modern parchments [42]. A similar trend is observed with SVM. However, there is a slight improvement with PCA-kNN, indicating that although the accuracy of the model diminishes, the model exhibits enhanced generalization capabilities following dimensionality reduction to 3 dimensions. From this benchmark analysis, it appears that both PCA and explicit CWT windowing contribute positively to feature selection, thereby influencing model performance.

Explainable classifications using SHAP values

Performing well in the classification task is only one aspect of our objective. This section aims to understand a posteriori the (black-box) classification process [43]. Approximations of Shapley additive explanations (SHAP) values [44] represent a widely recognized tool for the task. While SHAP values are better interpreted locally (sample-wise), a global representation of the sample-averaged SHAP values is provided in Fig. 3. As SHAP values can be obtained as 2D distributions on λ and σ, we will

Table. Optimal historical holdout and k-fold accuracies for a selection of classification procedures. The algorithms include convolutional autoencoders (CAE) in 1D and 2D using continuous wavelet transform (CWT) as well as gradient-boosting (GBC), random forest (RF), k-nearest neighbors (kNN), and support vector machine (SVM) classifiers. Methods using an optimized Savitzky–Golay filter contain the SG prefix. The parameters leading to optimal historical accuracy are indicated for each method when practical.

Method	Averaged 4-fold accuracy	Historical holdout accuracy	Parameters
CWT-CAE2D	0.85	0.79	≈27, 000
SG-CAE1D	0.82	0.69	≈2, 000
SG-GBC	0.40	0.38	$n_{\text{estimators}} = 100$
SG-SVM	0.69	0.48	
SG-kNN	0.82	0.59	$k = 9$
SG-RF	0.43	0.40	≈3, 000
SG-PCA-kNN	0.56	0.625	$n_{\text{PC}} = 3, k = 1$

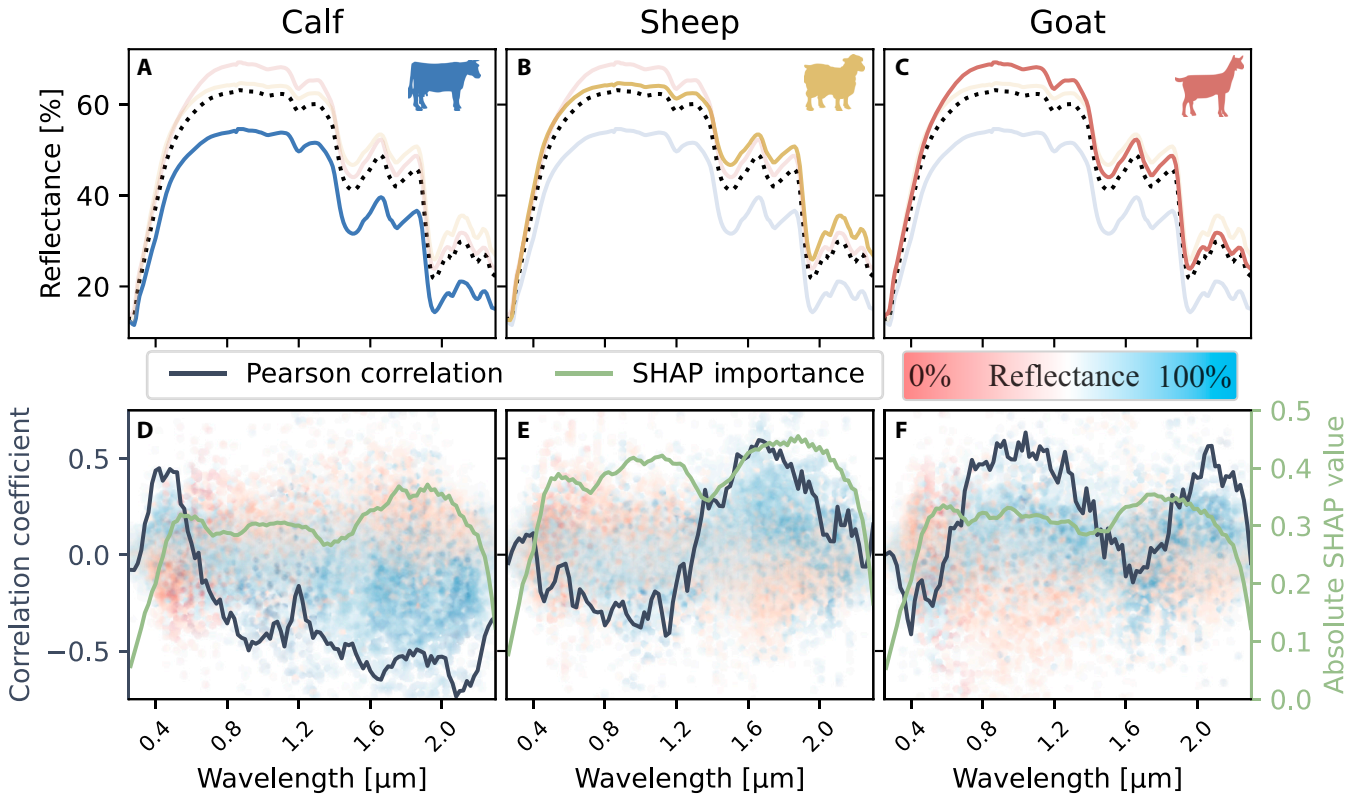

Fig. 3. Summary SHAP analysis on the whole dataset. (A to C) Illustrative spectra are shown for calf, sheep, and goat parchments. The average spectra over the dataset is charted as a dotted line in the background of each plot for reference. Illustrative spectra from the other species are also shown in dimmed colors. (D to F) The correlation (*C*) and importance (*I*) are represented on the left and right axes, respectively. In the background, unprocessed SHAP values are plotted as dots of color representing the reflectance.

refer to the SHAP values of sample i as a continuous function $S_i(\lambda,\sigma)$. Two global metrics were devised from this definition: the average correlation (Pearson coefficient function, ρ_P)

$$C(\lambda) = \frac{1}{N_{\text{samples}}} \sum_{i}^{N_{\text{samples}}} \rho_P\left(\int_{\sigma_0}^{\sigma_0+\Delta\sigma} S_i(\lambda,\sigma)d\sigma, R_i(\lambda)\right), \quad (3)$$

and the averaged importance:

$$I(\lambda) = \frac{1}{N_{\text{samples}}} \sum_{i}^{N_{\text{samples}}} \int_{\sigma_0}^{\sigma_0+\Delta\sigma} |S_i(\lambda,\sigma)|\, d\sigma. \quad (4)$$

Figure 3D to F shows mean tendencies out of the dataset by charting the SHAP values for the whole dataset against the wavelength. In the bottom charts, blue dots represent high reflectance and red dots denote low reflectance. To facilitate interpretation, the importance $I(\lambda)$ and correlation $C(\lambda)$ values are also charted for each species. Our interpretation focuses on spectral regions where the importance is high. For instance, the region with $\lambda \geq 0.6$ μm in Fig. 3 indicates anticorrelation ($C < 0$) between reflectance and the probability to be classified as calf.

The global analysis of averaged SHAP values (see the "Physicochemical discussion" section below) might initially suggest that classification is straightforward, as highlighted in the above example. However, our comparative benchmarks with simpler classification models indicate that it is not. While the averaged SHAP values in Fig. 3 provide a broad overview of the classification process, it is important to recognize that each sample possesses unique characteristics that influence its classification. This nuance is essential for understanding the complexity involved in accurately categorizing data: some samples may have SHAP values differing considerably from those reported in Fig. 3. With the examples of Fig. 3, our goal was to present a comprehensive overview of the classification process for practitioners in the field of cultural heritage to bridge the gap between complex data analysis techniques and their practical applications.

Discussion

Machine learning discussion

In this study, we developed a classification scheme that efficiently uses both labeled and unlabeled data to identify animal species from parchment UV–visible–NIR spectra. This method is designed to be robust to over-fitting and prepossessing, as it addresses the need for data-efficient classification in the absence of extensive labeled datasets. The classification model, which was trained exclusively on modern parchments, achieved an accuracy as high as 79% on holdout historical parchments. This outcome suggests the practicality of our approach, although the model was trained on a relatively narrow dataset. This result indicates that it is possible to extract species information from UV–visible–NIR spectral data, confirming that these spectra contain distinguishable features specific to different species [18].

However, the study also highlights the limitations we currently face, primarily the constraint imposed by the small size of the dataset available for training the model. This limitation is a common

challenge in specialized fields where collecting extensive and varied data can be difficult and time-consuming. To mitigate this issue for the present dataset, we tried removing the UV and visible parts of the spectra, which were the least critical in the obtained classification. This approach resulted in a slightly improved historical accuracy, up to 82% for $\sigma_0 < 0.2$ μm (see Supplementary Materials and related discussion). However, retaining the complete spectra is a more scalable approach in order to remain open to unforeseen correlations extracted from a bigger dataset. For this reason, the classification was operated in the full UV–visible–NIR range.

In a global, scalable approach, the best solution is a gradual increase in the dataset that incorporates more individuals (animals) among the collected parchments. We acknowledge that this expansion is a time-intensive process that would benefit considerably from external collaborations. Furthermore, classification would benefit from data augmentation that focuses on methods informed by the physicochemistry of parchment aging rather than relying on simple scaling and offsetting techniques. Such an approach would enrich the dataset in a manner that better reflects the complexities and nuances of realistic parchment aging processes, thereby enhancing the ability of the model to generalize from the training data to new, unseen historical parchment samples.

Physicochemical discussion

Our results show that species identification can be reached from reflectance spectra only, and that the classification algorithm focuses on different spectral ranges according to species. For measurements, the diffuse reflectance configuration was preferred as it offers the least invasive method for recording UV–visible–NIR spectra from parchment manuscripts, meeting the aforementioned requirement for contactless measurements. Furthermore, our method uses a limited amount of data due to the constraints imposed by the study of cultural heritage artifacts, yet is sufficient for species identification. In Fig. 4, the mean UV–visible–NIR spectra are replotted for better clarity and detailed physicochemical analysis, including band attribution to specific compounds (see also the Supplementary Materials).

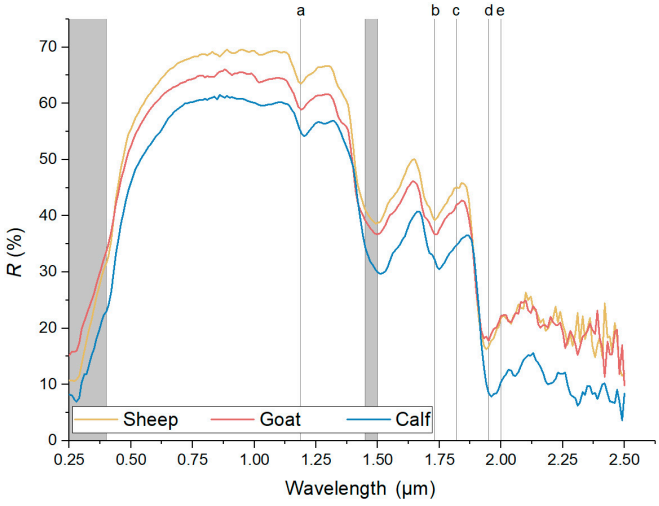

Fig. 4. Typical UV–visible–NIR spectra of calf, sheep, and goat parchments. Several specific bands and intervals are marked for interpretation: the left shaded area stretches from 0.250 to 0.400 μm while the right one stretches from 1.450 to 1.500 μm. The other marked bands are located at (a) 1.185 μm, (b) 1.730 μm, (c) 1.820 μm, (d) 1.950 μm, and (e) 2.000 μm.

The spectral specificity for classification is more striking for the calf skin, for which the correlation is high ($C > 0$), mostly in the visible region ($\lambda < 0.6$ μm), whereas goat and sheep skins have predominant features in the IR region (Fig. 3). Furthermore, calf skin is, in general, less reflective than the 2 other species. For sheep skin, the classification is characterized by a transition around 1.2 μm: sheep skin is more likely to be identified when the reflectance is low (below 1.2 μm), and vice versa above this wavelength threshold. Goat skin classification seems to be characterized by 4 different spectral regions: it is more likely to be identified when the reflectance is high (in the 0.5 to 1.5 μm and 1.7 to 2 μm ranges), or low (in the remaining wavelength ranges). The reflectance difference between species in the visible part of the spectrum is expected because parchment color is often used by experts as a method for distinguishing species using the naked eye or via colorimetry measurements through spectral analysis [45]. Nevertheless, this spectral region is often influenced by parchment degradation (aging), which could make identification more difficult. The influence of degradation extends to the UV region, where hyperchromicity is often attributed to collagen denaturation [46]. Spectral properties could be further exploited for parchment degradation assessments. They could also help differentiate older from more recent manuscripts, based on their natural degradation (see discussion in Supplementary Materials related to Fig. 4). Apart from the visible range, spectral differences observed in the species reflectance could be attributed to different skin surface properties, which induce differences in light scattering among the 3 species. Due to the complex nature of the surface and the composition of animal skins, it is difficult to predict how each species, and the related skin physicochemistry, really influence the scattering of light, especially in the IR region, where goat and sheep are differentiated by our algorithm. However, these spectral differences could be related to the animal phenotype, which manifests itself in the interaction of the parchment material with light. The collagen structure and the macroscopic properties of the skin (surface roughness, hair follicle sizes and densities, presence of fats, etc.), which are dependent on the species, influence the reflectance spectra, possibly explaining the variations described above.

A closer analysis of the spectra and a comparison with the algorithm output indicate that species differentiation solely based on specific bands is difficult. This could be explained as follows: (a) machine learning analysis requires a considerable amount of spectra, which requires speeding up recording by increasing the wavelength step, thus reducing spectral resolution and making band discrimination more difficult; (b) because transmission and absorption were not measured, information on the physicochemical properties of the skins is limited. More specifically, absorptance spectra could validate some hypotheses, mainly about the skin composition, which could validate some band attributions. While we limited ourselves to measuring reflectance, this choice did not impede the capability of the algorithm to discriminate between species. In addition to collagen bands, 2 other bands were attributed (-OH and C-H vibrations and deformations). For OH groups (see Fig. 4 and further analysis in the Supplementary Materials), bands near 1.5 and 1.95 μm correspond to hydroxyl groups in collagen or in water molecules, either in linked or free forms. The relative variation of these bands could indicate the state of the collagen structure,

i.e., its degradation [47]. The analysis of these bands, together with those in the UV region, could yield information on the degradation state of the parchment. C-H vibrations originate from the collagen molecules, but could also be present in other molecules inside the skins (fats, for instance). Nevertheless, C-H bands exhibit limited variations between the species, and do not seem to be considered by the algorithm. Indeed, these bands are not correlated to high local SHAP values and are mostly wound up among a wide spectral range. This could be explained as most C-H bands are often convoluted with other bands (especially in the 0.6 to 1.5 μm range), mainly from other moieties of the collagen molecules (see Fig. 4 and further analysis in the Supplementary Materials).

In conclusion, our study confirms that the species signature of the parchments is encoded in the UV–visible–NIR part of the scattered light spectrum. Detailed band attribution to specific moieties seems irrelevant for species discrimination: the species is not identified solely by the narrow-band signatures of one molecule or a specific moiety in the skin. This is to be expected, because few variations in the collagen sequence exist (only several percent of amino acids differ) between the 3 species (calf, sheep, and goat). These slight chemical structure differences would likely not be detected by the spectrometer, especially at the spectral resolution considered. Optical differences between the species probably lie within the general chemistry and structural differences that influence light scattering, more in line with the broad bands selected by the classification algorithm for species identification. The different light scattering properties depending on the species imply that the species signature is much more complex to interpret from a physicochemical point of view. We hope this study will help practitioners in cultural heritage with insights that are derived from a high-level analysis. Furthermore, the proposed methodology also highlights that, with the use of adapted data augmentation, preparation, and parameter regularization, neural networks remain practical in data-scarce scenarios.

Acknowledgments

Funding: Computational resources were provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under Grant No. 2.5020.11 and by the Walloon Region. This study was supported by the Pergamenum21 project of the University of Namur and by the NISM. A.M. was funded by the Fund for Scientific Research (F.R.S.-FNRS) of Belgium.
Author contributions: D.G., N.R., J.B., and H.P. carried out the experiments. D.G. and N.R. conceived the experiments. N.R. designed the model and the computational framework and analyzed the data. D.G. and H.P. analyzed the physicochemistry of the spectroscopic data. A.M. and O.D. helped supervise the project. All authors contributed to the final version of the manuscript.
Competing interests: The authors declare that they have no competing interests.

Data Availability

The spectra database and the code required for reproducing our numerical experiments are open and available in 2 separate repositories: https://github.com/Kaeryv/Light4Parch and https://github.com/Kaeryv/Manuspector.

Supplementary Materials

Sections S1 to S3
Figs. S1 and S2
References [48–50]

References

1. Trafela T, Strlic M, Kolar J, Litchblau DA, Andres M, Mencigar DP, Pihlar B. Nondestructive analysis and dating of historical paper based on IR spectroscopy and chemometric data evaluation. *Anal Chem*. 2007;16(79):6319–6323.
2. Martins A, Daffner LA, Fenech A, McGlinchey C, Strlic M. Non-destructive dating of fiber-based gelatin silver prints using near-infrared spectroscopy and multivariate analysis. *Anal Bioanal Chem*. 2012;402(4):1459–1469.
3. Fenech A, Strlic M, Cassar M. The past and the future of chromogenic colour photographs: Lifetime modelling using near-infrared spectroscopy & enhancement using hypoxia. *Appl Phys A*. 2012;106:411–417.
4. Mahgoub H, Gilchrist JR, Fearn T, Strlic M. Analytical robustness of quantitative NIR chemical imaging for Islamic paper characterization. In: *Optics for arts, architecture, and archaeology VI*. Munich (Germany): SPIE; 2017. p. 129–136.
5. Brown N, Lichtblau D, Fearn T, Strlic M. Characterisation of 19th and 20th century Chinese paper. *Herit Sci*. 2017;5:47.
6. Coppola F, Frigau L, Markelj J, Malesic J, Conversano C, Strlic M. Near-infrared spectroscopy and machine learning for accurate dating of historical books. *J Am Chem Soc*. 2023;145(22):12305–12314.
7. Xia J, Huang Y, Zhang J, du X, Yan H, Li Q, Li Y, Xiong Y, Min S. Development of a chemometric methodology based on FTIR spectra for paper dating. *Cellulose*. 2020;27:5323–5335.
8. Xia J, Du X, Xu W, Wei Y, Xiong Y, Min S. Non-destructive analysis the dating of paper based on convolutional neural network. *Spectrochim Acta A Mol Biomol Spectrosc*. 2021;248:Article 119290.
9. Hickinbotham S, Fiddyment S, Stinson TL, Collins MJ. How to get your goat: Automated identification of species from MALDI-ToF spectra. *Bioinformatics*. 2020;36(2): 3719–3725.
10. Thomas R, Holmes M, Morris J. "So bigge as bigge may be": Tracking size and shape change in domestic livestock in London (AD 1220-1900). *J Archaeol Sci*. 2013;40: 3309–3325.
11. Toniolo L, D'Amato A, Saccenti R, Gulotta D, Righetti PG. The silk road, Marco polo, a bible and its proteome: A detective story. *J Proteome*. 2012;75(11):3365–3373.
12. Gianvincenzo FD, Granzotto C, Cappellini E. Skin, furs, and textiles: Mass spectrometry based analysis of ancient protein residues. In: *The textile revolution in Bronze Age Europe*. Cambridge (UK): Cambridge University Press; 2019. p. 304–316.
13. Vnoucek J, Fiddyment S, Quandt A, Rabitsch S, Collins M, Hofmann C. The parchment of the Vienna Genesis: Characteristics and manufacture. In: *The Vienna genesis*. Vienna (Austria): Bohlau Verlag; 2020. p. 35–70.
14. Turner NK. The materiality of medieval parchment: A response to "the animal turn". *Revista Hispanica Moderna*. 2018;71(1):39–67.
15. Pohl B. Documenting the everyday in medieval Europe: The social dimensions of a writing revolution, 1250–1350. *Am Hist Rev*. 2021;126(1):370–371.

16. Clanchy MT. *From memory to written record: England 1066–1307*. 3rd ed. Hoboken (NJ): John Wiley & Sons; Hoboken (New Jersey); 2013.
17. Vilde V, Abel ML, Watts JF. A surface investigation of parchments using ToF-SIMS and PCA. *Surf Interf Anal*. 2016;48(7):393–397.
18. Alvarez AMF, Bouhy J, Dieu M, Charles C, Deparis O. Animal species identification in parchments by light. *Sci Rep*. 2019;9:1825.
19. Ruffini-Ronzani N, Nieus JF, Soncin S, Hickinbotham S, Dieu M, Bouhy J, Charles C, Ruzzier C, Falmagne T, Hermand X, et al. A biocodicological analysis of the medieval library and archive from Orval Abbey, Belgium. *R Soc Open Sci*. 2021;8(6):Article 210210.
20. Bischoff FM. Observation sur lemploi de differentes qualites de parchemin dans le livre medieval. In: Maniaci M, Munafò PF, editors. *Ancient and medieval book materials and techniques*. Vatican City: Biblioteca Apostolica Vaticana; 1993. p. 57–94.
21. Fiddyment S, Holsinger B, Ruzzier C, Devine A, Binois A, Albarella U, Fischer R, Nichols E, Curtis A, Cheese E, et al. Animal origin of 13th-century uterine vellum revealed using noninvasive peptide fingerprinting. *Proc Natl Acad Sci USA*. 2015;112(49):15066–15071.
22. Larsen R, editor. *Improved damage assessment of parchment*. Brussels (Belgium): European Commission; 2007.
23. Bainbridge A, editor. *Conservation of books*. 1st ed. London (UK): Routledge; 2023.
24. Kern MS, Pataki-Hundt A, Wouters J, Kirby DP. Accelerated ageing of parchment: Investigation of a photo catalysed, low-heat approach. *Restaurator*. 2018;39(1):33–69.
25. Axelsson KM, Larsen R, Sommer DV, Melin R. Degradation of collagen in parchment under the influence of heat-induced oxidation: Preliminary study of changes at macroscopic, microscopic, and molecular levels. *Stud Conserv*. 2016;61(1):46–57.
26. Buckley M. Species identification of bovine, ovine and porcine type 1 collagen; comparing peptide mass fingerprinting and LC-based proteomics methods. *Int J Mol Sci*. 2016;17(4):445.
27. Fiddyment S, Teasdale MD, Vnoucek J, Leveque E, Binois A, Collins MJ. So you want to do biocodicology? A field guide to the biological analysis of parchment. *Herit Sci*. 2019;7:35.
28. Gravis D, Roy N, Ruffini-Ronzani N, Houssiau L, Felten A, Tumanov N, Deparis O. Secondary ion mass spectrometry, a powerful tool for revealing ink formulations and animal skins in medieval manuscripts. *R Soc Open Sci*. 2023; doi.org/10.1098/rsos.230059.
29. Schmeltz M, Robinet L, Heu-Thao S, Sintès JM, Teulon C, Ducourthial G, Mahou P, Schanne-Klein MC, Latour G. Noninvasive quantitative assessment of collagen degradation in parchments by polarization-resolved SHG microscopy. *Sci Adv*. 2021;7(29):1090–1106.
30. Odlyha M, Theodorakopoulos C, Bozec L. Fourier transform infra-red spectroscopy (ATR/FTIR) and scanning probe microscopy of parchment. E-preservation. *Science*. 2009;6:138–144.
31. Ma, Shao. Continuous wavelet transform applied to removing the fluctuating background in near-infrared spectra. *J Chem Inf Comput Sci*. 2004;44(3):907–911.
32. Vnoucek J. Care and conservation of manuscripts 17. In: Driscoll MJ, editor. *Not all that shines like vellum is necessarily so*. Copenhagen (Denmark): Museum Tusculanum; 2021. p. 27–60.
33. Mozir A, Gonzalez L, Cigić IK, Wess TM, Rabin I, Hahn O, Strlič M. A study of degradation of historic parchment using small-angle X-ray scattering, synchrotron-IR and multivariate data analysis. *Anal Bioanal Chem*. 2012;402(4):1559–1566.
34. Kaneko H, Funatsu K. Fast optimization of hyperparameters for support vector regression models with highly predictive ability. *Chemom Intell Lab Syst*. 2015;142:64–69.
35. Hosseinpour-Zarnaq M, Omid M, Sarmadian F, Ghasemi-Mobtaker H. A CNN model for predicting soil properties using VIS–NIR spectral data. *Environ Earth Sci*. 2023;82:382.
36. Shang H, Shang L, Wu J, Xu Z, Zhou S, Wang Z, Wang H, Yin J. NIR spectroscopy combined with 1D-convolutional neural network for breast cancerization analysis and diagnosis. *Spectrochim Acta A Mol Biomol Spectrosc*. 2023;287(Part 1): Article 121990.
37. Hawkins DM. The problem of overfitting. *J Chem Inf Comput Sci*. 2004;44(1):1–12.
38. Decorte R, Polet C, Boudin M, Tilquin F, Matroule J-Y, Dieu M, Charles C, Carlier A, Lebecque F, Deparis O. An interdisciplinary study around the reliquary of the late cardinal Jacques de Vitry. *PLOS ONE*. 2019;14(2):Article e0201424.
39. Wu X, Zhang Y, Cheng C, Peng Z. A hybrid classification autoencoder for semi-supervised fault diagnosis in rotating machinery. *Mech Syst Signal Process*. 2021;149:Article 107327.
40. Kingma DP, Welling M. Auto-encoding variational Bayes. arXiv. 2013. https://doi.org/10.48550/arXiv.1312.6114.
41. Bengio Y, Lamblin P, Popovici D, Larochelle H. Greedy layer-wise training of deep networks. In: *Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference*. Cambridge (MA): The MIT Press; 2007.
42. Kalpakis K. Consensus–relevance kNN and covariate shift mitigation. *Mach Learn*. 2024;113(1):325–353.
43. Samek W, Montavon G, Lapuschkin S, Anders CJ, Muller KR. Explaining deep neural networks and beyond: A review of methods and applications. *Proc IEEE*. 2021;109(3): 247–278.
44. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, editors. *Advances in Neural Information Processing Systems 30*. Curran Associates, Inc.; 2017. p. 4765–4774.
45. Malissa A, Cappa F, Schreiner M, Marchetti-Deschmann M. Spectral features differentiate aging-induced changes in parchment—A combined approach of UV/VIS, μ-ATR/FTIR and μ-Raman spectroscopy with multivariate data analysis. *Molecules*. 2023;28(12):4584.
46. Na GC. UV spectroscopic characterization of type I collagen. *Coll Relat Res*. 1988;8(4):315–330.
47. Badea E, Miu L, Budrugeac P, Giurginca M, Mašić A, Badea N, Della Gatta G. Study of deterioration of historical parchments by various thermal analysis techniques complemented by SEM, FTIR, UV-vis-NIR and unilateral NMR investigations. *J Therm Anal Calorim*. 2008;91:17–27.
48. Zgarian RG, Tihan GT, Barbaresso RC, Rau I. Spectral characterization of some collagen based composite for dental application. *UPB Sci Bull Series B*. 2016;78(1):99–110.
49. Ghadam AGJ, Idrees M. Characterization of $CaCO_3$ nanoparticles synthesized by reverse microemulsion technique in different concentrations of surfactants. *Iran J Chem Chem Eng*. 2013;32(3):27–35.
50. Matei C, Berger D, Dumbrava A, Radu MD, Gheorghe E. Calcium carbonate as silver carrier in composite materials obtained in green seaweed extract with topical applications. *J Sol-Gel Sci Technol*. 2020;93:315–323.