



Pattern recognition techniques for provenance classification of archaeological ceramics using ultrasounds

Addisson Salazar^{a,*}, Gonzalo Safont^a, Luis Vergara^a, Enrique Vidal^b

^a Universitat Politècnica de València, Institute of Telecommunications and Multimedia Applications, Camino de Vera s/n, 46022 Valencia, Spain

^b Universitat Politècnica de València, PRHLT Research Center, Camino de Vera s/n, 46022 Valencia, Spain

ARTICLE INFO

Article history:

Received 15 July 2019

Revised 6 April 2020

Accepted 10 April 2020

Available online 21 May 2020

Keywords:

Decision fusion

Semi-supervised active learning

Feature ranking

Ultrasounds

Archaeological ceramics

Provenance

Classification

Pattern recognition

ABSTRACT

This paper presents a novel application of pattern recognition to the provenance classification of archaeological ceramics. This is a challenging problem for archaeologists, which involves assigning a making location to a fragment of archaeological pottery that was found along with other fragments of pieces made in different distant locations from the find. The pieces look very similar to each other and, often, other contextual information about the use of the pieces cannot be used due to the small size of the fragments. Current standard methods to solve this problem are limited since they are time consuming, require costly equipment, and can lead to the destruction of a part of the pieces. The proposed method overcome those limitations using non-destructive ultrasonic testing and incorporates versatile data analysis through advanced pattern recognition techniques. Those techniques include the following: feature ranking, sample augmentation, semi-supervision based on active learning; and optimal fusion. This latter is based in the concept of alpha integration, which allows optimal fitting of the fusion model parameters. Different provenance classification problems are showcased: provenance classification of *terra sigillata* ceramic pieces from Aretina, Northern Italy and Sud-Gaul origins; and provenance classification of Iberian ceramic pieces from archaeological sites of Paterna, and Les Jovaes in Valencia, Spain. We demonstrate that the proposed fusion-based method achieves the best results, in terms of balanced classification accuracy and F1 score, compared with competitive methods like linear discriminant analysis, random forest, and support vector machine. Experiments for simulating small sample sizes and uncertainty in labeling of the pieces are included. In addition, the paper provides a design of a practical specialized device that could be used in different applications of archaeological ceramic classification.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Classic methods of material characterization employed in archaeology can be divided in the following four categories: (i) Nuclear particle accelerator: proton induced X ray emission (PIXE), Rutherford backscattering spectroscopy (RBS), nuclear reaction analysis (NRA); (ii) Mineralogical analysis: binocular magnifying glass, thin sectioning, X-Ray Diffraction (XRD); (iii) Chemical analysis: scanning electron microscopy (SEM), atomic absorption spectrometry (AAS); and (iv) Physical analysis: porosity and density, mercury porosimetry; e.g. [1]. Contrary to non-destructive testing (NDT) methods, most of those techniques require some destruction of the analyzed materials, expensive equipment, and tube preparation, and thus, they are costly and time-consuming. Thus, we pro-

pose an ultrasound NDT method that involves evident practical advantages versus classic ones.

The objective of this paper is a method to classify the provenance of archaeological ceramic shards (broken pieces of pottery) dated in the same chronological period. The appearance of the pieces is very similar between them and it is not possible to use additional context information since the whole shape of the original ceramic object to which the pieces belonged is unknown. Besides, the pieces can be found together with other fragments of pieces made in distant places due to different circumstances that happened in time such as wars and migrations. Thus, identifying the provenance of archaeological ceramics can be a challenging pattern recognition problem, even for expert archaeologists [2]. We propose here a method to identify automatically the archaeological ceramic provenance of pieces drawn from different deposits. It is based on ultrasonic characterization of the ceramic material by estimating features from the measured ultrasonic signal to form a signature of the materials and process employed to make the

* Corresponding author.

E-mail address: asalazar@dcom.upv.es (A. Salazar).

ceramic piece. The ultrasonic signature consists of time, frequency, and statistical variables that are defined based on a material reflectivity model which is explained in Section 2. We hypothesize that the imprint of different proportions of raw materials and oven temperatures used to fire the ceramics should be revealed in the ultrasonic signature. Thus, we propose a method to process the ultrasound features extracted from a set of pieces considering classification based on the fusion of results from several classifiers. Moreover, the proposed method incorporates the following pattern recognition techniques: feature ranking [3]; sample augmentation [4]; semi-supervised active learning [5,6]; and optimal late fusion [7–9].

The single classifiers applied in the deployment of this work are the following: linear discriminant analysis (LDA), random forest (RF), and support vector machine (SVM). The mean of the posteriors from the single classifiers, a classic fusion method, is also calculated for comparison. We have incorporated semi-supervised learning based on active learning [10] in the classification stage to consider possible uncertainty of the expert in the labeling of the ceramic pieces. In addition, sample augmentation has been implemented to improve the classification performance facing the problem of data scarcity that could occur in real situations. Optimal fusion of classifiers is approached using a recent method called alpha integration, which allows optimal fitting of the fusion model parameters [7–9]. The indexes used to evaluate the classification results were accuracy, balanced accuracy (BAC), and F1 score. Besides the data analysis method, we include the design of an specialized unit for measuring ceramic pieces using NDT by ultrasounds adapted to the ceramic provenance classification problem, although it could be used in other related applications.

There are some references to using ultrasounds in archaeology for imaging and flaw detection [11,12]. Moreover, in archaeological material characterization, we can find the following applications of ultrasounds: examination of weathering processes and intensity acting of humidity and salt in the rocks of historical buildings [13] and determination of mechanical properties (Young's modulus and Poisson's ratio) of prehistory lithics [14]. Recently, an ultrasound-based method to classify archaeological ceramics by chronological dating using a single classifier [15] was proposed. The use of advanced pattern recognition techniques together with NDT techniques in archaeological provenance identification proposed in this work is a novel and practical contribution that might be used for museums, archaeologists, and archaeological ceramic restorers.

The rest of the paper is organized as follows. Section 2 describes the signal processing model, the features extracted from the ultrasonic signal, and a device to measure archaeological ceramics shards. Sections 3, 4, and 5 include the explanation of the pattern recognition method proposed for archaeological ceramics provenance classification. The experiments approached and results are included in Section 6. Finally, Section 7 contains the conclusions and future work derived from this research.

2. Features extracted from ultrasound signals

The measurement of ceramic pieces was made using through-transmission ultrasounds, i.e., an ultrasonic pulse is injected through an emitter transducer and the response of the material is collected by a receiver transducer. We have designed a practical specialized device that could be used in different applications of archaeological ceramic classification as explained in Section 2.1. The behavior of an ultrasonic pulse travelling inside a material is very complex, and thus difficult to be modeled. Ceramics are composite materials (clay, earthen elements...) where the ultrasonic waves suffer strong effects of attenuation, dispersion and backscattering. Instead of undertaking nonlinear wave propagation theory

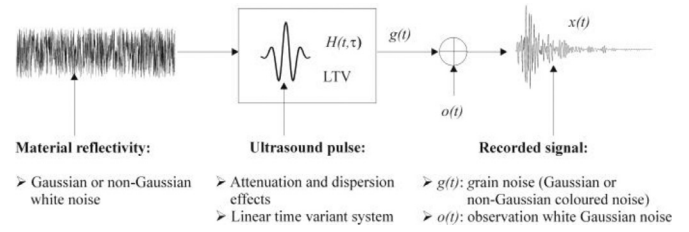


Fig. 1. LTV model for through-transmission ultrasound testing.

[16] to explain the differences in the recorded ultrasound signal, and thus differences in ceramic provenance, we devise a linear time variant (LTV) model, see Fig. 1 [17].

This simple model may assume that many small reflectors (reflectivity of the material) form the material. When the ultrasonic pulse affects a reflector, a new pulse is generated. The recorded signal would be the superposition of the many pulses thus generated by the material microstructure. This essentially relates to considering a linear system model where the input is the reflectivity of the material and the impulse response is the ultrasonic pulse excitation. Hence the output of the linear system (recorded sensor signal) would be the convolution of the reflectivity and the excitation pulse. In the simplest case, the pulse (impulse response) may be considered to be constant, which means that the pulse arriving at every reflector of the microstructure is always the same and that the only effect of the reflector is to generate a new replica with given amplitude, depending on the reflector size. This corresponds to a strictly linear time invariant (LTI) system in the context of system theory. More complex models may be gradually considered depending on the complexity of the material. Thus, a frequency independent attenuation of the pulse can be easily introduced into the model by assuming a LTV system. This can be generalized to frequency dependent attenuation by allowing a general distortion of the pulse as it propagates deeply into the material. The model explained above grounds the extraction of a set of features to relate the changes of the signal to differences of material so that they can be classified in a finite number of classes (see Table 1).

Table 1 includes features from several domains that define different ultrasound signatures from the inspected material. Time-domain features 1 to 4 correspond to the parameters A and β of an exponential model $\hat{x}(t) = Ae^{-\beta t}$ of the signal attenuation, the total signal power received in the receiver transducer $P = \int_0^T |x(t)|^2 dt / T$, and the propagation velocity of the measured signal. Frequency-domain features 5 to 8 (principal frequency, principal frequency amplitude, centroid frequency, and instantaneous centroid frequency

Table 1
Features extracted from ultrasound signal.

| No. | Feature | Definition |
|-----|--------------------------------------|---|
| 1 | Total signal attenuation | $\beta / \hat{x}(t) = Ae^{-\beta t}$ |
| 2 | Propagation velocity | $v = \frac{\text{piezethickness}}{\text{ultrasoundtimeofflight}}$ |
| 3 | Signal power | $P = \frac{\int_0^T x(t) ^2 dt}{T}$ |
| 4 | Attenuation curve initial value (dB) | $P_0 = 10 \log(A) / \hat{x}(t) = Ae^{-\beta t}$ |
| 5 | Principal frequency | $f_{\max} / X(f_{\max}) \geq X(f) \forall f$ |
| 6 | Principal frequency amplitude | $ X(f_{\max}) $ |
| 7 | Centroid frequency | $f_c = \frac{\int_0^T f \cdot X(f) df}{\int_0^T X(f) df}$ |
| 8 | Instantaneous centroid frequency | $f_c(t = t_0)$ |
| 9 | Time-reversibility | $\frac{1}{\sigma_x^2} \left(\left(\frac{dx(t)}{dt} \right)^2 \right)$ |
| 10 | Third order autocovariance | $\langle x(t) \cdot x(t-1) \cdot x(t-2) \rangle$ |

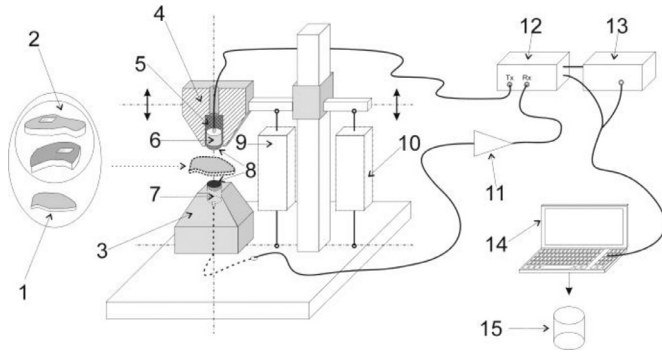


Fig. 2. Measurement device for ultrasound testing of archaeological ceramics pieces.

quency) are measures of the spectral content variations affected by the ultrasonic pulse travelling inside the material. They can be estimated by means of well-known smoothing techniques of time-frequency spectral analysis. Statistical-domain features can be used to consider special conditions of the through-transmission model. For example, higher-order statistics might detect the possible degree of non-Gaussianity of the material reflectivity. We use the so-called time-reversibility and the third order covariance (features 9 and 10 of Table 1) to measure departures from the linear model of Fig. 1.

2.1. Design of a measurement device for ultrasound measurement of archaeological ceramics pieces

Fig. 2 outlines a specialized device designed for measuring archaeological ceramics pieces using ultrasounds. The elements of the device are described below:

- Two housings (3) (4) to accommodate the transmitting and receiving ultrasound transducers (6) (7). A damping system (5) in each of the housings normalizes the pressure exerted on the transducers in each measurement. A cut of housing (4) in Fig. 2 shows a view of its components.
- Two ultrasound transducers: emitter (6) and receiver (7), associated with the emitter/receiver ultrasound card (12).
- A coupling material (8) for impedance coupling of the ultrasound transducers (6) (7) and the ceramic material (unlabeled (1), or labeled pieces (2)).
- A longitudinal measuring system (9) attached to the approach system (10) and housings (3) (4) to measure the thickness of the piece being evaluated.
- An approach system (10) that allows moving the assembly of housing (4) and measuring indicator of the longitudinal measuring system (9) to adjust them to the piece thickness.
- A preamplifier (11) between the receiver transducer (7) and the ultrasound card (12) to amplify the captured signal.
- Digitizing equipment (13), commonly an oscilloscope, that digitize the signals received by the receiver transducer (7).
- A processor (14) that processes the digitized signals, extracting and storing an ultrasonic signature that will be used to assign the ceramic piece provenance (15).

The ultrasound transducers will be 5–15 mm. diameter, for adapting to the small dimensions of the ceramic fragments. Usually, ceramic pieces are 3–15 mm. thicknesses and surface areas with irregular shapes and curvatures. Therefore, the coupling material must be flexible to guarantee the adaptation of the ultrasound transducers to the physical characteristics of the pieces and obtain a good signal-to-noise ratio in the transmitted signal. We use a piece of rubber as a coupling media. The ultrasound equip-

ment performs using a nominal working-frequency emitting transducer that generates an ultrasonic signal of sufficient wavelength to capture the material microstructure response. Normally, a working frequency in 1–5 MHz range is adequate for archaeological ceramics.

3. Basic classification methods

Sections 3 to 5 contain the description of the pattern recognition techniques implemented in the proposed method. Section 3 describes notation and feature selection and data augmentation algorithms. Sections 4 and 5 describe the semi-supervised active learning and optimal fusion algorithms.

We will assume the observations are denoted by $\mathbf{x}^j = [x_1^j, x_2^j, \dots, x_M^j]^T$, with M being the number of features and $j = 1 \dots N$ denoting the j th observation. The classification cases are two-class, therefore the labels y^j can be either 0 or 1.

As commented above, the single classifiers considered were LDA, RF, and SVM. These classifiers were chosen because of their successful application across many different fields and problems, including NDT. They are based on different and sometimes complementary principles and that fusion can be used to exploit such a complementarity. RF implemented 50 trees and SVM employed a Gaussian kernel. Furthermore, each of these classifiers is based on a different set of assumptions. LDA divides the space between the classes using a single hyperplane; SVM divides the space between the classes using a more complex hyperplane in a kernel space; and RF is based on the combination of a set of relatively simple classifiers that divide the space based on splits of the input features. We assume that those different solutions could be complementary; thus, an optimal fusion algorithm could yield a combined result that would improve over the results of the single classifiers. We also considered two decision fusion methods: the mean of the posteriors provided by each classifier (Mean), and alpha integration (α -INT) using the least mean squared error (LMSE) criterion [7–9].

We used the 10 features of Table 1, i.e., a 10-dimensional space for classification. We have implemented a ranker method ([3]) for feature selection instead of the classical principal component analysis (PCA). PCA involves whitening and Gaussianization of the data, which makes a possible interpretation of the results difficult. Moreover, isolating the noise in the lower order components can mask the particular classification power that some features may have. In addition, we tested in a rather experimental manner that the ranker yielded better results over PCA for this application. One could use more computationally-expensive feature selection methods (e.g., exhaustive search), given a low number of features as the proposed here in time, frequency, and statistical domains.

If we made an in depth analysis in time-frequency domain, we might approach the wavelet time-frequency decomposition where the number of features could be, for instance, as large as 100 or more. In that case, an exhaustive search would be unfeasible or impractical. For our proposed ranker method, as the number of features increases, the number of evaluations increases linearly. For the exhaustive approach, as the number of features increases, the number of evaluations increases exponentially. Dimensionality reduction is hence required, in particular considering overfitting prevention in the probable case of data scarcity in the provenance classification problem.

The algorithm followed for feature selection by ranking is shown in Algorithm 1. Briefly, the score of each input feature for classification is determined using a threshold-based classifier (e.g., a shallow decision tree), and features with a score below a empirically-set threshold are discarded. In this work, we considered a fixed threshold of 0.2.

Another data processing technique selected was sample data augmentation using synthetic replicates obtained by adding Gaus-

Algorithm 1 Ranker method for feature selection.

```

0  Given a set of features from  $N_{TRAIN}$  observations  $\mathbf{x}^j, j = 1 \dots N_{TRAIN}$ , and their associated labels  $y^j$ 
1  For each feature  $m = 1 \dots M$ 
2  Estimate a classifier based only on the  $m$ th feature,  $x_m^j$ 
3  Estimate the receiver operating characteristic (ROC) curve of the estimated classifier using  $x_m^j$  and  $y^j, j = 1 \dots N_{TRAIN}$ 
4  Compute the score of the  $m$ th feature as the maximum difference between the probability of detection and the probability of false alarm of the ROC curve
5  Sort (rank) features in order of descending score
6  Return the sorted scores and the rank of each feature

```

sian noise to an arbitrary number of samples from the available data [4]. This produces a smoothed version of the estimator, removing false minima when the sample size is small. We considered a noise variance of 0.1 times the average variance of the features.

4. Semi-supervised active learning (SSAL)

We were focused on analyzing the behavior of the proposed method with respect to the number of pieces available for training. Currently, an expert labels each piece by hand, a time-consuming task that requires expert knowledge. Therefore, it would be interesting to determine whether the proposed method is able to reduce the number of required labeled pieces in order to achieve a good classification. Thus, it would also be interesting to determine how many labeled pieces are necessary to reach a given performance. This represents a common real situation when the number of pieces found in the archaeological deposit is scarce. In addition, this grounds the proposal of using relatively simple classifiers with non-extensive data requirements. Furthermore, in real situations, data acquisition can be costly or unavailable, and thus, it is practical to minimize the amount of analyzed pieces. The device designed, however, helps to make the cost of acquiring new (unlabeled) data much lower than the cost of labeling the pieces. This scenario is suited to semi-supervised training, where the number of pieces is not so small but only some pieces are labeled. Thus, we could use the information provided by unlabeled pieces to improve classification results over that of purely supervised training.

We consider SSAL to simulate a situation of uncertainty or difficulty in labeling the data by the archaeologist. Active learning is an approach that can be used to assign iteratively unknown labels to an unlabeled dataset based on a labeled dataset and an information criterion [10]. The implemented SSAL algorithm is shown in Algorithm 2. Briefly, active learning starts with a randomly-chosen small amount of labeled pieces (the seed) and iteratively asks the expert to provide labels on the most informative unlabeled pieces, i.e., those unlabeled pieces where the classifier is less confident about its decision. Thus, the method is able to determine which pieces can be confidently classified using previously labeled pieces (no label required) and which pieces cannot be classified with confidence (a label would improve the result). In Section 6, we will prove that is possible to decrease the number of labeled pieces by using labeled and unlabeled together to obtain the best classification results.

5. Optimal fusion of classifiers

The proposed fusion method is founded on optimal fusion by alpha integration [7–9], a recent technique that has been successfully implemented in several applications. The objective is to improve the classification results by fusion of the scores (posterior probabilities) from single classifiers, i.e., optimally integrating those scores into a unique score $s_\alpha(\mathbf{s}^j)$:

$$s_\alpha(\mathbf{s}^j) = \begin{cases} \left[\sum_{i=1}^D w_i (s_i^j)^{(1-\alpha)/2} \right]^{2/(1-\alpha)} & , \alpha \neq 1 \\ \exp \left[\sum_{i=1}^D w_i \log(s_i^j) \right] & , \alpha = 1 \end{cases}, \quad j = 1 \dots N_{TRAIN}$$

Algorithm 2 Semi-supervised active learning (SSAL).

```

0  Given a set of features from  $N_{TRAIN}$  observations  $\mathbf{x}^j, j = 1 \dots N_{TRAIN}$ , and their associated labels  $y^j$ ; the number of pieces to consider at each step,  $q \leq N_{TRAIN}$ ; a classifier  $C$ ; and a set of  $N_{TEST}$  testing observations  $\hat{\mathbf{x}}^j, j = 1 \dots N_{TEST}$ , with their associated labels  $\hat{y}^j$ 
1  Initialize the set of labeled observations,  $L$ , by randomly subsampling  $m$  observations; the rest of the observations become the set of unlabeled observations,  $U$ 
2  Use the labeled pieces  $L$  to train the classifier  $C$ 
3  Apply the trained classifier  $C$  on the testing observations and determine performance
4  Apply the trained classifier  $C$  on the unlabeled observations  $U$  and obtain the posterior probabilities  $[1 - s^j, s^j], j \in U$ , where  $s^j$  is the probability that sample  $j$  belongs to class 1
5  For each unlabeled piece in  $U$ 
6  Calculate the entropy of the associated posterior probabilities  $e^j = -(1 - s^j) \log(1 - s^j) - s^j \log(s^j)$ 
7  Choose the  $q$  unlabeled observations with highest entropy,  $U'$ , and obtain their labels
8  Update the set of labeled observations to include the selected observations,  $L = L \cup U'$ 
9  Update the set of unlabeled observations to remove the selected observations,  $U = U - U'$ 
10 If there are still observations in  $U$ , repeat from step 2
11 Return the chosen pieces and the performance on the testing set at every iteration

```

Algorithm 3 Training alpha integration using gradient descent.

```

0  Given a set of scores from  $D$  classifiers for  $N_{TRAIN}$  archaeological ceramic pieces  $\mathbf{s}^j = [s_1^j \dots s_D^j]^T, i = 1 \dots D, j = 1 \dots N_{TRAIN}$ , and their associated labels  $y^j$ ; the learning rates  $\eta_w, \eta_\alpha$ ; and the starting values for the weights  $\mathbf{w}$  and the alpha value  $\alpha$ 
1  Apply alpha integration on the input scores using (1)
2  Determine the value of the LMSE cost function
3  Determine the value of the derivatives of the LMSE cost function using Eqs. (2) and (3)
4  Update the values of the weights,  $w_i = w_i + \eta_w \cdot \partial \varepsilon / \partial w_i, i = 1 \dots D$ , and alpha,  $\alpha = \alpha + \eta_\alpha \cdot \partial \varepsilon / \partial \alpha$ 
5  Until convergence, repeat from step 1
6  Return the final values of the weights  $\mathbf{w}$  and  $\alpha$ 

```

(1)

where $\mathbf{s}^j = [s_1^j \dots s_D^j]$ is a vector that contains the scores yielded by the D classifiers, and the parameters w_i and α must be learned from training data. The parameters of alpha integration were optimized to satisfy the LMSE criterion, $\varepsilon = \frac{1}{N_{TRAIN}} \sum_{j=1}^{N_{TRAIN}} (y^j - s_\alpha(\mathbf{s}^j))^2$.

The gradient descent algorithm considered to train alpha integration is shown, for completeness, in Algorithm 3. The derivatives of LMSE with respect to the parameters of alpha integration are the following:

$$\frac{\partial \varepsilon}{\partial w_i} = \frac{2}{N_{TRAIN}} \sum_{j=1}^{N_{TRAIN}} (y^j - s_\alpha(\mathbf{s}^j)) \begin{cases} \frac{2}{1-\alpha} \left(\frac{s_\alpha(\mathbf{s}^j) (s_i^j)^{(1-\alpha)/2}}{\sum_{l=1}^D w_l (s_l^j)^{(1-\alpha)/2}} \right) & , \alpha \neq 1 \\ s_\alpha(\mathbf{s}^j) \log(s_i^j) & , \alpha = 1 \end{cases} \quad (2)$$

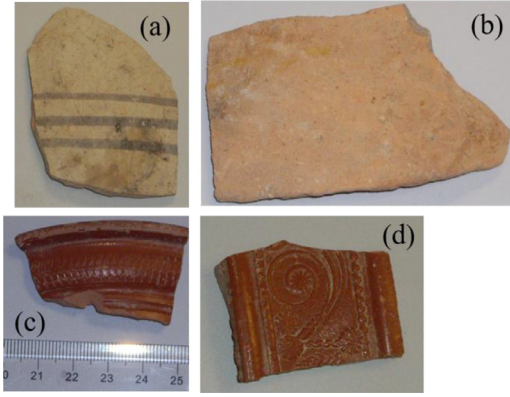


Fig. 3. Examples of the archaeological ceramic pieces: (a) Paterna; (b) Les Jovaes; (c) TSA; (d) TSSG.

$$\frac{\partial \mathcal{E}}{\partial \alpha} = \frac{-4}{N_{\text{TRAIN}}} \sum_{j=1}^{N_{\text{TRAIN}}} (y^j - s_{\alpha}(\mathbf{s}^j)) s_{\alpha}(\mathbf{s}^j) \quad (3)$$

$$\left[\frac{\log \left(\frac{\sum_{i=1}^D w_i (s_i^j)^{(1-\alpha)/2}}{1-\alpha} \right)}{1-\alpha} - \frac{\sum_{i=1}^D w_i \log(s_i^j) (s_i^j)^{(1-\alpha)/2}}{2 \sum_{i=1}^D w_i (s_i^j)^{(1-\alpha)/2}} \right]$$

6. Experiments

A total number of 945 archaeological ceramic pieces were available for the experiments. These pieces belonged to four classes, depending on the origin of the pieces: (i) 194 *terra sigillata* pieces from Aretina, Northern Italy (TSA); (ii) 200 *terra sigillata* pieces from Sud-Gaul origin (TSSG); (iii) 329 Iberian ceramic pieces from an archaeological site in Paterna, Valencia (Spain); and (iv) 222 Iberian ceramic pieces from an archaeological site in Les Jovaes, Valencia (Spain). Fig. 3 shows some specimens of the archaeological ceramic pieces analyzed. It can be seen that the visual appearance of the pieces does not allow to discern their provenance. Both *sigillata* ceramic pieces (Aretina and Sud-Gaul) are from Roman period and both Iberian pieces (Paterna and Les Jovaes) are from Iberian period.

6.1. Experimental setup

As was commented in Section 2, pieces were measured in transmission mode using a rubber adaptor. This coupling method was selected due to its good ultrasonic transmission and its lack of reactivity with the ceramic pieces, in comparison with methods such as water immersion and direct contact using coupling gel. The latter offers very good coupling, but the time available for measuring must be very short to avoid gel absorption by the piece. The pieces were excited using Gaussian ultrasonic pulses with a pulse width of 4 μs and a central frequency of 1.05 MHz, and the received signal was captured at a sampling rate of 500 MHz. Fig. 4 shows the ultrasonic signals extracted from the pieces of Fig. 3. To separate the pieces according to their provenance, the features in Table 1 were extracted from each signal. Since one set of features was extracted from each piece, in the following, we use “piece” and “sample” interchangeably.

6.2. Problem #1: *terra sigillata* versus non-*terra sigillata* pieces

In this problem, we considered whether the extracted features were enough to separate the available archaeological pieces into two groups: *terra sigillata* pieces (TSA, TSSG) versus non-*terra sigillata* pieces from Paterna and Les Jovaes. The ranker presented in Section 3 was used to determine the optimal features for the

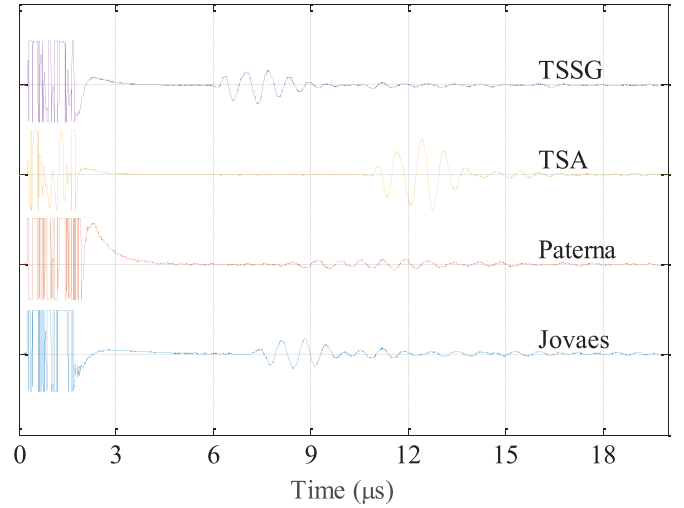


Fig. 4. Examples of ultrasonic signals from the pieces of Fig. 3.

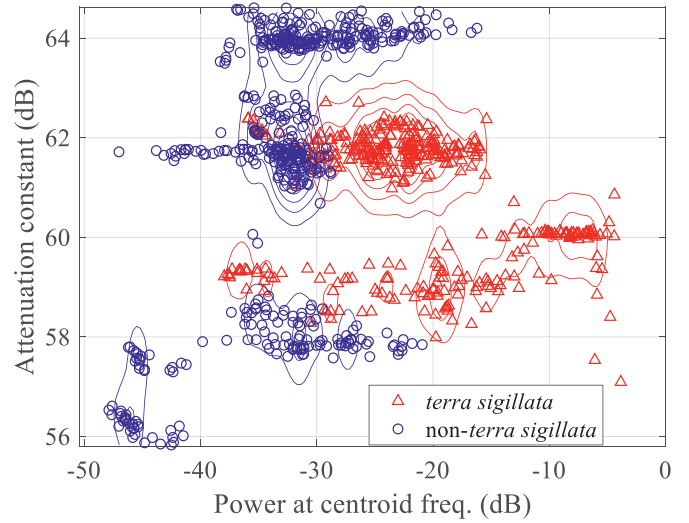


Fig. 5. Comparison of two feature values from *terra sigillata* versus non-*terra sigillata* pieces. Continuous lines mark areas with equal density.

experiment. Using cross-validation, it was determined that all 10 features presented in Table 1 were necessary for classification. Fig. 5 shows the distribution of each group of ceramics on the two highest-ranking features for this problem: signal power, centroid frequency, and attenuation constant. It can be seen that there is a fairly clear separation between *terra sigillata* pieces and the other pieces.

As commented above, LDA, RF, and SVM were considered as single classifiers. We also considered two decision fusion methods: the Mean and alpha integration (α -INT) using the LMSE criterion [7–9]. Furthermore, we also compared our results with those obtained using the independent component analysis mixture modeling (ICAMM), which is a competitive state of the art method proposed for chronological classification of archaeological ceramic shards [15]. One of the main advantages of the proposed ultrasound analysis is the time it requires. The ultrasound analysis for the complete experiment (measuring, processing, training, and automatic classification) took only about 10 hours. Conversely, given our prior experiences with classic techniques [15], SEM analyses (tube preparation and electron microscope analysis) for 120 representative pieces would have taken a bit over 400 hours. Likewise, porosity and density analyses for 120 representative pieces

Table 2

Average classification performance for the split between *terra sigillata* and non-*terra sigillata* pieces (values are percentages). The best final classification results are highlighted.

| Classifier | BAC | std. | F ₁ | std. | Acc | std. |
|---------------|--------------|-------------|----------------|-------------|--------------|-------------|
| LDA | 87.39 | 0.32 | 89.27 | 0.28 | 88.01 | 0.32 |
| RF | 95.83 | 0.23 | 96.78 | 0.18 | 96.32 | 0.20 |
| SVM | 94.92 | 0.30 | 96.38 | 0.20 | 94.93 | 0.33 |
| Mean | 95.54 | 0.26 | 96.49 | 0.21 | 96.01 | 0.23 |
| α -INT | 96.34 | 0.22 | 97.13 | 0.17 | 96.75 | 0.20 |
| ICAMM | 92.48 | 0.33 | 94.16 | 0.25 | 93.04 | 0.30 |

would have taken almost 450 hours. Furthermore, the equipment required for nondestructive evaluation by ultrasound is, in general, less costly, and experiments are easier to perform. Given these large differences in cost and the heritage nature of the archaeological pieces, in this work, no destructive analyses were performed. Thus, the pieces were not damaged in any way during testing and they remain intact for further analysis.

In order to verify the classification performance of the considered methods, a series of Monte Carlo experiments were run. For each Monte Carlo experiment, the pieces were separated equally into three sets: training, validation, and testing. The training pieces were used to train the single classifiers, and α -INT was trained using the scores of the trained single classifiers on the validation pieces. Finally, performance was estimated on the testing pieces. Three performance indicators were calculated: the classification accuracy (Acc); the balanced accuracy (BAC) defined as the average accuracy for each class; and the F1 score (F1), i.e., the harmonic average of precision and recall. The results were obtained as the average of 100 Monte Carlo experiments. We considered several statistical tests on the results of the Monte Carlo experiments. First, the Kolmogorov–Smirnov test has been considered to verify that the results were Normally-distributed ($p < 0.05$). Then, given the Normal distribution and the number of available samples, we considered one-way analyses of variance (ANOVA) and nonparametric Kruskal–Wallis ANOVAs to determine whether the differences in outcome were statistically significant.

Table 2 shows the average results of the experiment. Except for LDA, classifiers and fusions yielded very high results of separation between both groups of archaeological ceramic pieces, with around 97% in all indicators. The best performance was obtained by α -INT, which yielded the smallest error variance (standard deviation, std.). The difference with RF was not statistically significant, but both RF and α -INT performed significantly better than the rest of the methods. Conversely, the worst performance was yielded by LDA. Given that the decision boundaries provided by LDA are linear, unlike those of RF and SVM, this difference in performance might indicate nonlinear optimal boundaries between classes. This fits with the distribution of the values in Fig. 5, where there is no clear linear boundary that could split the two groups.

In addition, we performed an experiment where the number of pieces available for training was progressively reduced from 316 ($n=N_{\text{TRAIN}}$) pieces (all the pieces in the training set) to 5 ($n=5$) pieces (1.58% of the pieces in the training set). Furthermore, we also considered the addition of synthetic sample replicates to compensate for the missing training pieces, keeping constant the original prior probability of each of the classes. Every removed training piece was replaced by a synthetic sample replicate by adding white Gaussian noise to one sample randomly chosen from the labeled training n pieces, see Section 3. Other than these changes to the training set, we considered the same methods and number of iterations of Monte Carlo runs as the previous experiment. The average results of this experiment are shown in Fig. 6. Standard deviation was lower than 1.7 and 2.9 for Fig. 6a and b, respectively.

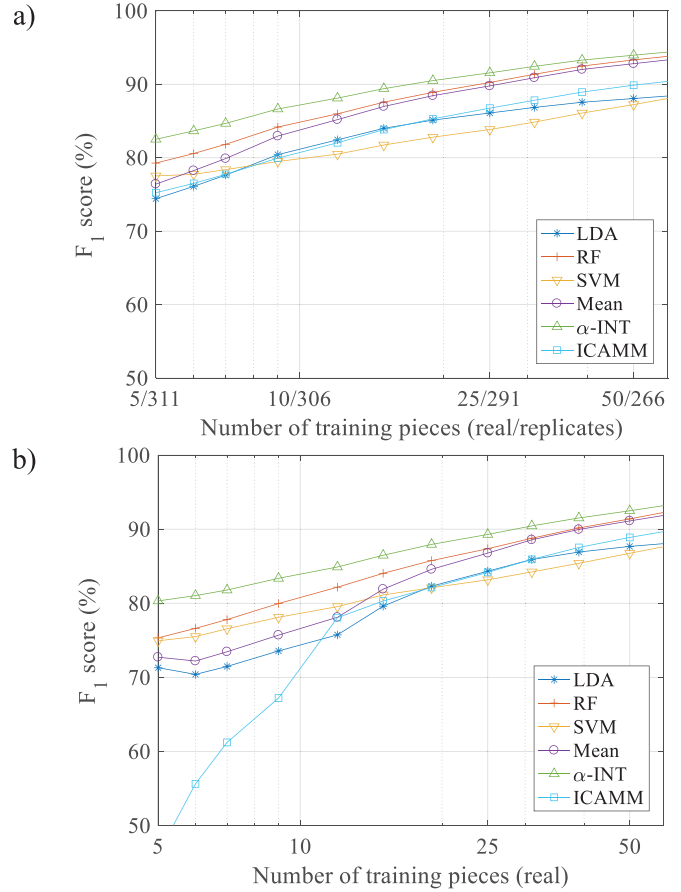


Fig. 6. Average classification for *terra sigillata* and non-*terra sigillata* pieces with a decreasing number of labeled training pieces: (a) F1 score after adding replicates; (b) F1 score before adding replicates. X axes are log scale.

For brevity, only the results for F1 score are shown; the results for Acc, BAC, and F1 were consistent. Furthermore, since the effect of the synthetic replicates was more noticeable for small numbers of training pieces, only the values up to 50 real training pieces are shown; above that threshold, the synthetic data did not affect performance. In all cases, performance decreased as the number of real training pieces decreased. This drop was more noticeable for LDA and especially ICAMM before the addition of replicates, as shown in Fig. 6b. In turn, both methods were the ones most improved by the synthetic replicates, which lessened the drop in performance that both methods experienced when using less than 25 training pieces (compare Fig. 6b with a). α -INT always yielded the best performance, with statistically significant improvements over the rest of the methods until 25 training pieces were used.

At any rate, all methods improved in performance with the addition of surrogate replicates for small numbers of training pieces (less than 50 pieces). SVM was the least affected by the synthetic replicates. α -INT yielded the best result for any number of labeled pieces for training, in accordance with the results for the case with all pieces (see Table 2), and yielded a more robust result, in terms of standard deviation. This first experiment is the simplest from a class separability standpoint, i.e., the cases of only 50 labeled pieces for training, Mean was the best option. Actually, as Mean is a particular case of α -INT, this later should yield similar results to ones of Mean. However, α -INT requires parameter estimation, which implies some degradation with respect to directly using the mean. The next experiments are more difficult and we will see that α -INT is the best option in all cases, even considering the mentioned degradation due to the required parameter estimation.

Table 3

Performance improvement using SSAL for the split between *terra sigillata* and non-*terra sigillata* pieces. The values indicate the base result obtained by the classifier using supervised learning (+/-) the improvement obtained over base result using SSAL. Two ratios of Real / Replicate pieces are shown: 50/316 and 100/316. The best final classification results are highlighted.

| Classifier | BAC (%) | | F1 (%) | | Acc (%) | |
|---------------|-------------------|-------------------|-------------------|-------------------|-----------------|-----------------|
| | 50/266 | 100/216 | 50/266 | 100/216 | 50/266 | 100/216 |
| LDA | 85.59-0.03 | 87.08-0.03 | 87.90+0.04 | 89.15+0.15 | 86.7+0.0 | 88.1+0.1 |
| RF | 91.24+1.30 | 94.55+1.05 | 92.63+0.95 | 95.67+0.87 | 91.9+1.2 | 95.1+1.0 |
| SVM | 81.51+0.89 | 87.89+1.02 | 86.71-0.28 | 91.58+0.57 | 84.1+0.3 | 89.7+0.8 |
| Mean | 91.29+1.60 | 94.22+1.40 | 92.79+1.26 | 95.33+1.22 | 92.0+1.5 | 94.8+1.4 |
| α -INT | 90.77+0.56 | 95.29+1.74 | 91.87+0.05 | 96.56+1.80 | 91.3+0.5 | 95.9+2.0 |
| ICAMM | 85.77+0.23 | 89.87+0.41 | 88.82-0.17 | 92.11+0.23 | 87.3-0.0 | 91.0+0.3 |

Another way to quantify the effect of the experiment is by determining the minimum amount of pieces required to reach a given performance level. If we set a minimum performance level of 80% F1 score, LDA required 18 pieces (10 with replicates); RF required 10 pieces (6 with replicates); SVM required 15 pieces (13 with replicates); the Mean required 16 pieces (9 with replicates); α -INT required 10 pieces (5 with replicates); and ICAMM required 17 pieces (11 with replicates).

In the previous experiment, we performed supervised training after assuming that only part of the training pieces was available. This experiment simulated a case where data acquisition is costly or unavailable, and thus, it would be interesting to minimize the amount of analyzed pieces. Furthermore, we implemented another experiment incorporating SSAL using the algorithm explained in Section 4. In practice, this experiment was similar to the previous one with supervised training, but instead of selecting random training pieces, the labeled pieces were selected by active learning. This experiment could be related to the difficulty or uncertainty of labeling certain ceramic pieces by the archaeologist. SSAL increased performance for relative large amounts of labeled training pieces (above 50) without affecting performance for small amounts of labeled training pieces. The improvements obtained by using SSA are summarized in Table 3.

The improvement was larger for Mean and α -INT, but all methods benefited from active learning. Furthermore, this improvement caused the maximum performance (previously obtained with the full labeled training set) to be reached with a small amount of labeled training pieces. Using semi-supervised learning, the considered methods required between 100 and 250 training pieces, and α -INT was able to reach the best performance with only 125 labeled training pieces. These results indicate that the proposed method was able to distinguish *terra sigillata* ceramic pieces from pieces having different provenance. Thus, the estimated ultrasonic signature allows material characterization to distinguish different making processes. In the following, this problem is split into two other problems where we classify the pieces by their provenance (i) separating pieces from the Paterna site from pieces from the Les Jovaes site; and (ii) separating TSA pieces from TSSG pieces.

6.3. Problem #2: Paterna vs. Les Jovaes

In this problem, we considered the separation of archaeological ceramic pieces from the Paterna site from those from the Les Jovaes site, using the same methods and experiment design explained in Section 6.2. In this case, the ranking method resulted in the removal of the two lowest-ranking features (time reversibility and principal frequency) before classification. The results of the experiment are shown in Table 4.

The best performance was obtained by α -INT, with an average 88.40% BAC. α -INT performed significantly better than the second-best method, the Mean. Compared with the results of Table 2, it would seem that separating pieces by their archeological site

Table 4

Average classification performance for the experiment on ceramic pieces from Paterna and Les Jovaes (values are percentages). The best final classification results are highlighted.

| Classifier | BAC | std. | F ₁ | std. | Acc | std. |
|---------------|--------------|-------------|----------------|-------------|--------------|-------------|
| LDA | 82.25 | 0.52 | 83.51 | 0.50 | 81.57 | 0.48 |
| RF | 86.12 | 0.60 | 87.69 | 0.46 | 85.57 | 0.49 |
| SVM | 86.27 | 0.50 | 87.17 | 0.43 | 85.55 | 0.46 |
| Mean | 86.66 | 0.51 | 87.73 | 0.44 | 87.41 | 0.46 |
| α -INT | 88.40 | 0.48 | 89.21 | 0.43 | 88.01 | 0.46 |
| ICAMM | 83.52 | 0.64 | 84.68 | 0.50 | 82.46 | 0.65 |

(Paterna versus Les Jovaes) was harder than separating pieces by whether they are *terra sigillata* or not. Despite the increased difficulty, the proposed method was still able to correctly separate most of the archaeological ceramic pieces between the two archaeological sites. As in Section 6.2, the performance of the proposed method with respect to the number of available training pieces was also considered. The number of available pieces for training was progressively reduced from 184 (the total number of pieces in the training set) to 5 (2.72% of the total number of pieces in the training set). When using synthetic data samples, the features for every missing training piece were replaced with synthetic replicates of labeled training pieces. The results of this experiment are shown in Fig. 7. Standard deviation was lower than 2 and 2.8 for Fig. 7a and b, respectively. As explained for Fig. 6, only the results for the F1 score up to 50 real training pieces are shown. In all cases, performance decreased as the number of real training pieces decreased. RF and ICAMM were the methods most affected by the decrease in the number of real training pieces, and α -INT yielded the best performance before (Fig. 7b) and after (Fig. 7a) adding synthetic replicates, and the difference was statistically significant in all cases.

The addition of synthetic data increased the overall performance of the methods for small numbers of labeled training pieces and did not decrease performance for larger numbers of real labeled training pieces, and yielded more stable results. If we set a minimum performance level of 80% F1 score, RF and LDA required an average of 47 labeled training pieces, and SVM required 24 training pieces. The Mean required an average number of 30 labeled training pieces, and α -INT only required 15 labeled training pieces. The addition of surrogate pieces resulted in a general decrease in the number of pieces required to reach the minimum performance level of 80% F1 score. This effect was more noticeable for RF, which reached the minimum performance with only 24 actual training pieces, and α -INT, which reached the minimum performance level with only 10 training pieces. We also considered the effect of semi-supervised learning, on the performance of the proposed method. The average improvement for a given amount of labeled training pieces is summarized in Table 5.

Improvement due to semi-supervised learning was larger for RF, but all methods benefited from active learning. Furthermore, this

Table 5
Performance improvement using SSAL for the split between Paterna and Les Jovaes pieces. The values indicate the base result obtained by the classifier using supervised learning (+/-) the improvement obtained over base result using SSAL. Two ratios of Real / Replicate pieces are shown: 50/184 and 100/184. The best final classification results are highlighted.

| Classifier | BAC (%) | | F1 (%) | | Acc (%) | |
|---------------|-------------------|-------------------|-------------------|-------------------|-----------------|-----------------|
| | 50/134 | 100/84 | 50/134 | 100/84 | 50/134 | 100/84 |
| LDA | 82.23+0.50 | 82.87+0.65 | 83.71+0.51 | 84.38+0.85 | 83.0-0.9 | 83.6-0.0 |
| RF | 85.14+2.02 | 86.44+1.10 | 86.86+1.26 | 88.14+0.97 | 86.0+0.2 | 87.3+0.3 |
| SVM | 84.71+0.89 | 86.31+0.65 | 86.06+0.03 | 87.55+0.61 | 85.4+0.4 | 86.9+0.6 |
| Mean | 86.49+1.31 | 87.53+0.95 | 87.70+0.95 | 88.70+0.99 | 87.1+0.2 | 88.1+0.5 |
| α -INT | 88.44+1.60 | 89.13+1.13 | 89.09+1.04 | 89.94+1.00 | 88.8+0.4 | 89.5+0.5 |
| ICAMM | 72.92+0.96 | 76.53+0.87 | 76.56-1.51 | 79.45-0.53 | 74.7-2.1 | 78.0+0.6 |

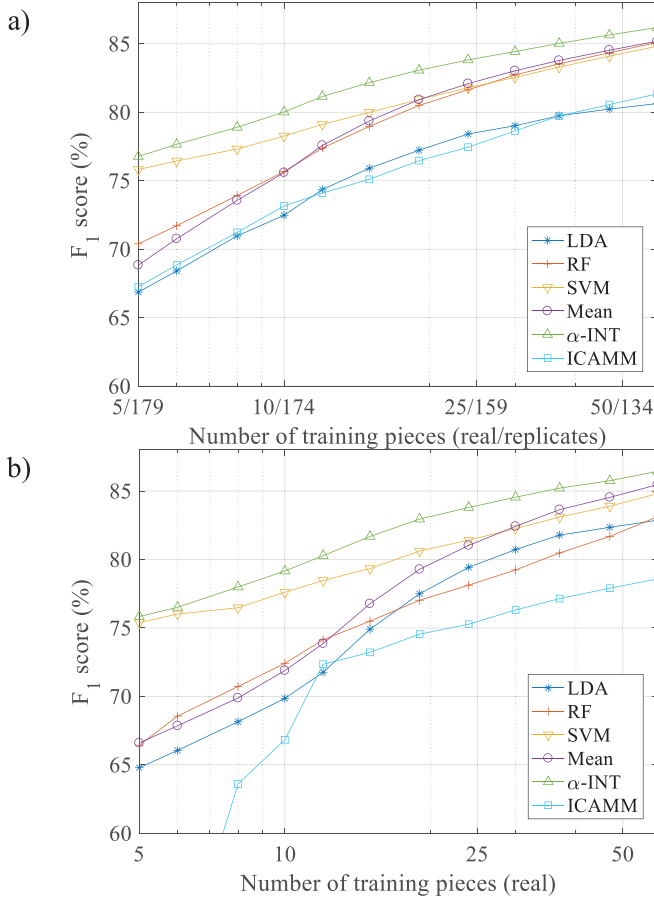


Fig. 7. Average classification for Paterna and Les Jovaes pieces with a decreasing number of labeled training pieces: (a) F1 score after adding replicates; (b) (d) F1 score before adding replicates. X axes are log scale.

improvement caused the maximum performance (previously obtained with the full labeled training set) was reached with a small amount of labeled training pieces. Using semi-supervised learning, the considered methods required less than 100 instead of 184 labeled pieces, with α -INT being able to reach the best performance with only 75 labeled training pieces.

6.4. Problem #3: TSA vs. TSSG

This problem consisted in separating *terra sigillata* archaeological ceramic pieces from Aretina from those of Sud-Gaul origin. We followed the procedure explained in Section 6.3. In this case, the ranking method resulted in the removal of the two lowest-

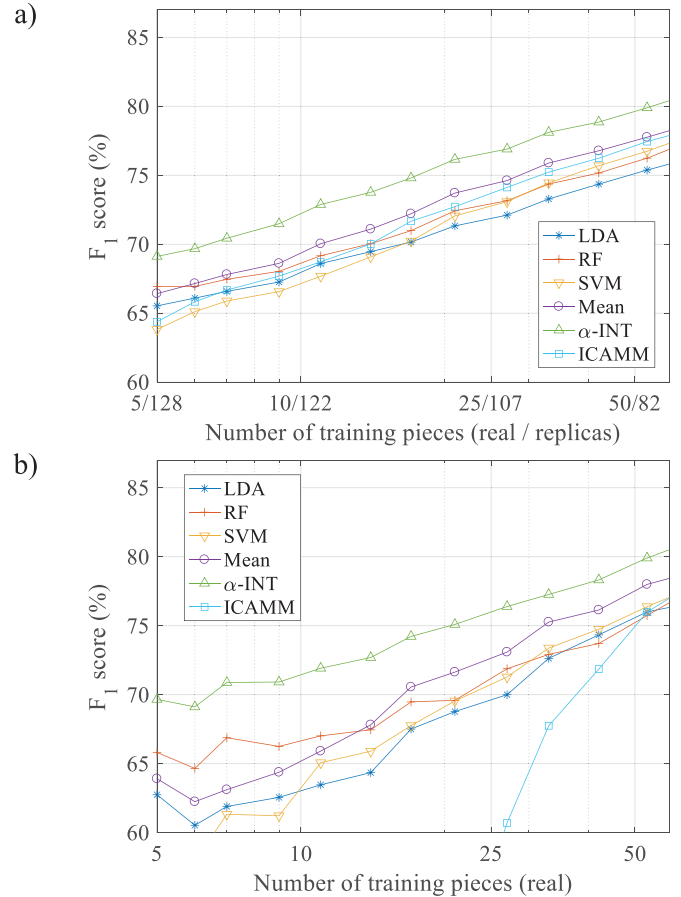


Fig. 8. Average classification for TSA and TSSG pieces with a decreasing number of labeled training pieces: (a) F1 score after adding replicates; (b) F1 score before adding replicates. X axes are log scale.

ranking features (centroid and maximum frequencies) before classification. Table 6 shows the results of the experiment. As in previous cases, α -INT obtained the best result, followed by RF and SVM. These differences were statistically significant. Mean was not able to improve on single classifier results, whereas α -INT did manage to combine single classifiers optimally. The performance decreased with respect to that obtained in previous experiments, showing the increased difficulty of the problem. Even after the decrease, performance was still relatively high, with a best value of 83.78% BAC.

As in Sections 6.2 and 6.3, we analyzed the performance of the proposed method with respect to the number of available training pieces. The number of available pieces for training was pro-

Table 6

Average classification performance for the experiment on TSA and TSSG pieces (values are percentages). The best final classification results are highlighted.

| Classifier | BAC | std. | F ₁ | std. | Acc | std. |
|---------------|--------------|-------------|----------------|-------------|--------------|-------------|
| LDA | 74.31 | 0.66 | 77.71 | 0.60 | 75.49 | 0.70 |
| RF | 80.71 | 0.63 | 81.31 | 0.64 | 80.67 | 0.66 |
| SVM | 80.11 | 0.67 | 81.00 | 0.63 | 80.27 | 0.63 |
| Mean | 78.97 | 0.72 | 80.76 | 0.62 | 80.06 | 0.75 |
| α -INT | 83.78 | 0.61 | 84.07 | 0.61 | 83.70 | 0.65 |
| ICAMM | 78.66 | 0.67 | 80.43 | 0.61 | 78.82 | 0.69 |

gressively reduced from 132 to 5 (3.16% of the total number of pieces in the training set). Additional synthetic sample replicates compensate for the missing training pieces, see results in Fig. 8. Standard deviation was lower than 1.3 and 2.1 for Fig. 8a and b, respectively. Only the results for F1 score up to 50 real training pieces are shown. All methods decreased their performance rapidly as the number of real training pieces decreased, and this decrease was generally more marked than it was for the case of Paterna vs. Les Jovaes (see Fig. 7). The effect of synthetic replicates was an improvement of the performance yielded by all methods for small amounts of real training pieces. In concordance with the previous experiments, α -INT performed significantly better than the rest of the methods in all cases, and the method most improved by adding synthetic replicates was RF (compare Fig. 7a and b). In this case, the minimum performance level of 80% F1 score was only reached by RF, SVM when using all the training pieces, and by α -INT when using an average of 65 real training pieces. Thus, it would seem like the split of *terra sigillata* pieces between TSA and TSSG is a tough problem that requires larger amount of labeled pieces. Finally, we also considered the effect of SSAL on classification performance. The results are summarized in Table 7. There was also a decrease in the sample number required to reach the maximum performance, with all methods requiring only 83–98 labeled training pieces rather than 132 pieces.

6.5. Verification of the results on chronological classification

For further verification of the proposed method performance, we compare its results with those obtained by the application in [15]: classification of archaeological ceramic pieces by their chronological period. Four periods were considered: Bronze Age, Iberian, Roman, and Middle Ages. The distribution of the pieces was 47 Bronze Age, 155 Iberian, 138 Roman, and 140 Middle Ages. All pieces were obtained from archaeological sites in the Valencian Community (Spain). The results of the considered methods are shown in Table 8. For this task, α -INT yielded the best result with 83.36% accuracy (82.86% BAC) and ICAMM yielded the second-best performance. Therefore, although the proposed method has been designed for provenance classification of archaeological ceramic shards, these results showed that it could also work for chronological classification.

Table 8

Average classification performance for the experiment on chronological classification of ceramic pieces (values are percentages). The best final classification results are highlighted.

| Classifier | BAC | std. | Acc | std. |
|---------------|--------------|-------------|--------------|-------------|
| LDA | 77.49 | 1.07 | 77.27 | 1.06 |
| RDF | 79.48 | 1.38 | 80.04 | 1.24 |
| SVM | 78.36 | 1.20 | 78.76 | 1.18 |
| Mean | 77.46 | 1.42 | 78.53 | 1.33 |
| α -INT | 82.86 | 1.09 | 83.36 | 1.06 |
| ICAMM | 82.50 | 1.34 | 83.00 | 1.32 |

7. Conclusions

A novel application of pattern recognition techniques for automated provenance classification of archaeological ceramics using ultrasounds has been presented. The proposed method is based on optimal fusion of scores from single classifiers and incorporates feature ranking, data augmentation, and semi-supervised active learning (SSAL). Results of different provenance classification problems have demonstrated the superiority of the proposed method over competitive ones such as LDA, RF, SVM, fusion of scores using the mean, and a method for classification of archaeological pieces in chronological period (ICAMM). Three problems of archaeological provenance classification of pieces of the same chronological period were approached. The first one was classifying between *terra sigillata* and non-*terra sigillata* ceramic shards from the same archaeological sites. The second one was classifying between Iberian ceramic shards from two sites in Spain (Paterna and Les Jovaes). The third one was classifying between Roman *sigillata* ceramic shards from two origins (Aretina and Sud-Gaul).

In addition, real situations that may occur in practice were simulated. The first real situation consisted of an experiment where the number of available pieces was decreased removing some pieces that were replaced with synthetic replicates. The addition of synthetic data increased the overall performance of the methods for small numbers of labeled training pieces and did not decrease performance for larger numbers of real labeled training piece. It also yielded more stable results. This experiment demonstrated the capabilities of the proposed method to obtain good results even for the probable case of data scarcity. This grounds the proposal of using relatively simple classifiers with non-extensive data requirements. SSAL allowed another real situation to be simulated, the uncertainty or difficulty in labeling the data by the archaeologist. It was demonstrated that it is possible to decrease the number of labeled archaeological pieces by using labeled and unlabeled pieces together to obtain the best results of classification. Future lines of research from this work include the extension of the proposed method to other related applications of material characterization and classification of archaeological ceramics such as conservation and restoration.

Table 7

Performance improvement using SSAL for the split between TSA and TSSG pieces. The values indicate the base result obtained by the classifier using supervised learning (+/-) the improvement obtained over base result using SSAL. Two ratios of Real / Replicate pieces are shown: 50/132 and 100/132. The best final classification results are highlighted.

| Classifier | BAC (%) | | F1 (%) | | Acc (%) | |
|---------------|-------------------|-------------------|-------------------|-------------------|-----------------|-----------------|
| | 50/82 | 100/32 | 50/82 | 100/32 | 50/82 | 100/32 |
| LDA | 73.07-0.60 | 74.42+0.02 | 75.75-1.33 | 77.81-0.08 | 74.4-0.9 | 76.1-0.0 |
| RF | 78.74+0.92 | 80.56+0.50 | 78.65+0.03 | 80.91+0.23 | 78.7+0.2 | 80.7+0.3 |
| SVM | 78.43+0.79 | 80.05+0.74 | 78.85+0.24 | 80.74+0.56 | 78.6+0.4 | 80.4+0.6 |
| Mean | 78.47+0.95 | 79.61+0.65 | 79.60+0.05 | 81.24+0.43 | 79.0+0.2 | 80.4+0.5 |
| α -INT | 82.39+0.93 | 83.80+0.62 | 82.06+0.14 | 83.87+0.42 | 82.2+0.4 | 83.8+0.5 |
| ICAMM | 74.54-1.13 | 78.93+0.70 | 72.03-3.95 | 80.29+0.52 | 73.3-2.1 | 79.6+0.6 |

Conflict of Interest

None.

Acknowledgments

This work was supported by Spanish Administration and European Union under grant TEC2017-84743-P.

References

- [1] T. Rose, et al., Questioning Fe isotopes as a provenance tool: Insights from bog iron ores and alternative applications in archeometry, *J. Archaeolog. Sci.* 101 (2019) 52–62.
- [2] D. Angelici, et al., μ -XRF analysis of trace elements in Lapis Lazuli-forming minerals for a provenance study, *Microsc. Microanal.* 21 (2) (2015) 526–533.
- [1] U. Stańczyk, L.C. Jain, *Feature Selection for Data and Pattern Recognition*, Springer, BerlinGermany, 2011.
- [2] E.G. Learned-Miller, J.W. Fisher III, ICA Using spacings estimates of entropy, *J. Mach. Learn. Res.* 4 (2003) 1271–1295.
- [3] O. Chapelle, *Semi-Supervised Learning*, MIT Press, 2003.
- [4] A. Salazar, G. Safont, L. Vergara, Semi-supervised learning for imbalanced classification of credit card transaction, *Proc. Int. Joint Conf. Neural Netw. IJCNN* (2018) 4976–4982.
- [5] S. Amari, *Information Geometry and its Applications*, Springer, 2016.
- [6] A. Soriano, L. Vergara, B. Ahmed, A. Salazar, Fusion of scores in a detection context based on alpha integration, *Neural Comput.* 27 (9) (2015) 1983–2010.
- [7] G. Safont, A. Salazar, L. Vergara, Multiclass alpha integration of scores from multiple classifiers, *Neural Comput.* 31 (4) (2019) 806–825.
- [8] B. Settles, in: *Algorithms for Active Learning*, in *Cost-Sensitive Machine Learning*, CRC Press, 2011, pp. 18–45.
- [9] A.R. Jiménez, F. Seco, Precise location of archaeological findings with a new ultrasonic 3D positioning sensor, *Sens. Actuators* 123 (2005) 224–233.
- [10] P. Capizzi, L.P. Electromagnetic and ultrasonic investigations on a Roman marble slab, *J. Geophys. Eng.* 8 (3) (2011) 117–125.
- [11] G.M.E. Kamh, Environmental impact on construction limestone at humid regions with an emphasis on salt weathering, *Alhambra islamic archaeological site, Granada City, Environ. Geol.* 52 (2007) 1539–1547.
- [12] R.A. Tsofigou, Mapping Mesolithic and Neolithic cultures behaviours and interactions with nature and properties of rocks in Western France, *J. Archaeol. Sci.* 36 (2009) 1615–1625.
- [13] A. Salazar, L. Vergara, ICA mixtures applied to ultrasonic nondestructive classification of archaeological ceramics, *Eurasip J. Adv. Signal Process.* 2010:125201 doi:10.1155/2010/125201.
- [14] J.D. Cheeke, *Fundam. and App. of Ultrasonic Waves*, CRC Press, 2002.
- [15] Y. Shmaliy, in: *Linear Time-Varying Systems*, in *Continuous-Time Systems*, Springer-Verlag, 2007, pp. 349–423.