STAR: Science & Technology of Archaeological Research

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ysta20

Automatic ceramic identification using machine learning. Lusitanian amphorae and Faience. Two Portuguese case studies

Joel Santos, Diogo A.P. Nunes, Ruslan Padnevych, José Carlos Quaresma, Martim Lopes, Joana Gil, João Pedro Bernardes & Tania Manuel Casimiro

To cite this article: Joel Santos, Diogo A.P. Nunes, Ruslan Padnevych, José Carlos Quaresma, Martim Lopes, Joana Gil, João Pedro Bernardes & Tania Manuel Casimiro (2024) Automatic ceramic identification using machine learning. Lusitanian amphorae and Faience. Two Portuguese case studies, STAR: Science & Technology of Archaeological Research, 10:1, e2343214, DOI: 10.1080/20548923.2024.2343214

To link to this article: https://doi.org/10.1080/20548923.2024.2343214

Published online: 15 May 2024.

STAR: SCIENCE & TECHNOLOGY OF ARCHAEOLOGICAL RESEARCH
2024, VOL. 10, NO. 1, e2343214 (17 pages)
https://doi.org/10.1080/20548923.2024.2343214

Routledge
Taylor & Francis Group

OPEN ACCESS

Check for updates

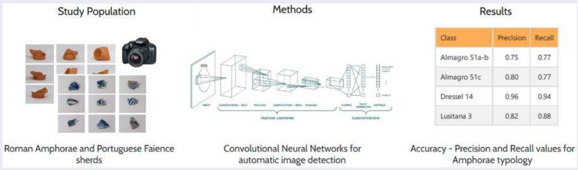
Automatic ceramic identification using machine learning. Lusitanian amphorae and Faience. Two Portuguese case studies

Joel Santos [a], Diogo A.P. Nunes [b], Ruslan Padnevych [c], José Carlos Quaresma [d], Martim Lopes [d], Joana Gil [e], João Pedro Bernardes [f,g] and Tania Manuel Casimiro [h]

[a]University of Leicester, Leicester, Portugal; [b]Instituto Superior Técnico, Lisbon, Portugal; [c]NOVA School of Science and Technology, Lisbon, Portugal; [d]CHAM-Centre for Humanities, FCSH-NOVA University of Lisbon, Lisbon, Portugal; [e]FCSH-NOVA University of Lisbon, Lisbon, Portugal; [f]University of Algarve, Faro, Portugal; [g]CEAACP, Faro, Portugal; [h]NOVA University of Lisbon FCSH HTC CFE, Lisbon, Portugal

ABSTRACT

This article presents a novel approach to classifying archaeological artefacts using machine learning, specifically deep learning, rather than relying on traditional, time-consuming human-based methods. By employing Convolutional Neural Networks (CNNs), this approach aims to expedite and enhance the identification process, making it more accessible to a wider audience. The study focuses on two types of artefacts- Roman Lusitanian amphorae (2nd-5th centuries) and Portuguese faience (16th-18th centuries)- chosen for their diversity. While Lusitanian amphorae lack decoration, Portuguese faience poses challenges with subtle colour variations. The study demonstrates the potential of this approach to overcome these hurdles. The paper outlines the methodology, dataset creation, and model training, emphasizing the importance of extensive data and computational resources. The ultimate objective of this research is to develop a mobile application that utilizes image classification techniques to accurately classify ceramic sherds and bring about a significant transformation in archaeological classification.



KEYWORDS
Deep learning; machine learning; artificial intelligence; Amphorae; Faience; ceramic identification

Introduction

Collecting, identifying, and classifying artefacts is probably the basis of the archaeological method since its beginnings (Harris and Cipolla 2017, 16). This identification and classification task, which varies in difficulty depending on the context and object, is fundamentally based on archaeological expertise, which is very time-consuming and limited to archaeologists' knowledge regarding specific types of artefacts. The project detailed in this article aims to tackle this problem precisely by changing how objects are classified using machine learning, specifically through automatic image classification. Our main goal is to make artefact (in our case, ceramic sherds) identification and classification faster, more accessible, and more accurate.

Machine learning is an application of Artificial Intelligence encompassing the learning of patterns within large datasets, useful for making predictions or decisions over new, related data. In recent decades, the increase in data availability and computing power has enabled the fast development of a set of machine learning models called deep learning (Sejnowski 2018). Using large amounts of data and computing power, these models have shown excellent results in many fields, such as medicine (Li et al. 2014), autonomous driving (Li, Chen, and Shen 2019), and in archaeological work. These uses include object and bone recognition, spatial identification of new tombs and cremations, old language translation and many more (Anichini et al. 2020; 2021; Berganzo-Besga et al. 2021; Brandsen and Lippok 2021; Navarro et al. 2021; Cobb 2023; Domínguez-Rodrigo et al. 2020; Gualandi, Gattiglia, and Anichini 2021; Gutherz et al. 2023; Lucena et al. 2016; Orengo et al. 2020; Pawlowicz and Downum 2021; Resler et al. 2021; Sharafi et al. 2016). Artificial intelligence use in archaeology has grown exponentially in recent years.

CONTACT Joel Santos joelrosantos@gmail.com Rua Manuel Pinheiro Chagas, 3-2Esq 2780-069 Oeiras Portugal

Our proposal is to apply these models specifically for ceramic identification, based solely on captured images of its sherds. Other works that use similar approaches may be dependent on additional information, such as the expected profile of the entire object based on the sample of the object's morphological curves (Navarro et al. 2021; Lucena et al. 2016; Parisotto et al. 2022), or as in the Archaide project (Anichini et al. 2020; 2021; Gualandi, Gattiglia, and Anichini 2021), which based its image identification of ceramics without design elements, like the Roman plainware Terra Sigillata, on a "shape-based recognition (… learning model …) integrating shape information from points along the inner and outer profile of a sherd (…) and a decoration classification, where the (…) decoration classifier is based on relatively standard architectures used in image recognition" regarding Majolica decorated pottery (Gualandi, Gattiglia, and Anichini 2021, 140). Other works, like Cardarelli (2022), used machine learning to extract information regarding ceramic profiles.

Using a deep learning approach based on Convolutional Neural Networks (CNN), our model establishes its classification solely on images previously classified by experts without additional information. This approach was already used by Pawlowicz and Downum (2021) regarding Tusayan White Ware from northeast Arizona. However, as these authors state, "[d]ue to the lack of design elements, this specific method of type classification using two-dimensional imagery is likely not applicable to plainware ceramics" (2021, 12). They were talking specifically about American southwest pottery, but we challenged ourselves to verify whether this statement would be true in other ceramic typologies. Our project wants precisely to make that step by demonstrating the potential of deep learning in archaeology with a model that accurately associates an image of an archaeological sherd with a specific ceramic typology, even in plainware ceramics or in ceramics without the colour variation that Pawlowicz and Downum identified as the limits in the American southwest ceramics.

This paper focuses on two kinds of ceramics: Roman Lusitanian amphorae, dated between the second and fifth centuries, and Portuguese faience, dated between the sixteenth and eighteenth centuries. These typologies were chosen owing to several factors. First, we wanted to apply our model to types of ceramics that would overcome the limitations suggested by Pawlowicz and Downum (2021). Lusitanian amphorae, our first experience, are plainware ceramics without decoration, and Portuguese faiences, our second experience, are decorated ceramics with slight colour variations, but without the contrast of the Tusayan White Ware that Pawlowicz and Downum considered fundamental in pattern identification (2021, 5). Secondly, we wanted to explore two completely different ceramics regarding chronology, and productive and aesthetic characteristics. Finally, and considering that we needed, according to our hypothesis, a considerable number of ceramic sherds of each typology, we chose two objects that are widely available and accessible in the Portuguese territory. During the Roman period, the Portuguese territory was an important commercial area of the Roman Empire, with its own amphorae production widespread in the Atlantic and the Mediterranean (see, for instance, Bombico 2016; Quaresma 2023). Following this, from the sixteenth to the eighteenth centuries, Portugal was the head of a transcontinental empire. Thousands of faience objects were produced and globally distributed to support that endeavour (Casimiro, Gomes, and Gomes 2015). In this sense, although separated by many centuries, these two ceramic productions are commonly found during archaeological interventions within and outside the Portuguese territory, and a tool that can aid in their identification is of utmost importance.

Our multidisciplinary team, composed of several archaeologists and engineers, established the following strategic objectives for the project:

- The development, preparation, and maintenance of an extensive and standardised database of correctly classified images of archaeological sherds, which can be used in the future for the most diverse studies and purposes;
- The application of machine learning algorithms to classify archaeological sherds based solely on photographs contained in the database, and;
- The development of a user-friendly mobile application to facilitate the work of field archaeologists.

Our work includes the collection and preparation of a database with correctly classified images corresponding to the following:

- More than 11,000 images of Lusitanian amphorae sherds from four different classes: Almagro 51a-b (Pinto and Magalhães 2016), Almagro 51c (Viegas, Raposo, and Pinto 2016), Dressel 14 (Raposo and Viegas 2016) and Lusitana 3 (Quaresma and Raposo 2016) (Figure 1a–d).
- More than 5,000 images of Portuguese faience sherds from six different classes: Aranhões, Coats of Arms, Beads, Lace, Semi-circles, and Floral (Casimiro 2013) (Figure 1e–j)

We could have used more classes (more amphorae and faience typologies), but limited workforce availability, together with limited access to other sherd typologies, forced us to make a compromise, reducing the number of classes included in this project. In the future, we expect to overcome this limitation.

Figure 1. (a) Almagro 51a-b; (b) Almagro 51c; (c) Dressel 14; (d) Lusitana 3 (all images are from Southampton City Council 2013). (e) Aranhões; (f) Coat of Arms; (g) Beads; (h) Lace; (i) Semi-circles; (j) Floral (all images are from Cabral Moncada Leilões). Photos are not to scale

In the first part of this paper we will contextualize the work, discussing the traditional process for classifying archaeological artefacts and the benefits of using machine learning. We will describe the materials used, including the database creation, preparation, and maintenance, and explain the methodology for developing, training, and testing the image classification model. Finally, we will present and discuss our results and a proposal for future work. The final goal will be a ready-on-hand mobile app with an algorithm to classify a ceramic sherd based solely on its photograph.

Traditional archaeological classification

As is well known, the identification of ceramic sherds, the traditional archaeological process for classifying ceramic artefacts, depends essentially on the experience and knowledge of experienced archaeologists. This limits, among other things, the collection of data, the speed of its collection, access to the collected data, and sometimes even the quality. Since this project is based in Portugal, we used Portuguese experience and considered the numerous excavations throughout the Portuguese territory as an example. Almost all produce considerable amounts of ceramic evidence that, for site report purposes, needs to be classified. However, there are several limitations that we believe could be overcome with the use of the work we present in this paper.

The first limitation concerns the availability of archaeologists with ceramic expertise in each excavation. This non-immediate availability produces an amount of material kept in archaeological warehouses without proper, or just generalist, interpretations (Figure 2). Unfortunately, the time elapsed between the discovery of evidence, its correct classification, and successive publication can often be several years.

Specialists cannot be everywhere, and the material is usually classified by field archaeologists who might not be specialists in the cultural contexts they are excavating. This increases the probability of misinterpretations, generalisations, and wrong classifications. However, even expert analyses can become outdated owing to the evolution in typological classifications that occurs over time and which is not always immediately learned. As noted by Heilen and Altschul, "analysis of artifacts in the field or through digital photographs in the laboratory at two sites yielded results that were neither accurate nor reliable" (2013, 132). Artificial Intelligence would be a revolution in these cases since automatic recognition would reclassify all the sherds in seconds, as long as the training set is correctly parameterized.

A second limitation is the enormous time required for classifying ceramic sherds and the time invested in inventories, which tend to become generic. Machine learning would facilitate the classification of these ceramic sherds in terms of speed and accuracy, improving the quantity and quality of available data. These limitations, even when overcome, typically delay the investigation process since the time invested is spent on material classification rather than on the interpretation of the results. Sherd classification should be a means, not an end.

In summary, the current limitations in traditional identification and classification would thus be reduced, accelerating the classification process, reducing variability in typological categorisation errors, and increasing the availability of information since all the photos would be available in a public repository. However, it is essential to mention that Artificial Intelligence does not eliminate the need for the human knowledge on which it is based. Automatic classification of images is only possible according to an essential human parameterization (supervised learning) which, once carried out, has the advantage of being available to everyone, whether teachers, students, or the archaeologists who face this situation daily.

Automatic image classification

Automatic image classification is the process of accurately assigning a class from a set of pre-established classes to a given input image. This process is carried out by a classification model, commonly called a classifier. Like any other machine learning model, the

Figure 2. Quantity of sherds waiting to be classified (Photo by Tânia Casimiro).

classifier is trained, validated, and tested on disjunct sets of the available data. Should the available data represent the real world (as it would not make sense to use it otherwise), the classifier's performance on the test set indicates its probable performance in the real world since the classifier has never seen those images while training. Given the nature of this task, we identified three key points that may determine its success:

1. The universe of classes and instances of those classes;
2. How well the training, validation, and test sets represent this universe;
3. The classifier's ability to discriminate between classes.

The first point refers to the feasibility of the task in light of the universe of study. Suppose we intend to apply this task to the automatic distinction between cats and apples. In this case, we can be very confident that the job is feasible and that it has the potential to be very accurate: the instances of each class are of such a different nature, with very different apparent characteristics, that we humans, as a reference, are easily capable of distinguishing them. On the other hand, if we try to differentiate between instances of very similar classes, such as images of the index finger and the ring finger, we cannot start with such high confidence because the task is intrinsically more complex: the classes of this universe are so similar that, even for a human with several years of experience, it would be challenging to execute it accurately.

The second point concerns collecting a set of images representative of the universe of study. Suppose a classifier only has access to a limited class representation during training. In this case, it will have difficulties generalising to other instances of the same class because it is unaware of the possible differences. Thus, whether identifying and extracting features that distinguish instances of different classes (a process called feature extraction in the context of machine learning) manually or automatically, the quality of this step is always limited by the representativeness of the training, validation, and test sets.

Finally, as suggested in the previous point, we should be concerned with the quality of the extracted features and the classifier's ability to use them to distinguish instances of different classes. The feature extraction process is traditionally done semi-manually (Guyon et al. 2008). Researchers have devoted a lot of time and study to the training set to identify these features and develop methods for extracting them for each image, effectively characterizing the images not by the set of pixels that define them but by a vector of features. More recently, the field of machine learning has evolved towards the use of deep learning models. These models, based on multi-layer architectures, are capable of learning increasingly abstract representations of their inputs and how to weigh them to make a final classification (LeCun, Bengio, and Hinton 2015). Instead of using a predetermined vector of features, the image itself (the set of pixels) is used. The clear advantage of this type of approach is the delegation of the feature extraction step to the model itself, rather than being based on the researchers' study. However, this approach requires a training set with abundant artefacts so these models can learn to identify the most relevant features to distinguish classes. Convolutional Neural Networks, an architectural specification for the layers of a deep learning model, have revolutionized the ability of a computational algorithm to distinguish images of different classes. At multiple levels of depth, these networks can learn increasingly abstract filters of the various concepts in the image, such that the last layer can distinguish images of different classes without direct reference to low-level features; for example, the colouring of a particular region of pixels. A visual representation of this architecture in action can be seen in Figure 3.

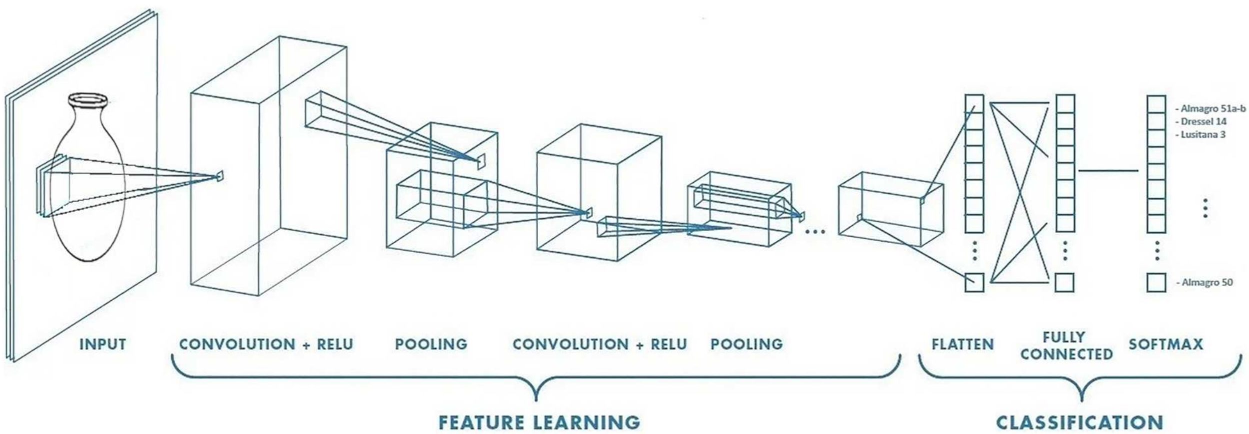


Figure 3. Visual representation of a Convolutional Neural Network. Adapted from https://towardsdatascience.com/a-vomprehensive-guide-to.-convolutional-neural-networks-the-eli5-way-3bd2b1164a53.

Dataset

To perform automatic image classification, the most essential data to be provided are images paired with their corresponding classes. These images are the basis for the classification process, and they need to be accurately labelled with their respective classes for the system to work effectively. Without this data, the system cannot accurately classify and categorise the images provided, making the process unreliable and ineffective. In our case, these images are of ceramic sherds, while the classes are the corresponding archaeological typology and type. In our experiments we worked with two typologies, amphorae and faience, which will be described below. Besides the information pairing images with classes, we collected further information for each ceramic sherd to avoid narrowing the dataset's future potential. Our data was collected from February 2020 to February 2023.

Object of study

The amphorae ceramic sherds are remains of the Romans' presence in the ancient Lusitania province (part of the Portuguese territory) since the second century BC. However, the first amphorae production in this territory is believed to have started around the second half of the first century BC with the type "Ovoid Lusitan" (Filipe 2023) (Western Lusitania). Nonetheless, only around 100 years later this ceramic typology was produced and became widespread throughout the territory with the massive production of the Dressel 14 type of amphora, used to transport salt-fish sauce. Dressel 14 was made in Western and Meridional Lusitania (Figure 4 – Map 1), two parts of the Lusitanian territories (in Portugal) with known amphorae production.

According to the project Amphora ex-Hispania, ten types of amphorae were produced in "Portuguese" Lusitania, some exclusively in certain zones, such as the Lusitana 3 type, which was produced only (according to present-day knowledge) in Western Lusitania, and the Almagro 51a-b type, produced mainly (again according to present-day knowledge) in Meridional Lusitania (Bernardes et al. 2013). These amphorae were produced during different periods – for example, the Dressel 14 type between the first and third century AD, and the Almagro 51a-b type between the fourth

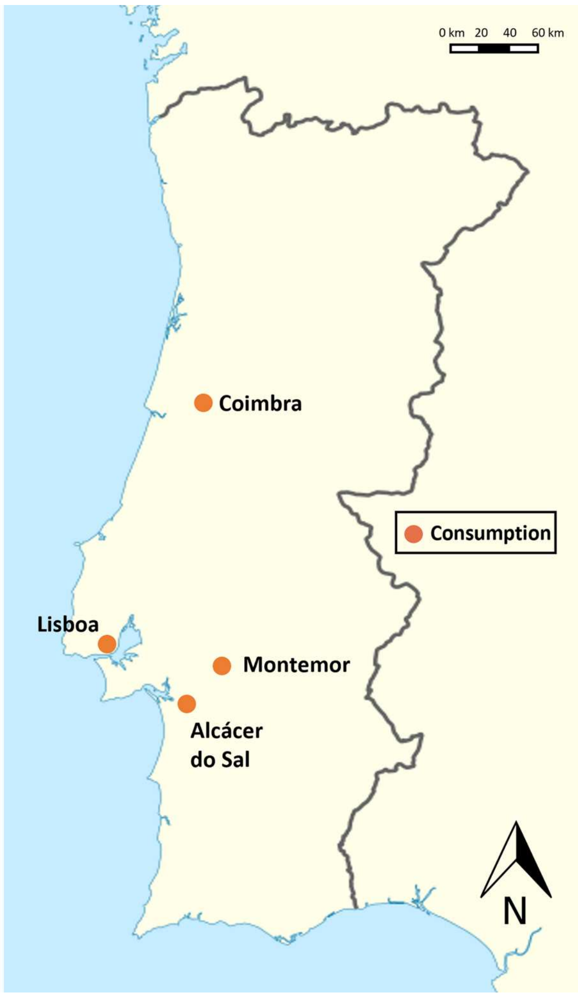


Figure 4. Maps with the provenance of the sherds used to take amphorae photos (Map 1) and provenance of the sherds used to take faience photos (Map 2).

and the fifth century AD – and can be used to roughly date the contexts where they are found. The final objective of this project is to have an App that can identify all ten types of amphorae and their respective variants, such as the 3 Lusitana 3 type variants (Quaresma et al. 2023). The results we present here focus on four types of amphorae (Dressel 14, Lusitana 3, Almagro 51c, and Almagro 51a-b), whose sherds were mainly identified in two production centres: Porto dos Cacos and Martinhal. Sherds from other production centres were also used (Rouxinol) (Figure 4 – Map 1), although in very small amounts.

The choice to take pictures of ceramic sherds mainly from production centres was taken owing to the vast quantity of sherds needed to do a correct classification using the shape of an object and the fact that the production centres usually have a lot of amphorae sherds that were discarded, which is precisely what we need to build our database. The Dressel 14, Lusitana 3, and Almagro 51c photos came mainly from Porto dos Cacos, while the Almagro 51a-b came mainly from Martinhal. Pictures taken from sherds collected in two consumption centres, Almoínhas and Miróbriga (Figure 4 – Map 1), were also used to diversify our sample and check how this would change the final results.

The four types of amphorae were chosen based partly on the available quantity collected in the production centres, but mainly due to their shape. The Dressel 14 type has no similarities with the other three types. The Lusitana 3 type is similar to the Almagro 51c's rims and handles, although very different in the bottom. The Almagro 51c is similar to the Almagro 51a-b on the rim and bottom (sometimes), while the handles differ. One of our objectives with these choices was to understand how the algorithm would behave with such similarities.

Regarding the Portuguese faience, although this was manufactured in Portugal from the fifteenth to the nineteenth century, the ceramics used for this project are dated roughly between the 1580s and 1770s. This chronological choice is related to the fact that we only wanted to use six classes of material for the first experiment on this ceramic typology classification, and enlarging the chronology would imply duplicating that number. These classes can be classified essentially, although not exclusively, as tableware, with elevated amounts of plates, bowls, and bottles and less frequent forms such as drug jars, pots, and boxes, among others (Casimiro 2013). Contrary to what was said about amphorae, the most distinguishable feature of these productions is not their shape but their decoration. During the almost 300 years of its production, different types and styles were created, and considering their holistic presence in Portuguese and the once-colonial territories, these have been critical elements in the chronological interpretation of contexts in the Atlantic world (Casimiro 2022).

These objects were made in three production centres, producing similar decorations: Lisbon, Coimbra, and Vila Nova, whose production supplied internal and external trade. Although many different decorations exist, the colour is usually the same, lacking the contrast of the Tusayan White Ware (Pawlowicz and Downum 2021, 5). The decoration, mainly in small sherds, can create some identification confusion and cause mistakes even for expert archaeologists. The six types of this ceramic typology (used in this project as the classes) integrate roughly 90% of all existing faience in the studied chronology and can be used to date different chronological periods, namely Aranhões, Coats of Arms, Beads, Lace, Semicircles, and Floral (Casimiro 2013). Photographs were taken from collections corresponding to Lisbon, Coimbra, Montemor-o-Novo, and Alcácer do Sal consumption sites (Figure 4 – Map 2).

Photographs

The two ceramic typologies, as mentioned, are identified through different characteristics. While amphorae are identified through shape, faiences are identified through decoration. These will impact the way pictures are taken. While an amphora sherd must be photographed from several angles (Figure 6), faience needs only one photo per sherd (Figure 7). Owing to this difference, but mostly since different people would be taking the photos, an internal protocol had to be created to reduce the variance among photographs. This protocol controlled the following variables: (1) background, (2) lighting, (3) camera positioning and orientation, (4) photo dimension, and (5) sherd position and angle. Our setup for photographing a single ceramic sherd is shown in Figure 5.

To ensure that the ceramic sherd is the main focus of the image, the background (1) should have a neutral

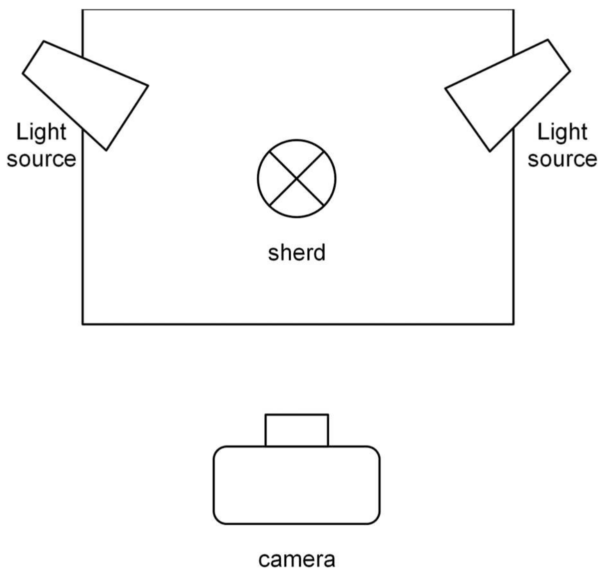


Figure 5. Diagram of our setup for photographing ceramic sherds from a top-view.

Figure 6. Images of the same amphora sherd from multiple angles.

role. This means that it should be white or any other neutral colour (Figures 6 and 7) and devoid of any other objects besides the sherd, especially the ready-to-use archaeologist's scale. The lighting (2) definition was essential since we intended to minimize shadows and highlight specific characteristics of the sherd by

Figure 7. Images of several sherds taken from the same angle.

having multiple light sources from various angles. The camera (3) was fixed in the exact location throughout the collection process, except for big pieces (e.g. complete amphorae), where we needed to create a greater distance between the camera and the sherd (Figure 5). (4) All images must have a 2:3 ratio to avoid converting differently, as we will see, the same characteristic on different sherds of the same type.

The positioning and the angle of the sherd (5) were of utmost importance. First, the sherd was always positioned in the center of the photograph. Second, for the amphorae, where the shape was the most critical aspect, a sherd must be photographed from several angles since, even for an archaeologist expert, specific perspectives are more informative than others, thus also facilitating automatic image classification. In the database, we identified which angles would allow easier recognition by an expert archaeologist and used only those photos for the model training. Figure 6, as an example, shows the same sherd (a Lusitana 3 amphora-type bottom) photographed from multiple angles. As for the faience, owing to the decoration-based identification, only one photo where the drawings on the sherd are perceptible was needed, eliminating the necessity of taking pictures from different angles (Figure 7). Although we only use images tagged with "easier angle" in this work, we keep all other images in the database so future work can develop a model robust to angles less informative of shape and other fragment characteristics.

Metadata

Each image in the database was accompanied by a metadata vector with information about the sherd itself and its context. These are described in Table 1. The "sherd" part of the metadata contains the archaeologist's previous classification of the sherd in the image. This classification was done by specialists in the field (amphora and faience) and successively checked by a second specialist. A final meeting was held to resolve situations in which the specialists did not agree. The "context" part of the metadata includes logistical information for future checking, plus the "image ID" parameter, which is the critical parameter that associates a photograph with its metadata vector. In this paper's experiments, the "Type" parameter was our target prediction class, although we might evolve into the "sub-type" or the "fragment" identification in the future.

Table 1. Metadata associated with each image.

Axis	Parameter	Description
Sherd	Typology	Ceramic typology regards a certain recognised group of ceramics (E.g. "Roman Amphora" or "Portuguese Faience")
	Type	Each typology can be divided into different types (E.g. "Dressel 14" or "Lusitana 3" for the "Roman Amphora" typology)
	Sub-type	Each type can eventually be subdivided (E.g. "V1", "V2" or "V3" from the "Lusitana 3" type)
	Fragment	The part of the ceramic that was photographed (E.g. "Rim" or "bottom")
	Fragment notes	Notes in case the fragment has some particularity (E.g. "Handle connecting with rim")
	Is easier angle	This is a binary parameter ("True"/"False"). True if the image is deemed an easier angle for recognition by an expert archeologist
Context	Location	Where the photograph was taken. (E.g. "Porto dos Cacos, Portugal")
	Warehouse	Name of the warehouse where the sherd was. (E.g. "CAA")
	Date	Date of the photograph (E.g. "2020-04-18")
	Image ID	Image identification number (E.g. "IMG_165")
	Container ID	Sherd container identification number for future validation (E.g. "441")

Table 2. Distribution of images per target class – "Type" (Roman Amphora example).

Total	Almagro 51a-b	Almagro 51c	Dressel 14	Lusitana 3
11,486	442	2,132	5,833	3,079

Data description

The Roman Amphora dataset comprises 11,486 images. All photos belong to one of 4 classes ("Type" parameter): "Almagro 51a-b", "Almagro 51c", "Dressel 14" and "Lusitana 3". Table 2 shows the corresponding image distribution, highlighting the unbalanced nature of our dataset in light of the target class ("Type") – for example, comparing the Almagro 51a-b and the Dressel 14 samples. Table 3 shows the distribution of images for the remaining "sherd" metadata parameters (see Table 1 for reference).

The "Portuguese Faience" dataset comprises 5,108 images. All photos belong to one of 6 classes ("Type" parameter): "Aranhões", "Coat of Arms", "Beads", "Lace", "Semi-circles" and "Floral". Table 4 shows the corresponding image distribution, highlighting the unbalanced nature of our dataset in light of the target class ("Type") – for example, comparing the Floral and the Beads samples. Table 5 shows the distribution of images for the existing "sherd" parameters (see Table 1 for reference). As you can see, the extra information collected for the amphorae, "Sub-type" and "Fragment", was not needed for the "Portuguese faience". Moreover, since each sherd has only one photo, the "Is easier angle" parameter is never set as "False".

Methodology

This chapter describes the methodology used step-by-step, from the architecture of the machine learning

Table 3. Distribution of images per "Sherd" metadata parameter (Roma Amphora example).

Parameter	Distribution
Typology	Amphora (11,486)
Sub-type	VR1 (943), VR2 (512), VR3 (185), N/A (9,846)
Fragment	Bottom (4,735), Rim (4,829), Handle (1,893), Body (29)
Is easier angle	True (8,050), False (3,436)

Table 4. Distribution of images per target class – "Type" (Portuguese faience example).

Total	Aranhões	Coat of arms	Beads	Lace	Semi-circles	Floral
5,108	986	325	277	632	611	2,277

Table 5. Distribution of images per "Sherd" metadata parameter (Roma Amphora example).

Parameter	Distribution
Typology	Faience (5,108)
Sub-type	N/A
Fragment	N/A
Is easier angle	True (5,108), False (0)

approach to the evaluation of the results, passing through the data pre-processing and the model training and validation.

Architecture

As previously stated, our work aims to contribute to integrating machine learning techniques for archaeological sherd classification based on deep learning Convolutional Neural Networks. Besides the input layer of size (264, 264, 3), our model is composed of 13 hidden layers, depicted in Figure 8. These follow the standard approach (Elhassouny and Smarandache 2019): a convolution layer, a pooling layer, and a normalization layer. All convolution layers have size filters (3, 3), with 32, 64, 128, and 64 filters, respectively. All of these layers also have a ReLU (Elhassouny and Smarandache 2019) activation unit. Finally, the dense layer at the end has a softmax (Elhassouny and Smarandache 2019) activation unit to output target class scores that can be interpreted as probabilities (i.e. for a given input image, the sum of the output target class scores is equal to 1). Our model was implemented with Keras (Chollet 2015) in Python. The model is compiled with the Adam optimizer with a learning rate of 0.0001, beta_1 of 0.9, beta_2 of 0.999, and epsilon of 1e-08. The loss function used is categorical cross-entropy. We used the same architecture and trained it on two different tasks (amphorae and faience). It is arguable whether is the best approach, but we were precisely trying to check whether using the same architecture in different typologies of sherds would present the same behaviour. Future work will include the use of different models. Importantly, owing to the available computing resources, we opted to use a smaller network and train it from zero, instead of loading a pre-trained model. Given resource availability, future work should include a pre-trained model such as ImageNet.

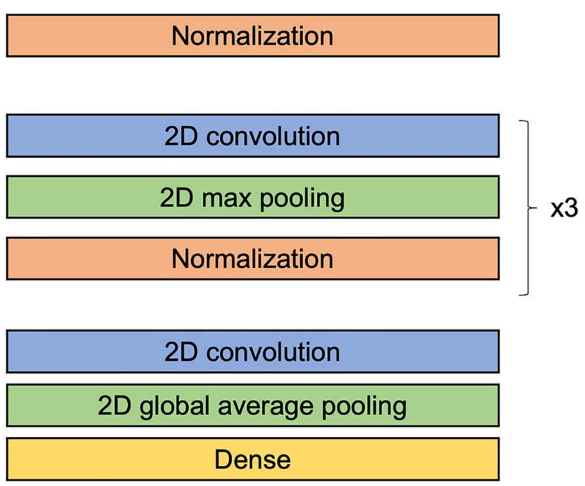


Figure 8. Model's architecture.

Data preprocessing

Since the input layer of our model is of size (264, 264, 3), all images needed to be standardised to that shape. Although the 2:3 ratio of the original resolution is lost during the conversion to 1:1, since we made sure that all images had the same ratio, the relative dimensions of the sherds were kept intact. Moreover, we limited the dataset to images with the "Is easier angle" parameter set to "True". Table 6 shows the distribution of these images per target class for the "Roman Amphora" typology example after the "Is easier angle" reduction. Since, as previously stated, there are no limitations in the "Is easier angle" parameter for the "Portuguese faience", Table 4 quantities will be used for this ceramic typology.

Model training and validation

We split the dataset into training, validation, and test sets (Tables 7 and 8). This split was randomised and stratified, with a fixed random seed for experiment comparability. The model training epochs were set to 50 since we trained the model with more epochs, but they did not increase its accuracy. In each epoch, for every image in the training set in a randomised order, the model tries to predict its target class, adjusting its parameters as needed, according to the loss function (which measures how far off the model's prediction is from the actual class, in this case, using categorical cross-entropy (Murphy 2022)). At the end of each epoch the model is validated on the validation set, an independent set of images the model has not seen during training. This validation set is used to adjust any hyperparameters and determine the quality of the training, especially regarding overfitting (i.e.

Table 6. Distribution of images per target class with the "Is easier angle" parameter set to "True" for the Roman Amphora typology.

Total	Almagro 51a-b	Almagro 51c	Dressel 14	Lusitana 3
8,050	442	1,604	4,147	1,857

Table 7. Training, validation, and test sets for the "Roman Amphora" typology example.

Total	Training	Validation	Test
8,050 (100%)	7245 (90%) 6,521 (90% of training)	724 (10% of training)	805 (10%)

Table 8. Training, validation, and test sets for the "Portuguese faience" typology example.

Total	Training	Validation	Test
5,108 (100%)	4597 (90%) 4,138 (90% of training)	459 (10% of training)	511 (10%)

when the model loses the capacity to generalise beyond the training set). The test set comprises 10% of all images. The remaining 90% are split between the training set (90% of the 90%) and the validation set (10% of the 90%). Importantly, after all training epochs, we keep only the model with best performance on the validation set.

Evaluation

The test set is exclusively used to test the model and report its performance.

The metrics used to evaluate the performance of our model are those most used in the multiclass classification literature (Albawi, Mohammed, and Al-Zawi 2017; Guyon et al. 2008; LeCun, Bengio, and Hinton 2015; Li et al. 2014), namely accuracy, precision, and recall. Accuracy measures the percentage of test images that were correctly classified. Precision measures the percentage of the class's true positives in all images classified as class-positive. Recall measures the percentage of the class's true positives in all images that belong to that class. These metrics vary between 0 and 1, where 1 is the best possible score. Due to class imbalance (Tables 2 and 4) we were explicitly interested in balanced accuracy and weighted precision and recall (Murphy 2022), a version of the standardised metrics, since they consider each class's relative size in a dataset. We also observed the model prediction's confusion matrix to better understand which classes were confused in the prediction and where to improve.

Results and discussion

Each epoch took, on average, 25 s to process. Figures 9 and 10 show the training and validation accuracies for all 50 Roman amphorae and Portuguese faience epochs. Tables 9 and 10 show the accuracy and balanced accuracy of the test set of each ceramic typology. Tables 11 and 12 show the precision and recall for each class of the ceramic typology. Tables 13 and 14 show the corresponding confusion matrix.

This project's first experiment was with Roman amphorae, and the final results precisely show the maturity achieved. Not only is the general accuracy much higher (Table 9), both in standard accuracy (88%) and in balance accuracy (84%), when compared with the same metrics for the Portuguese faience (Table 10), 57% and 49%, but that maturity can also be observed in the training graphs (Figures 9 and 10), where the Portuguese faience still presents considerable instability. Let us start by analysing the results of the Roman amphorae.

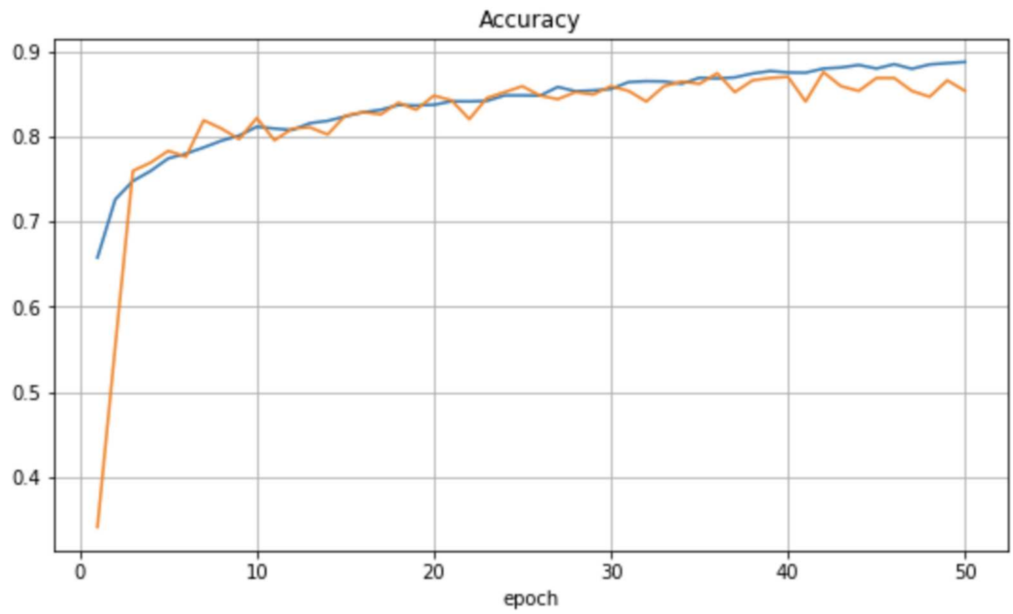


Figure 9. Training and validation accuracies for all 50 epochs (Training (blue) and Validation (orange) accuracies in each epoch) for the "Roman Amphora" typology example.

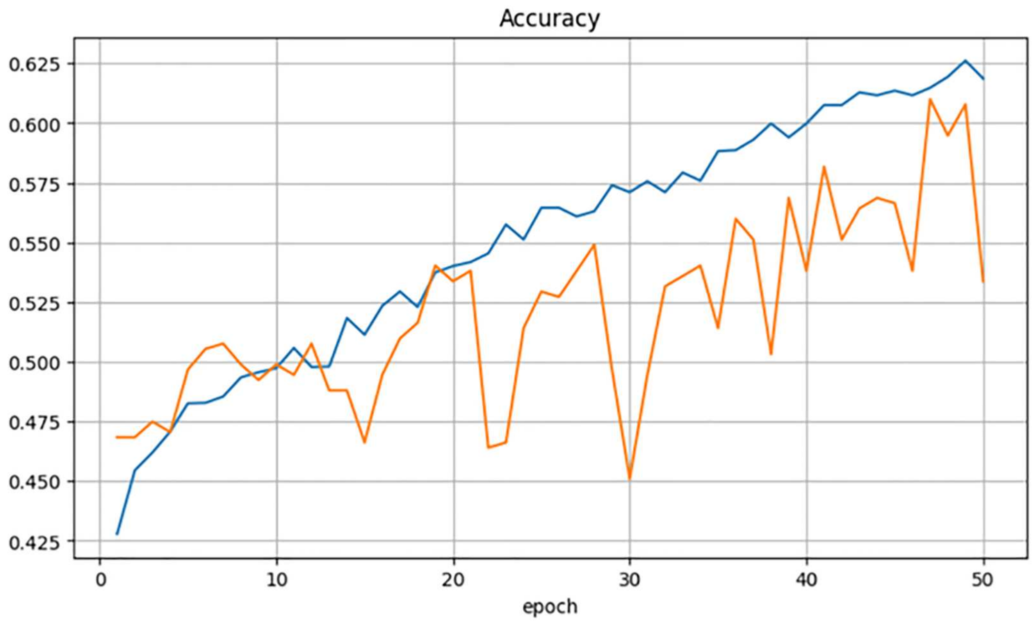


Figure 10. Training and validation accuracies for all 50 epochs (Training (blue) and validation (orange) accuracies in each epoch) for the "Portuguese faience" typology example.

Table 9. Accuracy and balanced accuracy on the test set for the "Roman Amphora" typology example.

Accuracy	Balanced accuracy
0.88	0.84

Table 10. Accuracy and balanced accuracy on the test set for the "Portuguese faience" typology example.

Accuracy	Balanced accuracy
0.57	0.49

Regarding the model training and validation in the 50 epochs, we see no overfitting to the training set (Figure 9). Overfitting occurs when the model is so tuned to the training set that it loses generalizability capacity. This is commonly observed when, at a given epoch, the model's performance is very high for the training set and suddenly drops for the validation set. This did not happen during the training of our model. Indeed, the model seems to be relatively stabilised by the 50th epoch.

Although we reported both accuracy and balanced accuracy in the test set, given our class imbalance (e.g. Dressel 14 has almost ten times more images than Almagro 51a-b), we think it was wise to restrict our analyses to the balanced accuracy, which accounts for these differences. Given the results during training and validation (Figure 9), the test performance results were expected (Table 9). The 88% and 84% results in each of the accuracies are excellent. They showed that our model could learn the differences between amphorae classes and take advantage of this to recognise images it has never seen before.

Regarding the precision and recall metrics (Table 11), Almagro 51a-b and Almagro 51c have the lowest scores. This probably correlates with Almagro 51a-b being the least represented class in the training data (Table 6) and with Almagro 51c being the type most similar to the other types (see below). We noticed that Dressel 14 has the highest score, close to 1, which is the best score possible. This also correlates with this class being the most represented in the training data. Given these observations, and since no significant discrepancies exist between precision and recall for any of the classes, we argue that these scores could be improved first by solving class representation by using more balanced training data rather than changing the model's architecture.

Observing the confusion matrix for the test set (Table 13), we can immediately see the most confusion between Almagro 51a-b and Almagro 51c and slight confusion between Almagro 51c and Lusitana

Table 11. Precision and recall in the test set for each class for the "Roman Amphora" typology example.

Class	Precision	Recall
Almagro 51a-b	0.75	0.77
Almagro 51c	0.80	0.77
Dressel 14	0.96	0.94
Lusitana 3	0.82	0.88

Table 12. Precision and recall in the test set for each class for the "Portuguese faience" typology example.

Class	Precision	Recall
Aranhões	0.43	0.60
Coat of Arms	0.51	0.50
Beads	0.48	0.43
Lace	0.62	0.54
Semi-circles	0.67	0.16
Floral	0.63	0.70

Table 13. Confusion matrix in the test set for the "Roman Amphora" typology example.

		Predicted class			
		Almagro 51a-b	Almagro 51c	Dressel 14	Lusitana 3
Actual class	Almagro 51a-b	0.77	0.23	0	0
	Almagro 51c	0.07	0.77	0.04	0.13
	Dressel 14	0	0.02	0.94	0.04
	Lusitana 3	0	0.08	0.04	0.88

3. This is expected because (1) Almagro 51a-b is the least represented class in the training data, and (2) because, as mentioned previously, there are some similarities between Almagro 51a-b and Almagro 51c (on the rim and the bottom) and between the Almagro 51c and Lusitana 3 (on the rim).

Notably, a possible mismatch might exist between our test set and the images captured in the field using a smartphone device. If this mismatch occurs, the performance results are expected to be a ceiling performance in the field since the model is trained and tested in what we might call a laboratory (i.e. ideal) setting. These results still needed to be confirmed. Nonetheless, the results stand independently and prove the model's validity for the Roman amphorae.

Given the excellent performance results and the reasons for the model's validity and capabilities, we decided to test the application of the same pipeline to a similar problem: automatically identifying sherds of faience using the same model.

We can observe some overfitting to the training set in the model training and validation in the 50 epochs (Figure 10). This can be observed in epochs 30 or 50, for example, when the model's performance is very high for the training set and suddenly drops for the validation set. We used a fixed learning rate. In the future variable learning rates can be explored, decreasing as we reach the last epochs to improve our results. Given the results during training and validation (Figure 10), the test performance results of 57% and 49% were expected (Table 10), showing that the model did not perform so well in this setting, although still much better than the random baseline (which would have an accuracy of 17%) and better than previous experiments (Anichini et al. 2020; 2021). These results are probably due to the size of our sample. While Roman amphorae used more than 8,000 images split into four classes, Portuguese faience used only 5,000 images divided into six classes, with one of those classes, the Floral, comprising more than one-third of the images (2,277). The Floral type (see Table 12) was precisely the class with the best results. We argue that before changing the model's architecture, we should first try to improve these scores by solving class representation.

Notably, the class semi-circles have the lowest recall (Table 12), even though it is not one of the least represented in the training set. The reason may be related to the fact that it is very similar to the Floral class, especially in small sherds where the paint strokes made by the potter are challenging to recognise. This is confirmed by the confusion matrix (Table 14), where 70% of the Semi-circles sherds were confused as being Floral.

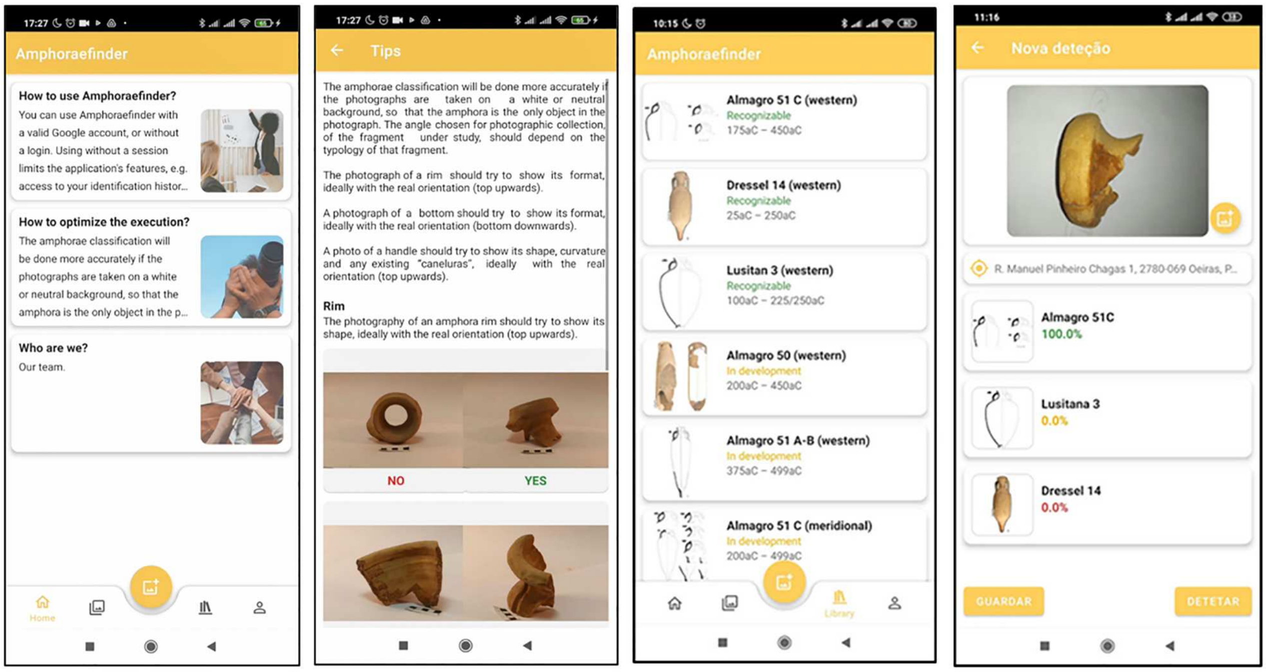
These two experiments showed us that, even though our approach accurately identifies amphorae sherds with extraordinary results, it does not directly translate the same results to faience sherds. We believe this is related to the nature of the faience samples. First, as mentioned, this is due to the size of the sample and its unbalanced distribution in the six classes. Second, the most relevant characteristic distinguishing the various Portuguese faience classes is the drawn patterns, and all other differences are irrelevant overall. Thus for this specific problem, after increasing the size of the sample, we suggest an approach based on pattern matching.

Another step forward – mobile application for field use

In case of success, one of our main objectives was to use the developed tools to help archaeologists in the field and create something that could make the work described in this document practical and accessible for them. We believe the Roman Amphorae's excellent results are mature enough to take a step forward. We wanted something a non-expert archaeologist could use in the field or the lab to identify a sherd automatically. To this end, we developed Amphoraefinder, an Android app. Amphoraefinder is still a beta experiment with three sections: landing page, library, and identification.

Table 14. Confusion matrix in the test set for the "Portuguese faience" typology example.

		Predicted class					
		Aranhões	Coat of arms	Beads	Lace	Semi-circles	Floral
Actual class	Aranhões	0.60	0.01	0.05	0.13	0	0.21
	Coat of arms	0.22	0.50	0.06	0.08	0	0.14
	Beads	0.25	0	0.43	0.04	0	0.29
	Lace	0.19	0.01	0.06	0.54	0.01	0.18
	Semi-circles	0.05	0.03	0	0.05	0.16	0.70
	Floral	0.19	0.06	0.01	0.02	0.02	0.70



(a) (b) (c) (d)

Figure 11. (a) Landing page; (b) Advice on taking photographs of sherds; (c) Library section; (d) Identification of a sherd and saved image in the device's gallery.

The landing page provides an overview of how the app works. It also aggregates various pieces of advice on how to photograph sherds in the field to achieve better identification results (e.g. lighting, sherd positioning, angle, background, etc.) (Figure 11a and b). The library section lists all amphorae that the application can identify, those currently under development, or those under consideration for future work. It is also connected with the Amphorae ex-Hispania site (https://amphorae.icac.cat/), which allows for quick consultation of details regarding a particular amphora (Figure 11c). Finally we have the app's main functionality, the identification section. Here it is possible to photograph a sherd, or choose one from the device's gallery (previously photographed), and identify the sherd type using the machine learning model presented and discussed in this document. It is also possible to save the photograph with the classification result in the device's gallery (Figure 11d).

As for the faience identification, we still do not have the necessary recognition percentages to create such an identification tool.

Conclusion

The work described in this document has many stages, from the collection of photographs to the final app, all to simplify an essential task in archaeology, specifically the identification of archaeological artefacts, by suggesting the inclusion of machine learning models. These, among other features, can be designed to classify images in any given universe of classes automatically. In line with this, we presented our experimental setup, including the collection, preparation, and extension of the database, which is identified as one of the most critical points of the work. We successfully specified the model's architecture, training, validation, and testing methodologies.

We wanted to take a step forward regarding the work by Pawlowicz and Downum (2021), which identified some limitations in the use of this approach for plainware ceramics or ceramics lacking the colour variation and contrast of the Tusayan White Wares from northeast Arizona. Given our results and their discussion, we can state that our model presents an exciting and promising ability to classify plainware ceramics, namely Roman amphorae. In contrast, ceramics without the colour variation and contrast of the Portuguese faience still need to be deepened. However, we believe an increase in the sample size and balance of the Portuguese faience classes would be enough to improve the results.

One of the most significant obstacles to leveraging Artificial Intelligence for archaeological research is the scarcity of data. Limited data availability makes it challenging for researchers to train Artificial Intelligence models and develop algorithms that can accurately analyse and interpret archaeological artefacts and data. Adding to this scarcity, existing data (photographs of artefacts taken by archaeologists used for field and find reports) do not always have the necessary requirements (quality, size, etc.), suitable for use in the deep learning models. This is why an essential part of this project was to take a considerable number of photographs to guarantee that the traditional lack of data would not be an issue.

Our model can be used in multiple contexts, from a mobile app for the fast classification of ceramic sherds in the field to the classification of large amounts of ceramic images for almost instantaneous database indexing. However, to be even more helpful, we should expand the number of classes in the Roman amphorae typology and improve the accuracy of the results regarding Portuguese faience. Our next steps will be precisely in this direction. We will include the Almagro 50 amphora type by taking photos in the production centre of Quinta do Rouxinol, and we will start testing the capacity to identify sub-types of the Lusitana 3 amphora type. Regarding the Portuguese faience, we are thinking of expanding our approach to capture visual explanations of the sherd. This, most probably, could enhance comprehension of the classifier's reasoning.

In this work we present the first iteration of this application on Android, called Amphoraefinder. Archaeologists are eager to use instruments that can improve their work, namely the identification of ceramics, giving them more time for other activities. Ideally this time would be used to conceptualise new ideas for making archaeology; however, even if the algorithm is only used to improve the number of correctly identified sherds, we would be satisfied and consider our work a success.

Acknowledgments

We would like to thank everyone who facilitated our access to their ceramic collections: Centro de Arqueologia de Almada, Jorge Raposo (CAA), Ana Cristina Oliveira (Almoínhas), Marisol Ferreira (Alcácer do Sal), Manuela Pereira (Montemor), António Marques (CAL Lisboa), Diogo Teixeira Dias (Coimbra), Raquel Santos and Nuno Dias (Neoépica) and Filipe Caldeira (CPAS). We would also like to thank those who helped us in the process of taking photographs: John, Marcelo, Joana, Lucas, Andreia, Rute, Filipe, and Raquel.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Fundação para a Ciência e a Tecnologia.

Notes on contributors

Joel Santos graduated in archaeology from NOVA University of Lisbon. He then obtained an MA in archaeology and heritage at the University of Leicester, where he developed a dissertation about the Archaeology of Loneliness and is currently pursuing a PhD at the University of Leicester. He is passionate about solving problems and mainly focuses on critical analysis and theoretical debates in archaeology, applying previous knowledge obtained in a prior electronics and computer engineering graduation from the Instituto Superior Técnico in Lisbon.

Diogo A. P. Nunes holds a Master's and Bachelor's in informatics and computer engineering from the University of Lisbon, Instituto Superior Técnico. He has been pursuing a Ph.D. in informatics and computer engineering at the University of Lisbon, IST since February 2021. He has been a early stage researcher at the Institute of Systems and Computer Engineering Research and Development in Lisbon (INESC-ID), at the Human Language Technologies lab since 2019.

Ruslan Padnevych completed his Master's in informatics and computer engineering in February 2021 from the University of Lisbon, NOVA School of Science and Technology (FCT NOVA), and a Bachelor's degree in informatics and computer engineering in 2018 from the same institution. His dissertation was focused on machine learning with the theme of Robust Facial Biometrics for Virtual ldentification using Machine Learning. After that, he followed his passion and started his professional career as a Mobile developer.

José Carlos Quaresma is an assistant professor at Universidade Nova de Lisboa, a vice president of SECAH (Sociedad de Estudios de la Cerámica Antigua en Hispania), and an evaluator for the European Science Foundation. He has published more than a hundred works, mainly in the fields of ceramology, ancient economy, and trade relations in the imperial and late antique periods. He has carried out research in Portugal, Spain, France, and Italy.

Martim Lopes is pursuing a Ph.D. in classical archaeology at Universitat Rovira i Virgili. He is researching ceramic and glass taxonomies from waste dump contexts of Lusitania between the 3rd and 6th centuries AD. He is also involved in research projects-TabMir II and ProCir-for the study of goods production and circulation in Ciudad Real. His research focuses on Roman and Late Antique archaeology, especially commercial dynamics, with a strong emphasis on ceramics.

Joana Gil is an undergrad in archaeology at the NOVA University of Lisbon.

João Pedro Bernardes has a Ph.D. in archaeology from the University of Coimbra and is a full professor at the University of the Algarve. He has participated in and led national and international research projects on Roman archaeology and the enhancement of cultural heritage. He has published around 150 titles, including books and collaborations in collective works and scientific journals.

Tânia Manuel Casimiro graduated in history and archaeology at NOVA University of Lisbon took an MA in artefact studies at the University College of London, and returned to NOVA for her PhD in 2011. Her research focuses on the early modern period and contemporary global contacts of people and objects. Theoretically, she deals with frameworks concerning how individuals and things relate and interact in the formation of identities and how they can reflect globalized ontologies.

ORCID

Joel Santos http://orcid.org/0000-0002-5796-9213
Diogo A.P. Nunes http://orcid.org/0000-0002-6614-8556
José Carlos Quaresma http://orcid.org/0000-0003-3139-1975

Martim Lopes http://orcid.org/0000-0001-9261-7240
João Pedro Bernardes http://orcid.org/0000-0002-1086-2128
Tania Manuel Casimiro http://orcid.org/0000-0002-9471-6194

References

Albawi, S., T. A. Mohammed, and S. Al-Zawi. 2017. "Understanding of a Convolutional Neural Network." 2017 International Conference on Engineering and Technology (ICET), 1–6.

Anichini, F., F. Banterle, J. Buxeda i Garrigos, M. Callieri, N. Dershowitz, N. Dubbini, D. L. Diaz, et al. 2020. "Developing the ArchAIDE Application: A Digital Workflow for Identifying, Organizing and Sharing Archaeological Pottery Using Automated Image Recognition." *Internet Archaeology* 52. https://doi.org/10.11141/ia.52.7.

Anichini, F., N. Dershowitz, N. Dubbini, G. Gattiglia, B. Itkin, and L. Wolf. 2021. "The Automatic Recognition of Ceramics from Only one Photo." *Journal of Archaeological Science: Reports* 36. https://doi.org/10.1016/j.jasrep.2020.102788.

Berganzo-Besga, I., H. A. Orengo, F. Lumbreras, M. Carrero-Pazos, J. Fonte, and B. Vilas-Estévez. 2021. "Hybrid MSRM-Based Deep Learning and MultitemporalSentinel 2-Based Machine Learning Algorithm Detects Near 10k Archaeological Tumuli in North-Western Iberia." *Remote Sensing* 13: 4181. https://doi.org/10.3390/rs13204181.

Bernardes, J. P., R. Morais, I. Vaz Pinto, and R. Dias. 2013. "A olaria baixo-imperial do Martinhal, Sagres (Portugal)." In *Hornos, talleres y focos de producción alfarera en Hispania*, edited by D. Bernal, L. C. Juan, M. Bustamante, J. J. Díaz, and A. M. Sáez, 317–329 (Monografías Ex Officina Hispana; 1),. Cadiz: Universidad de Cadiz.

Bombico, S. 2016. *Economia marítima da Lusitânia romana: exportação e circulação de bens alimentares. Tese de doutoramento policopiada*. Évora: Un. de Évora.

Brandsen, A., and F. Lippok. 2021. "A Burning Question – Using an Intelligent Grey Literature Search Engine to Change our Views on Early Medieval Burial Practices in the Netherlands." *Journal of Archaeological Science* [Online] 133: 105456. https://doi.org/10.1016/j.jas.2021.105456.

Cardarelli, L. 2022. "A Deep Variational Convolutional Autoencoder for Unsupervised Features Extraction of Ceramic Profiles. A Case Study from Central Italy." *Journal of Archaeological Science* 144: 105640. https://doi.org/10.1016/j.jas.2022.105640.

Casimiro, T. M. 2013. "Faiança Portuguesa: datação e evolução crono-estilística." *Revista Portuguesa de Arqueologia* 14: 355–373.

Casimiro, T. M. 2022. Porque importa saber de cerâmica portuguesa no colonialismo atlântico, 230–249. https://doi.org/10.15210/LEPAARQ.V19I37.22651.

Casimiro, T. M., R. V. Gomes, and M. V. Gomes. 2015. "Portuguese Faience Trade e Consumption Across the World (16th -18th Centuries)." In *Global Pottery Proceedings – 1st International Conference for Historical Archaeology e Archaeometry for Societies in Contact. BAR International Series 2761*, edited by J. Buxeda I Garrigós, M. Madrid I Ferneez, and J. Iñañez, 67–80. Oxford: Hadrian Books.

Chollet, F., and Others. 2015. *Keras*. Retrieved from https://keras.io

Cobb, P. 2023. "Large Language Models and Generative AI, Oh My!: Archaeology in the Time of ChatGPT, Midjourney, and Beyond." *Advances in Archaeological Practice* [Online] 11: 363–369. https://doi.org/10.1017/aap.2023.20.

Domínguez-Rodrigo, M., G. Cifuentes-Alcobendas, B. Jiménez-García, N. Abellán, M. Pizarro-Monzo, E. Organista, and E. Baquedano. 2020. "Artificial Intelligence Provides Greater Accuracy in the Classification of Modern and Ancient Bone Surface Modifications." *Scientific Reports* 10 (1): 2045–2322. https://doi.org/10.1038/s41598-020-75994-7.18862.

Elhassouny, A., and F. Smarandache. 2019. "Trends in Deep Convolutional Neural Networks Architectures: A Review," *International Conference of Computer Science and Renewable Energies* (ICCSRE), Agadir, Morocco, 2019, 1–8. https://doi.org/10.1109/ICCSRE.2019.8807741.

Filipe, V. 2023. *Olisipo, o grande porto romani da fachada atlântica. Economia e comércio entre a República e o Principado*. Lisboa: Tese de doutoramento. Universidade de Lisboa.

Gualandi, M., G. Gattiglia, and F. Anichini. 2021. "An Open System for Collection and Automatic Recognition of Pottery through Neural Network Algorithms." *Heritage* 4 (1): 140–159. https://doi.org/10.3390/heritage4010008.

Gutherz, G., S. Gordin, L. Sáenz, O. Levy, and J. Berant. 2023. "Translating Akkadian to English with Neural Machine Translation." *PNAS Nexus* [Online] 2 (5): pgad096. https://doi.org/10.1093/pnasnexus/pgad096.

Guyon, I., S. Gunn, M. Nikravesh, and L. A. Zadeh. 2008. *Feature Extraction: Foundations and* Applications, vol. 207. Berlin: Springer.

Harris, O., and C. Cipolla. 2017. *Archaeological Theory in the New Millennium. Introducing Current Perspectives*. London: Routledge.

Heilen, M., and J. Altschul. 2013. "The Accuracy and Adequacy of In-Field Artifact Analysis: An Experimental Test at Two Archaeological Sites in the Western United States." *Advances in Archaeological Practice* 1 (2): 121–138. https://doi.org/10.7183/2326-3768.1.2.121.

LeCun, Y., Y. Bengio, and G. Hinton. 2015. "Deep Learning." *Nature* 521: 436–444. https://doi.org/10.1038/nature14539.

Li, Q., W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen. 2014. "Medical Image Classification with convolutional neural network." 2014 13th international conference on control automation robotics & vision (ICARCV), 844–848. https://doi.org/10.1109/ICARCV.2014.7064414.

Li, P., X. Chen, and S. Shen. 2019. "Stereo R-CNN Based 3D Object Detection for Autonomous Driving." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

Lucena, M., A. Martínez-Carrillo, J. M. Fuertes, F. Carrosca, and A. Ruiz. 2016. "Decision Support System for Classifying Archaeological Pottery Profiles Based on Mathematical Morphology." *Multimedia Tools and Applications* 75: 3677–3691. https://doi.org/10.1007/s11042-014-2063-6.

Murphy, K. P. 2022. *Probabilistic Machine Learning: An Introduction*. Cambridge, MA: MIT Press.

Navarro, P., C. Cintas, M. Lucena, J. Fuertes, C. Delrieux, and M. Molinos. 2021. "Learning Feature Representation of Iberian Ceramics with Automatic Classification Models." *Journal of Cultural Heritage* 48: 65–73.

Orengo, H., F. Conesaa, A. Garcia-Molsosaa, A. Lobob, A. Greenc, M. Madellad, and C. Petrie. 2020. "Automated Detection of Archaeological Mounds Using Machine-Learning Classification of Multisensor and Multitemporal Satellite Data." *Proceedings of the National Academy of Sciences* 18240–18250. https://doi.org/10.1073/pnas.2005583117.

Parisotto, S., N. Leone, C. Schönlieb, and A. Launaro. 2022. "Unsupervised Clustering of Roman Potsherds via Variational Autoencoders." *Journal of Archaeological Science* 142: 1055598.

Pawlowicz, L. M., and C. E. Downum. 2021. "Applications of Deep Learning to Decorated Ceramic Typology and Classification: A Case Study Using Tusayan White Ware from Northeast Arizona." *Journal of Archaeological Science* 130. https://doi.org/10.1016/j.jas.2021.105375.

Pinto, I., and A. P. Magalhães. 2016. "Almagro 51A-B (Western Lusitania)." *Amphorae ex Hispania. Landscapes of Production and Consumption*. http://amphorae.icac.cat/amphora/almagro-51a-b-western-lusitania.

Quaresma, J. C. 2023. "Lusitanian Food-Stuff Trade at Arles from 1st to 5th Century AD." In *LRCW 6: Late Roman Coarse Wares, Cooking Wares and Amphorae in the Mediterranean: Archaeology and Archaeometry. Land and Sea: Pottery Routes. Agrigento. 24–28th May 2017*, edited by V. Camineci, E. Giannitrapani, M. C. Parello, and M. S. Rizzo, 19–26. Oxford: Archaeopress.

Quaresma, J. C., and J. Raposo. 2016. "Lusitana 3 (Western Lusitania)." *Amphorae ex Hispania. Landscapes of production and consumption*. http://amphorae.icac.cat/amphora/lusitana-3-western-lusitania.

Quaresma, J. C., J. Santos, M. Lopes, and P. Calaveiras. 2023. "A Ânfora vinária Lusitana 3: Diferenças morfológicas Entre Tejo e Sado." *SAGVNTVM (P.L.A.V.)* 55: 183–200.

Raposo, J., and C. Viegas. 2016. "Dressel 14 (Western Lusitania)." *Amphorae ex Hispania. Landscapes of production and consumption*. http://amphorae.icac.cat/amphora/dressel-14-western-lusitania.

Resler, A., R. Yeshurun, F. Natalio, and R. Giryes. 2021. "A Deep-Learning Model for Predictive Archaeology and Archaeological Community Detection." *Humanities and Social Sciences Communications* [Online] 8 (1): 1–10. https://doi.org/10.1057/s41599-021-00970-z.

Sejnowski, T. J. 2018. *The Deep Learning Revolution*. Cambridge, MA: MIT Press.

Sharafi, S., S. Fouladvand, I. Simpson, and J. Alvarez. 2016. "Application of Pattern Recognition in Detection of Buried Archaeological Sites Based on Analyzing Environmental Variables, Khorramabad Plain, West Iran." *Journal of Archaeological Science: Reports* 8: 206–215. https://doi.org/10.1016/j.jasrep.2016.06.024.

Southampton City Council. 2013. *Chapel, Royal South Hants Hospital, Fanshawe Street, Southampton (SOU1570)* (data-set). York: Archaeology Data Service (distributor). https://doi.org/10.5284/1018074.

Viegas, C., J. Raposo, and I. Pinto. 2016. "Almagro 51C (Western Lusitania)." *Amphorae ex Hispania. Landscapes of Production and Consumption.* http://amphorae.icac.cat/amphora/almagro-51c-western-lusitania.