Classification of Pottery Shards from Diverse Geographical Regions Based on XRF Profiles[1]

Ankita Nandy
Department of Computing,
Asia Pacific University,
Kuala Lumpur, Malaysia

DOI:10.37648/ijrst.v13i03.005

Received: 12 June 2023; Accepted: 29 July 2023; Published: 30 July 2023

ABSTRACT

Pottery fragments found in archaeological sites across the world provide insights into the prevalent manufacturing technology, commercial usage of wares and the socio-politico-economic fabric of the societies which crafted them. Their chemical profile can be used to characterise the clay used in their making, and thus, locate their origins. The technologies for the generation of such geochemical profiles have been around for decades, and several researchers have published the results for their samples. However, such data has undergone just basic statistical analysis. This work collates such data from multiple sources and performs a comparative analysis of multiple machine learning classifiers, to showcase the potential utility of bringing up such datasets for further exploration. It can speed up the segregation and mapping of historical artefacts and add value to archaeological teams working in developing countries of Asia and Africa.

Keywords: *pottery shards; XRF; machine learning; classification; provenance*

INTRODUCTION

As the clay used in pottery varies from region to region, analysis of such shards, as described in [1], can establish if the fragments are of local origin or have been brought in from elsewhere through trade or migration [2]. Finer details, such as components used for coloring or physical features such as the decorative elements, can indicate the socio-economic hierarchies of the inhabitants. These shards have been playing a valuable role in the reconstruction of the historical events based on available evidences. Such analysis has been traditionally undertaken through visual observation and mapping of physical features like the color, texture, and patterning [3]. Petrological microscopes were employed in studying the mineral composition in such shards [1], however, this is a laborious, time-consuming activity. Further advancements led to the inclusion of technologies like the neutron activation analysis, particle induced x-ray emission [3] et cetera, which generate the elemental profile of the sample.

X-ray fluorescence (XRF) is a standard in the geochemical analysis of rock samples in geology. It is a non-destructive method and demonstrates high spectral resolution. Decades of research and commercial usage has led to diversification of XRF, such as Energy Dispersive XRF (ED-XRF) and Wavelength Dispersive XRF (WD-XRF). When a sample is radiated with X-rays, the excitation results in emissions of certain wavelengths from the sample, which can be mapped to the characteristic emissions of certain elements; thus, confirming their presence, and relative concentrations. The preparation of the sample for XRF is elaborated in [4]. The samples are thoroughly cleaned to remove mud and contaminants from the surface. They are dried and heated, and ground into fine powder. The powders are systematically organized as homogenous pellets. Spectroscopy instrument can be expensive and requires skilled handling. There are limitations to any technique, therefore additional and alternative methods are used to complement the results.

[4] obtained their samples from the archaeological sites in northeastern Tamil Nadu, which was an important center under the Mughal rule in the early 18th century, and later annexed by the British under the Doctrine of Lapse. [5]

[1] How to cite the article: Nandy A. (July 2023); Classification of Pottery Shards from Diverse Geographical Regions Based on XRF Profiles; *International Journal of Research in Science and Technology*, Vol 13, Issue 3, 60-64, DOI: http://doi.org/10.37648/ijrst.v13i03.005

analyses samples from Dahan-e Ghulaman archaeological site in Sistan, Iran which date back to 5th and 6th century BCE. This site is believed to be the capital of Sistan in the heights of glory. [6] analyses ancient pottery shards obtained from archaeological sites of Conjunto Vilas and São João in Amazonia. [2] compares the samples from Mleiha in the Emirate of Sharjah to samples from archaeological sites in present day Gujarat and Maharashtra in India. These samples date back to late 3rd century BCE to 6th century CE. [7] applies XRF to samples gathered from the northern coast of the Black Sea, which date back to 12th century BC to 7th century BCE. To study other aspects like the non-plastic inclusion and the matrix composition, other forms of analyses have been incorporated.

Data gathered through XRF were visualized, clustered, or transformed through principal component analysis (PCA) et cetera [2][4] to find assess the similarities and find distinguishing features.

The rapid advancement in computational systems has propelled a surge of interest in machine learning across all domains, archaeology not being an exception. [8] presents a provenance analysis system built on R to classify and cluster pottery shards based on their chemical properties and demonstrates it on 208 samples. With advancement in spectroscopy techniques, the analysis of the chemical profile of pottery fragments has been widely adopted, with statistical techniques for detecting compositional similarities and distinguishing features. [9][10] train convolutional neural networks to categorize artefacts by their appearance, however, an expert's keen eye is essential in placing a historical artifact in the right context, and its importance cannot be overlooked [11]. But automation of some aspects using machine learning tools provides the initial categorization which expert knowledge can improve. Therefore, in provenance analysis, machine learning can be a useful assistant.

In this work, the XRF generated profiles of ancient pottery samples collected from four regions, diverse spatially and chronologically, are visualized and then classified based on the region of origin, using the percentage composition of major constituents as features.

MATERIAL AND METHODS

Data. Data for this work has been gathered by consolidating the compositions of samples as published by [2][4][5][6][7]. The assimilation of data from past works facilitates their reuse for additional analysis without reinventing the wheel. The samples have been labelled as in Table 1.

Visualization is the window to the data. Using the ggplot2 library in R, the violin plots of the chemical compositions are generated. The data is scaled using max-min scaling. The objective of scaling is to avoid the dominance of elements like silica and aluminum over elements in lower quantities like titanium.

$$X_{scaled} = \frac{X_{original} - minimum}{maximum - minimum}$$

Four supervised classification techniques are implemented using scikit-learn in Python.

Decision Tree. In the building of a decision tree, measures such as Gini Index, Entropy, et cetera are employed to partition the data recursively and segregate the records into classes, such that the intermixing of records corresponding to different classes can be minimized. Although a weak classifier, the decision tree algorithm owes its popularity to the understandability of the classification. The tree generates a set of rules which explain the segregation, which can be used for decision making in practical scenarios.

Support Vector Machines. This algorithm projects the existing data to higher dimensional space, which facilitates the separability of classes. The records critical to creating the decision boundaries are referred to as support vectors.

k Nearest Neighbors. The most intuitive of machine learning algorithms, the majority class of the k neighbors nearest to the object of study is assigned to it. The distance between objects is calculated using Euclidean distance, while k is a parameter specified by the user. The drawback is the high memory requirement for calculating pairwise distance between all records in large datasets.

Random Forest. Though decision tree is a weak classifier, a collection of trees, hence the name forest, trained on randomly generates subsets of the data, when work together in assigning a class to the object under study, generate impressive results.

(IJRST) 2023, Vol. No. 13, Issue No. 3, Jul-Sep **e-ISSN: 2249-0604, p-ISSN: 2454-180X**

Oversampling. As classes with samples in minority might get overlooked by the classifier as noise, Synthetic Minority Oversampling Technique (SMOTE) is used to populate those classes and thus fix the imbalance. The implementation of SMOTE in the imblearn library is used.

Cross Validation. As the data is limited, instead of splitting it into test and train, k fold cross validation with k set to 5 is carried out. In this method, the available data is split into k sections, with iteratively using any (k-1) of them for training and the k^{th} one for validation. The mean accuracy of these iterations is used to assess the performance of the classifier. Python and R are used for this work.

RESULTS

Figure 1 presents the element-wise comparisons of samples across locations. From the element-wise comparisons, samples from Tamil Nadu, are low in aluminosilicates in comparison to other samples. In terms of Calcium Oxide, Iron Oxide, Potassium Oxide and Titanium Oxide, the samples from Western India, have a large interquartile range. The samples from Amazonia show the lowest percentages of Calcium Oxide and Potassium Oxide in this collection.

The default settings as per the scikit-learn library in Python are used for building the classification models. The mean accuracy scores as listed in Table 2 shows that SVM Classifier and Random Forest deliver the best results, with an average of 98.9%. It is followed by k-Nearest Neighbors at 96% and the Decision Tree has been the worst performer of the set at 88.6%.

Table I. Data labels and corresponding locations

Label	*Locations*
A	Melpadi, Paiyampalli, Thirumani, Karivadu, Arcot, Vallimalai, Kavaripakkam, PanchaPandavarmali, Walajah, Udayandiram in Vellore, Tamil Nadu (Chandrasekaran, Naseerutheen and Ravisankar, 2017)
B	Dahan-e Ghulaman, Sistan (Sarhaddi-Dadian, Moradi, Zuliskandar and Purzarghan, 2017)
C	Dwarka, Prabhas Patan, Padri, Nevasa, Junnar and Nasik (Reddy, Attaelmanan and Mouton, 2012)
D	Conjunto Vilas and São João in Amazonia (Oliveira et al., 2020)
E	Dykyj Sad, Subotiv, Glinjeni II-La Șanț, Kartal and Nemyriv in the northern coast of Black Sea (Daszkiewiczm et al., 2020)

Table II. Model vs Mean Cross Validation Accuracy

Model	*Mean Cross Validation Accuracy*
Decision Tree	0.886
SVM Classifier	0.989
K Nearest Neighbors	0.960
Random Forest	0.989

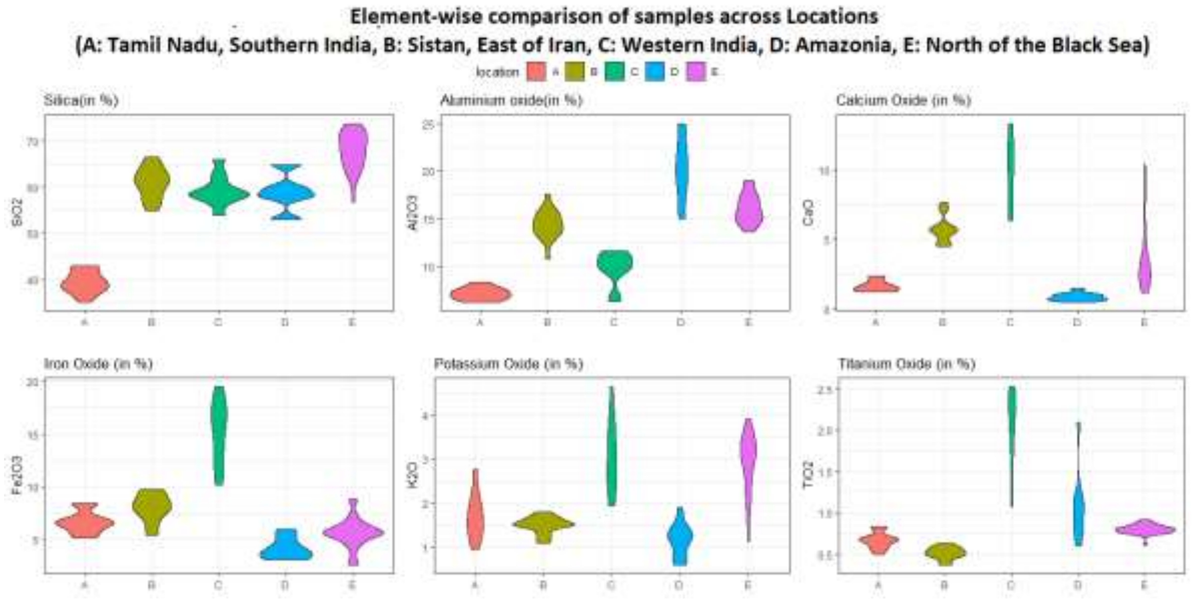

Figure 1. Violin Plot: Element-wise comparisons of samples across locations

CONCLUSION

The geographically diverse data was classified with accuracies over 90% by three out of four classifiers. These results encourage the increased adoption of machine learning in archaeological segregation. This can add value to explorations of ancient civilizations in India and African countries where a plethora of artefacts are discovered, some locally produced, some traded or carried over by nomadic groups, but lie in wait of dedicated analysis. Machine learning tools can automate part of the process in identification and mapping.

REFERENCES

1. Pillay, A. E., Punyadeera, C., Jacobson, L., & Eriksen, J. (2000). Analysis of ancient pottery and ceramic objects using x-ray fluorescence spectrometry. *X-Ray Spectrometry: An International Journal*, *29*(1), 53-62.
2. Reddy, A., Attaelmanan, A. G., & Mouton, M. (2012, July). Pots, plates and provenance: sourcing Indian coarse wares from Mleiha using X-ray fluorescence (XRF) spectrometry analysis. In *IOP Conference Series: Materials Science and Engineering* (Vol. 37, No. 1, p. 012010). IOP Publishing.
3. Wu, Q. Q., Zhu, J. J., Liu, M. T., Zhou, Z., An, Z., Huang, W., ... & Zhao, D. Y. (2013). PIXE-RBS analysis on potteries unearthed from Lijiaba Site. Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, 296, 1-6.
4. Chandrasekaran, A., Naseerutheen, A., & Ravisankar, R. (2017). Dataset on elemental concentration and group identification of ancient potteries from Tamil Nadu, India. *Data in brief*, *10*, 215-220.
5. Sarhaddi-Dadian, H., Moradi, H., Zuliskandar, R., & Purzarghan, V. (2017). X-ray fluorescence analysis of the pottery shards from Dahan-e Ghulaman, the Achaemenid site in Sistan, east of Iran. *Interdisciplinaria Archaeologica*, *8*(1), 35-41.
6. Oliveira, L. S. S., Abreu, C. M., Ferreira, F. C. L., Lopes, R. C. A., Almeida, F. O., Tamanaha, E. K., ... & Souza, D. N. (2020). Archeometric study of pottery shards from Conjunto Vilas and São João, Amazon. *Radiation Physics and Chemistry*, *167*, 108303.
7. Daszkiewicz, M., Gavrylyuk, N., Hellström, K., Kaiser, E., Kashuba, M., Kulkova, M., ... & Winger, K. (2020). Possibilities and limitations of pXRF as a tool for analysing ancient pottery: a case study of Late Bronze and Early Iron Age pottery (1100–600 BC) from the northern Black Sea region. *Praehistorische Zeitschrift*, *95*(1), 238-266.
8. Anglisano, A., Casas, L., Queralt, I., & Di Febo, R. (2022). Supervised Machine Learning Algorithms to Predict Provenance of Archaeological Pottery Fragments. *Sustainability*, *14*(18), 11214.

(IJRST) 2023, Vol. No. 13, Issue No. 3, Jul-Sep **e-ISSN: 2249-0604, p-ISSN: 2454-180X**

9. Iyer, A., & Franklin, M. (2022). AI-Powered Archaeology: Determining the Origin Culture of Various Ancient Artifacts Using Machine Learning. *Journal of Student Research*, *11*(1).
10. Pawlowicz, L. M., & Downum, C. E. (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*, 105375.
11. Bickler, S. H. (2021). Machine learning arrives in archaeology. *Advances in Archaeological Practice*, *9*(2), 186-191.