Automatic identification of Dong Son antique glass artifacts using evolving learning

Ngo-Ho Anh-Khoi[1*], Bui Hoang-Bac[2], and Pham Van-Trieu[3]

[1]*Nam Can Tho University, Viet Nam*

[2]*Hanoi University of Mining and Geology, Viet Nam*

[3]*Vietnamese Institute of Archaeology, Viet Nam*

*Corresponding author (ngohoanhkhoi@gmail.com, nhakhoi@nctu.edu.vn)

ABSTRACT

Regarding the Dong Son culture, given the diverse range of artifacts discovered, we propose the utilization of an artificial intelligence system for the automated and comprehensive identification of Dong Son glass jewelry through SEM gemological analysis. This approach, which has gained prominence in the field of archaeology worldwide over the past five years, aims to integrate advanced technology into Vietnamese archaeology. Our research is motivated by the unique conditions present in archaeology, where we seek to employ evolving learning algorithms to archaeological databases, comparing and selecting the most suitable model that aligns with the archeological dataset's performance. We have developed the Recognition Automatic System for Dong Son Antique Glasses (RAS-DSA), capable of accurately distinguishing between Dong Son and non-Dong Son glass ornaments, and is freely distributed to experts and archaeologists. This collaborative research involves Nam Can Tho University, Hanoi University of Mining and Geology, the Vietnamese Institute of Archaeology, and the UNESCO Center for Research and Conservation of Vietnamese Antiquities.

1. INTRODUCTION

Throughout ancient times and even in the present day, ancient Vietnamese glassware such as Dong Son, Sa Huynh, and Oc Eo artifacts (relative with Dong Son Culture, Sa Huynh Culture, and Oc Eo Culture) have been widely traded and can be found in various regions of Viet Nam and across the globe. However, within private antiquities collections, there is a prevalent and intricate issue of misclassifying and confusing these ancient jewels. Specifically, with the Dong Son culture, jewelry is often sold under different names, while ornaments from other cultures like Sa Huynh and Oc Eo are frequently mislabeled as Dong Son.

This paper aims to address this challenge by analyzing the distinct identification characteristics of Dong Son glass jewelry using a combination of gemological methods and advanced artificial intelligence. The ultimate aim is to develop a valuable tool that enables scientists to automatically differentiate between Dong Son glassware and other groups of ancient glassware within the region, as well as identify modern glass jewelry that imitates antiquated styles.

In this paper, our aim is to present our research through three main sections. First, we will summarize the interdisciplinary research conducted at the intersection of gemology, archaeology, and machine learning. The subsequent section will delve into the details of our archaeology dataset, the

gemological measuring equipment used, and the experimental procedures implemented. We then will offer a concise yet comprehensive overview of our learning methods, evaluation metrics, and experimental protocols, accompanied by a detailed discussion of our experimental results. Finally, we will conclude the paper by emphasizing the overall significance and value of our findings.

1.1. State of Art of gemological ancient glass analysis research

The study of ancient glass using gemological methods emerged in the 1990s, with J. Henderson's series on Roman glass in England being a notable contribution (Henderson, 1991). The gemological approach for categorizing ancient glass is based on their characteristics and composition-related components. Scholars recognize two primary categories of antique glass based on the glassmaking process: lead glass and natron glass. Lead glass, invented by the Mesopotamians, reached China through the Silk Road and became known as Han glass or Oriental glass. Natron glass, invented by the Romans, appeared in various Mediterranean cultures such as Egypt, Africa, and Arabia, and is often referred to as ancient Roman or Western ancient glass (Tait, 2004).

Studies on ancient glass in Asia are relatively limited and have only recently gained attention in a few studies conducted in Thailand and China. In Viet Nam, there have been only a few separate French studies on this topic, with little local research. Most of the research in this field has been conducted in Europe, primarily focusing on the natron glass category, and the research framework has been relatively mature for over 30 years. One of the earliest taxonomies, proposed by Sayre (Sayre, 1961) divided natron glasses into two main categories based on their components. The first category is lmlp (low-magnesia, low-potash glasses), which were popular in ancient cultures, including the Romans, from the 1st millennium BC until the 9th century AD. The second category is hmhp (high-magnesia, high-potash glasses), found in Central Asia before the 7th century BC and in the Arab cultural area before the 9th century BC. Another classification, proposed by Freestone (Freestone et al., 2002) and Wedepohl (Wedepohl et al., 2000), is based on the calcium (Ca) content and divides the glasses into two smaller categories. However, for the convenience of studying ancient glass with Vietnamese characteristics, we have chosen to divide them into seven categories based

on suggestions by scholars such as Wedepohl, despite disagreement from Freestone. Rehren (Rehren, 2015) proposed a method of dividing the mixtures of Freestone's and Wedepohl's categories into six groups of ancient glass based on the material source, encompassing both natron and lead glass families. Recent research, such as (Ngo-Ho, 2019a, 2019b, 2020) has published gemological analyses of Oc Eo glass specimens with hybrid propositions of six categories.

While categorizing ancient glass into distinct groups can be a simplistic method, it does not meet modern requirements. Each group of ancient glass can be further divided into smaller subgroups, each exhibiting distinct differences despite sharing common characteristics within the broader category. This complexity becomes evident when different types of glass are created within the same tradition but in different regions, resulting in significant archaeological variations. However, from a gemological perspective, these differences may not be pronounced enough to be discerned using rudimentary mathematical tools. On the other hand, in many cases in archaeology, an initial group of artifacts classified as Group A belonging to Culture A may be reclassified into another group in the future, depending on the available data at that time and subsequent discoveries. Classifying an artifact or a group of artifacts into a specific category or classification relies on various interdisciplinary factors, many of which are related to future excavations, and some of these factors can be completely unexpected. This makes classification studies highly complex because previous research data on gemology, for example, can be completely different, rendering those studies useless later on. This also poses challenges for data science disciplines, which will be discussed further in the next section, regarding the phenomenon of "*concept drift*".

Therefore, it is natural that there is a growing expectation for artificial intelligence methods with improved classification capabilities in many research institutions, especially during the recent explosive development in this field. The next section of the study will present these expectations.

1.2. State of Art of artificial intelligence applied for identification of antique artifacts

Currently, the application of artificial intelligence technologies for the classification, appraisal, assessment, and judgment of antiquities, sites, and heritages has gained significant popularity in recent

years. This trend has been observed in studies such as Bickler's examination of antique porcelain (Bickler, 2018), Jones' classification of ancient plants (Jones et al, 2019), and Ngo-Ho's works on the classification and evaluation of ancient documents (Ho et al., 2013, 2014, 2015, 2017). The application of AI also includes aerial archaeology, employing various techniques for detection through aerial images (Alexandra, 2022; Tran et al., 2022). Recently, as a result, we are gradually establishing the first automatic identification systems for Vietnamese antiquities using advanced technologies (Ngo-Ho, 2019c).

In our previous papers, we utilized a neural network artificial intelligence system that simulates the human brain system, a method proven effective in similar systems presented in Bhuvaneswari's, Karlik's, and Truong's studies (Bhuvaneswari et al., 2013; Bekir et al., 2011; Truong et al., 2014). After obtaining the feature sets extracted from the SEM system, which includes a selection of 36 features, we trained the neural networks for recognition, comparing them to select the best network. The selected network in this application comprised 36 input values, 120 hidden layer neurons, and one output layer neuron. Through the training process, we obtained a set of networks suitable for identification. During the identification process, the application continuously collects data directly entered by the user, extracts features, and feeds them to the network for further identification. As a result, the results and efficiency gradually improve through continuous learning, enhancing the accuracy of the final outcome. The attributes of the analyzed rays are selected based on characteristic properties, vectorized, and transformed into a 36-dimensional vector $\{x1, x2, ..., x36\}$. Although we have installed a system that builds a single network for specimen recognition, it can be extended to handle multiple types of specimens simultaneously. The neuron set comprises three layers: the input value layer, the hidden neuron layer, and the output neuron layer. There are two sets of neuron weights corresponding to the links from the input layer to the hidden layer and from the hidden layer to the output layer. Considering the characteristics of artifact identification, we have chosen the feedforward neural network architecture combined with the error backpropagation algorithm as the network architecture type.

In this article, our main objective is to improve upon a recently published system by integrating additional learning methods through the utilization of evolving approaches. The inspiration behind this concept arises from the distinctive characteristics of stream data in the field of archaeology. Unlike traditional datasets, archaeological data accumulates gradually over time as new discoveries arise from excavations. This research presents significant challenges due to the scarcity of data, which can be attributed to three primary reasons.

Firstly, the real-life scenarios of archaeological investigations make it challenging to find artifacts, and conducting experimental excavations to extract features from these artifacts can be prohibitively expensive. As a result, data is acquired sporadically over time and in limited quantities. Secondly, although archaeological research institutions allow the use of experimental feature extraction results and the final model, they do not permit the storage of the actual archaeological data. Consequently, the algorithms employed in the experiments must prove their effectiveness under these unique circumstances. And the last one, in the field of archaeology, there are instances where artifacts initially classified as belonging to a group may be reclassified in the future based on new data and discoveries. The classification of artifacts into specific categories or classifications depends on various interdisciplinary factors, many of which are associated with future excavations. Some of these factors can be completely unexpected. As explaned in the previous section, this also presents challenges for data science disciplines, particularly concerning the phenomenon known as "*concept drift*" which will be explored further in the subsequent section.

Therefore, given the limitations of traditional learning methods, evolving continuous learning methods become more relevant and intriguing for addressing this problem. In this study, we will implement and compare the results of evolving continuous learning methods using our archaeological dataset. A comprehensive overview of these evolving continuous learning methods has been extensively studied in Ho et al. (2014, 2015), especially in archaeology (Ngo-Ho, 2023a, 2023b). Considering the specific context of archaeology, we acknowledge that the system can effectively utilize lightweight methods based on data selection approaches due to the limited samples obtained from archaeological excavations. This approach ensures that important information is not lost due to the generalization concept employed by many evolving methods. To achieve this, we propose employing a simple "*sliding windows*" approach, similar to the FLORA approach (Widmer & Kubat,

1996). This principle involves updating the model at each moment '*t*' using the most recent training data, defined by a sliding window of a predetermined size (either based on time scale or number of data points). This approach can involve either batch retraining using the selected data within the sliding window or updating the model if an online learning method allows for it.

Typically, these methods comprise three steps (Bitfet & Gavalda, 2007):

1. Detecting concept changes using statistical tests on different windows.

2. If a change is observed, selecting representative and recent examples to adapt the models.

3. Updating the models.

The size of the window is determined a priori by the user, and each window overlaps the previous one by sharing a batch of data. At each step, a new model is learned, which represents an updated set of classes. The crucial aspect of these approaches lies in determining the most appropriate window size. While most methods employ a fixed-size window configured for each real-world problem, there are approaches aimed at automatically detecting the optimal size of the analysis window.

For instance, Bitfet & Gavalda, (2007) with ADWIN (ADaptive WINdow), tested a set of window sizes by dividing each window into sub-windows of minimal size. If the sub-windows exhibit sufficiently different distributions, a statistically significant size is considered a good choice. Lazarescu et al., (2003), propose using two models at each step, each trained with a different window size: *S* (a predefined standard size) and *2S*. The smaller window with size *S* was utilized to detect new concept spaces using a statistical test, while the larger window with size *2S* was employed to update the model upon detecting a new concept space. Last (2002) used OLIN (On Line Information Network), to suggest dynamically adjusting the window size based on performance achieved on a validation dataset. The new data was divided into two parts: one for training and the other for validation. Multiple windows with varying sizes were independently applied for learning and testing, and the size that yields the best result on the validation data was selected for the current step. However, to implement this approach effectively, it requires conducting learning phases on sufficiently large batches of data. Lastly, in (Klinkenberg,

2004), the author applies a consecutive increment of window sizes. At each step, the performance (in terms of *error rate*) is calculated for different window sizes, and the size that yields the best performance is chosen (e.g., size *No1* represents the last batch, size *No2* represents the last two batches, size *No3* represents the last three batches, and so on). In this article, we will explore the evolving approaches based on Klinkenberg's idea (Klinkenberg, 2004), applied to our archaeological dataset, to dynamically detect the best window size for analysis. Hence, it is essential for the algorithms used in the experiments to show their efficacy within these exceptional conditions. Considering the specific context of archaeology, as explained previously, we recognize the potential effectiveness of employing lightweight methods that leverage data selection approaches. This is important because of the limited number of samples obtained from archaeological excavations. The underlying rationale is straightforward: we can exercise control over information loss by managing the "*density*" of learning data within the optimal size selected. By utilizing Klinkenberg's methods, with only a simple parameter *n* as the size of the batches, we can control the loss of valuable information caused by the generalization concept commonly employed in various evolving techniques.

2. MATERIAL AND METHODS: ARCHAEOLOGICAL DATASET

The Oc Eo, Sa Huynh, and Dong Son specimens utilized in this study have been showcased in numerous national exhibitions, authorized by the National Appraisal Council, and sourced from the collection of the UNESCO Center for Research and Conservation of Vietnamese Antiquities. To extract the gemological characteristics of these antique specimens, we employed the gemological Scanning Electron Microscope (SEM) technique. A comprehensive description of the SEM technique can be found in our previous studies (Ngo-Ho, 2019a, 2019c). In summary, this method revolves around the utilization of Energy Dispersive X-ray Spectroscopy (EDXS) or Wavelength Dispersive X-ray Spectroscopy (WDXS) analysis. These techniques involve analyzing the chemical composition of solids by recording the X-ray spectrum emitted when the solid interacts with radiation, typically high-energy electron beams in electron microscopes (Figure 1). In our study, we employed the SEM machine combined with the EDS exploitation machine (Model: Quanta 450;

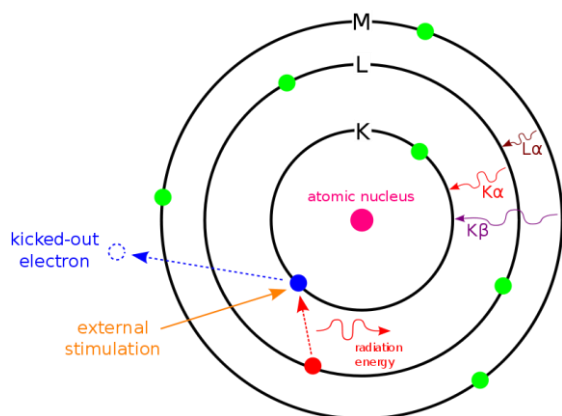Manufacturer: FEI-USA) to facilitate this physical technique.

Figure 1. The technique is described in terms of electron interactions.

The SEM technique provides a combination of 10 values for each chemistry element in the results. The effective reflectance (Net Int) represents the reflectance of the electron beam, while Weight and Atomic Mass (%) and Atoms (%) indicate the composition of the specimen. The Kratio value represents the ratio of the reflected electron density, and R signifies the resolution in microns after determining the reflected electron density. The elemental composition is calibrated by comparing the spectrum of the standard sample with the spectrum of the measured sample. Calibrating

values such as Z (atomic calibration), A (absorption correction), and F (fluorescence calibration) are used to ensure accuracy and adjust for specific factors. Table. 1 depicts the Structure of SEM Measurement Results: for each detected chemical element, SEM technique will calculate 9 indicators; Table 1 is a demo result for iron (Fe).

In this particular experiment, the antique Oc Eo glass specimens consist of eight chemical elements: Calcium (Ca), Potassium (K), Iron (Fe), Sodium (Na), Magnesium (Mg), Aluminium (Al), Silicon (Si), and Oxygen (O). The location and number of analytical processing times are crucial for identification. Given the expense of each analysis shot, only necessary positions are analyzed to avoid data redundancy. Therefore, we follow specific conditions in our measurements: depending on the sample quality, each sample is analyzed five times at different locations on the specimen. By selecting distinct sites for analysis, we aim to achieve maximum variation. Homogeneous samples may require fewer analyses, while larger samples may undergo fewer shots. For Dong Son specimens, we analyze five different locations within each specimen, with five shots at each position to ensure analytical quality. For other antique specimens, we analyze one position in each specimen with five shots or less depending on the specificity of the artifacts.

Table 1. Structure of SEM Measurement Results

Element[1]	Weight[2] (%)	Atomic[3] (%)	Net Int.[4]	Error[5] (%)	Kratio[6]	Z[7]	R[8]	A[9]	F[10]
FeK (Iron class K)									

[1]*Element names according to electron subshell*

[2]*Mass (%) in specimen*

[3]*Atoms (%) in specimen*

[4]*Effective intensity obtained in the specimen*

[5]*Errors rate*

[6]*Density of reflected electrons in the specimen*

[7]*Atomic calibration*

[8]*Resolution in microns*

[9]*Absorption correction*

[10]*Fluorescence calibration*

At present, our dataset solely consists of information sourced from the UNESCO Center for Research and Conservation of Vietnamese Antiquities. However, we are soon expecting a data package from the Vietnamese Institute of Archaeology, which will be transferred under technology transfer conditions that exclude direct data transfer, as previously

mentioned. With a total of 108 different analyses collected, including 25 Dong Son specimens, the remaining samples consist of Sa Huynh, Oc Eo, and antique imitations. Each sample is characterized by 40 features representing eight chemistry elements with 5 indicators ('W,' 'R,' 'Z,' 'A,' and 'F,') for each element, different with the previous works with 36

selected features (Ngo-Ho et al., 2019c). For tiny datasets, using any normalization before classification can lead to significant distortion of the true class concept underlying the data, so the study decides not to use any normalization.

3. RESULTS AND DISCUSSION

In this article, we will explore nine different machine learning methods with the incorporation of evolving approaches based on the Klinkenberg concept idea for dynamically detecting the optimal window size, utilizing our archaeological dataset. AdaBoost Classifier (Freund & Schapire, 1995): An ensemble estimation tool that adjusts a classifier on the original dataset and fits additional copies to focus on difficult cases. Gaussian Naïve Bayes (Chan et al., 1979): A variant of Naïve Bayes that follows the Gaussian normal distribution and supports continuous data. Decision Tree Classifier (Breiman et al., 1984): A tree-like flowchart structure that uses decision rules to partition data based on attribute values. K-Neighbors Classifier (Goldberger et al., 2005): A method that finds a

predefined number of training samples closest in distance to a new point and predicts the label based on these neighbors. Extra Tree Classifier (Geurts et al., 2006): A meta estimator that fits randomized decision trees on subsets of the dataset to improve predictive accuracy. Bagging Classifier (Breiman, 1996): An ensemble meta-estimator that fits base classifiers on random subsets of the dataset and aggregates their predictions. Multi-layer Perceptron Classifier (Hinton, 1989): Relies on a neural network to perform classification, optimizing the log-loss function. Random Forest Classifier (Breiman, 2001): A meta-estimator that fits multiple decision tree classifiers on different subsets of the dataset and uses averaging to improve prediction accuracy. Bernoulli Naive Bayes (Lewis, 1998): A variant of Naive Bayes that works well with a transformation for limited binary data. These learning methods are combined with the "*sliding windows*" approach and incorporate Klinkenberg's idea for dynamically detecting the optimal window size to conduct evolving continuous learning.
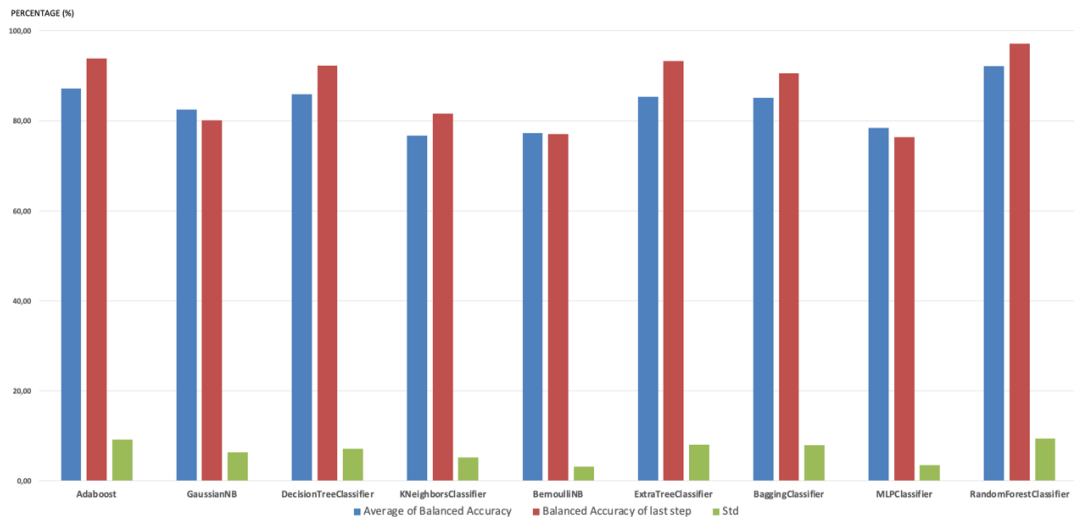

Figure 2. The chart presenting the results of Average of Balanced Accuracy, Balanced Accuracy of last step and Standard deviation of algorithms in the archaeology dataset

The algorithms utilized in this experiment were implemented using version 0.24.2 of the scikit-learn library, combined with the "*sliding windows*" technique. The data is processed in a streaming fashion using mini "*sliding windows*" that consider the last n samples. We begin with *n*=1, which corresponds to classic online learning, and then gradually increase *n* > 1, representing batch learning. For each method, only the best results of *n*

will be selected for comparison with other learning methods.

The archaeology dataset comprises a total of 108 samples, with 25 belonging to Dong Son specimens and the remaining samples belonging to other categories. Each sample is characterized by 40 features representing 8 chemical elements, each having 5 associated indicators. This differs from previous works where only 36 features were

selected (Ngo-Ho et al., 2019c). The dataset is divided into a training dataset (70%) and a testing dataset (30%) with a cross-validation. The positions of the data in the training and testing datasets are randomly assigned for each experiment, which is repeated 10 times. The training dataset is shuffled randomly 10 times to create different scenarios of

stream learning, demonstrating the model's adaptability to evolving data. This means that each learning method with each parameter variation will undergo 100 random experiments before obtaining the final comparison data. The results will be averaged to ensure the reliability and generalizability of the experiments.
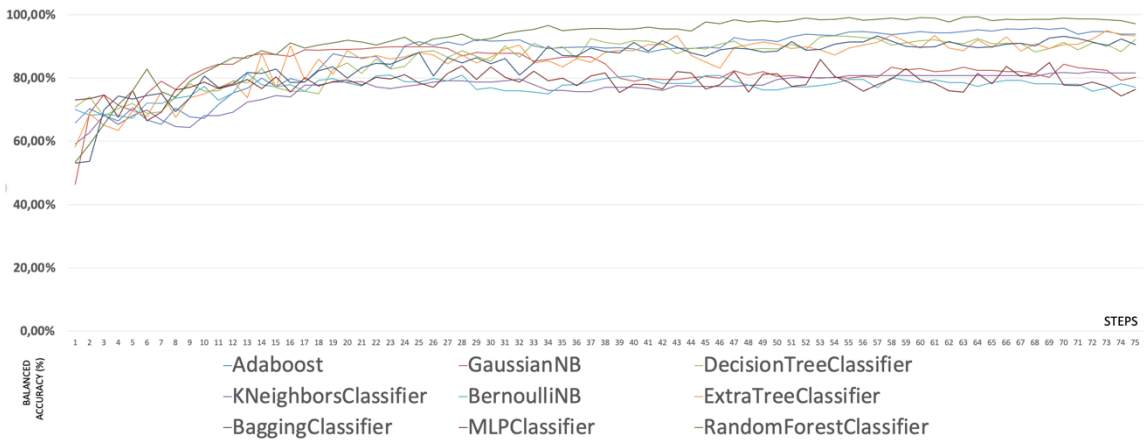

Figure 3. The line graph of learning process of the algorithms

All obtained results are predicated upon the utilization of Balanced Accuracy (BA). The selection of various metrics depends on the specific problem's objectives and the composition of the dataset. In pronounced data imbalance, where, for instance, one group contains only a single data point while another has 999 data points, traditional accuracy computations prove unreliable. Consequently, the area under a receiver operating characteristic (ROC) curve, abbreviated as AUC, and Balanced Accuracy (BA) are often preferred in such scenarios. Metrics like precision, recall, specificity are less appropriate when processing imbalanced class data. If the problem aims to discover consensus, it relies on metrics generated by true positive/false positive, such as precision, recall, and F-Score. Conversely, although less prevalent in real-life scenarios, if the problem seeks to detect non-consensus rather than consensus, it relies on true negative/false negative, such as specificity. The use of Recall, Precision, and F-score is criticized for their limitations as they disregard the true negative cell of the contingency table and are susceptible to manipulation through prediction bias (Powers, 2011). In contrast, Balanced Accuracy considers both true positive and true negative, providing a more balanced assessment for both consensus detection (true positive) and non-consensus detection (true negative). Therefore, when aiming to

address both consensus and imbalanced data conditions, Balanced Accuracy emerges as the most appropriate metric (Ho et al., 2015). Balanced Accuracy (BA) serves as an essential and straightforward metric for evaluating the performance of binary classifiers in situations of class imbalance, where one class is significantly more prevalent than the other. The balanced accuracy formula provides a means to calculate the most realistic and optimal assessment percentage. The formula for balanced accuracy is Balanced Accuracy (BA) = ½ (Specificity + Precision).

Overall, all methods performed well, with the average balanced accuracy exceeding 80% for each method (except k-Neighbors Classifier, MLP Classifier and Bernoulli Naive Bayes). The rankings are Random Forest Classifier (92.12%), Adaboost (87.13%), Decision Tree Classifier (85.87%), Extra Tree Classifier (85.34%), Bagging Classifier (85.13%), Gaussian Naive Bayes Classifier (82.45%), MLP Classifier (78.39%), Bernoulli Naive Bayes Classifier (77.26%), and k-Neighbors Classifier (76.68%); they almost achieved the best results with a window size around 70. It is worth noting that Ensemble Learning methods such as Random Forest Classifier, Adaboost, Extra Tree Classifier, and Bagging Classifier, achieved excellent results compared to other classical

methods. Regarding the final balanced accuracy results, although there were differences in magnitude and some positions changed, the results were similar to the average balanced accuracy. In terms of results, the Random Forest Classifier performed the best, although the Bagging Classifier also achieved excellent results.

Table 2. Table of experiment results

Algorithms	Average (%)	Last Step (%)	Std (%)
k-Neighbors Classifier (*n*=74)	76,68	81,58	5,18
Bernoulli NB Classifier (*n*=73)	77,26	77,00	3,10
MLP Classifier (*n*=63)	78,39	76,34	3,42
Gaussian NB Classifier (*n*=75)	82,45	80,15	6,27
Bagging Classifier (*n*=74)	85,13	90,55	7,86
Extra Tree Classifier (*n*=70)	85,34	93,29	7,98
Decision Tree Classifier (*n*=65)	85,87	92,25	7,13
Adaboost Classifier (*n*=73)	87,13	93,90	9,18
Random Forest Classifier (*n*=57)	92,12	97,11	9,40

The Naive Bayes methods (Bernoulli Naive Bayes and Gaussian Naive Bayes) achieved unsatisfactory results, showing an unclear separation between Dong Son Specimens and the remaining samples. The outcome of the MLP Classifier may appear unexpected initially, but it can be rationalized. Neural network methodologies typically exhibit robust performance in scenarios characterized by abundant data, yet their behavior often diverges in situations marked by data scarcity. The observed results of the MLP Classifier, while initially perplexing, can be explained. In accordance with the theoretical underpinnings of the MLP Classifier and also deep learning methods, it exhibits a consistent and minimal variance in its performance. However, it struggles to achieve efficient classification outcomes when confronted with limited data, mirroring the negative impact observed with the k-Neighbors Classifier. This suggests that representing the concept of Dong Son and other ancient glass as a single category is not effective, possibly due to the inefficiency of representing different types of ancient glass within the same category because of their significant differences. Conversely, the separation between Dong Son and others could be easily solved with "*divide-and-conquer*" approaches like Ensemble Learning or Decision Trees (Random Forest Classifier, Adaboost, Extra Tree Classifier, Decision Tree Classifier, and Bagging Classifier), which all achieved the top 5 performance results. This shows that the structure of the Dong Son class has distinct components that align with the components of other classes. Therefore, with limited data, this alignment becomes a challenge for probabilistic-based machine learning approaches, while it becomes advantageous for "*divide-and-conquer*" approaches, as evidenced by the overall results in this study.

Observing the plotted lines representing the learning progression of the methods, two main groups can be identified. After an initial period, the groups stabilize, with most of them achieving stability around steps 10-12, with results hovering around 80%. Subsequently, the groups gradually diverge and distinctively separate around steps 37 to 40: one group steadily increases and reaches high levels of stability, becoming one of the top 5 most effective methods, while the other group maintains stability. The breakthrough group comprises methods based on "*divide-and-conquer*" approaches; the group that maintains stability comprises methods based on probability factors, as analyzed in the previous section. As expected, the ensemble learning methods exhibit less stability with limited data compared to probabilistic-based methods. Therefore, although these ensemble methods rank among the top 5 most effective methods, they also fall into the top 5 least stable methods when considering the overall standard deviation.

Overall, the Random Forest Classifier, despite having a relatively high overall standard deviation (primarily influenced by the initial phase when seeking stability points), outperforms all other methods with high stability and distinctively superior results.

4. CONCLUSION

This study was conducted within the unique limitations of the archaeology field, where data scarcity poses a significant challenge. This scarcity arises from three primary factors: the difficulty of locating artifacts in real-world contexts, the high

costs associated with conducting excavations for feature extraction, and the sporadic nature of data acquisition. It should be noted that, in archaeology, artifacts initially classified under a specific group belonging to a culture may be subject to reclassification based on additional data and new discoveries. The classification of artifacts or groups into specific categories relies on interdisciplinary factors, some of which may emerge unexpectedly during future excavations. This complexity makes classification studies challenging, as previously conducted research in relevant fields may become outdated and lose relevance. Consequently, these circumstances present challenges for data science disciplines, particularly in relation to the phenomenon of "*concept drift*" which becomes a prominent factor driving the adoption of these methods in archaeological datasets.

The results obtained from the study have been integrated into a software suite called the Recognition Automatic System of Dong Son Glass (RAS-DSG), freely available on the system's website for archaeological experts. While the system shows commendable performance in

recognition, further support is required to expand and enhance the feature dataset.

This paper introduces an advanced artificial intelligence system specifically designed for the automatic identification of Dong Son glass jewelry, utilizing SEM gemological analysis parameters. Implementing this system represents a significant technological advancement in the evaluation and identification of antiquities, surpassing traditional methods and aligning with the latest developments in archaeological centers worldwide. The integration of artificial intelligence technology into archaeological practices is a natural progression in the field.

REFERENCES

Bhuvaneswari, S., & Sabarathinam, J. (2013). Defect analysis using artificial neural network. *IJ Intelligent Systems and Applications*, *5*, 33-38.

Bickler, S. H. (2018). Machine learning identification and classification of historic ceramics. *Archaeology*, *20*, 20-32.

Bifet, A., & Gavalda, R. (2007, April). Learning from time-changing data with adaptive windowing. *In Proceedings of the 2007 SIAM international conference on data mining* (pp. 443-448). Society for Industrial and Applied Mathematics.

Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*, 123-140.

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. CRC press.

Chan, T. F., Golub, G. H., & LeVeque, R. J. (1982). Updating formulae and a pairwise algorithm for computing sample variances. *In COMPSTAT 1982 5th Symposium held at Toulouse 1982: Part I: Proceedings in Computational Statistics* (pp. 30-41). Physica-Verlag HD.

Freestone, I. C., Greenwood, R., & Gorin-Rosen, Y. (2002). Byzantine and early Islamic glassmaking in the Eastern Mediterranean: production and distribution of primary glass. *In Hyalos-Vitrum-Glass: Proceedings of the First Hellenic Glass Conference. Athens: Glasnet Publications* (pp. 167-174).

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, *55*(1), 119-139.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*, 3-42.

Goldberger J., Roweis S., Hinton G., Salakhutdinov R. (2005). Neighbourhood Components Analysis. *Advances in Neural Information Processing Systems*, *17*, 513-520.

Henderson, J. (1990). Chemical characterization of Roman glass vessels, enamels and tesserae. *MRS Online Proceedings Library (OPL)*, *185*, 601.

Hinton, G. E. (1990). Connectionist learning procedures. *Artificial Intelligence*, *40*(1-3), 185 234, 1989. reprinted in j. carbonell, editor. Machine Learning: Paradigms and Methods", MIT Press.

Ho, A. K. N., Ragot, N., Ramel, J. Y., Eglin, V., & Sidere, N. (2013, August). Document classification in a non-stationary environment: A one-class SVM approach. In *2013 12th International Conference on Document Analysis and Recognition* (pp. 616-620). IEEE.

Ho, A. K. N., Ragot, N., Eglin, V., & Ramel, J. Y. (2014, July). Multi one-class incremental SVM for both stationary and non-stationary environment. In *16ème Conférence d'Apprentissage automatique (Cap'2014)*.

Ho, N., Khoi, A. (2015). *Méthodes de classifications dynamiques et incrémentales: application à la numérisation cognitive d'images de documents* (Doctoral dissertation, Tours).

Karamitrou, A., Sturt, F., Bogiatzis, P., & Beresford-Jones, D. (2022). Towards the use of artificial intelligence deep learning networks for detection of archaeological sites. *Surface Topography: Metrology and Properties*, *10*(4), 044001.

Karlik, B., & Olgac, A. V. (2011). Performance analysis of various activation functions in generalized MLP architectures of neural networks. International *Journal of Artificial Intelligence and Expert Systems*, *1*(4), 111-122.

Jones, B., & Bickler, S. H. (2019). Multi-scalar and semi-automatic approaches to detect archaeological features in NZ using Airborne LiDAR data. *Archaeology*, *10*.

Klinkenberg, R. (2004). Learning drifting concepts: Example selection vs. example weighting. *Intelligent data analysis*, *8*(3), 281-300.

Last, M. (2002). Online classification of nonstationary data streams. *Intelligent data analysis*, *6*(2), 129-147..

Lazarescu, M. M., Venkatesh, S., & Bui, H. H. (2004). Using multiple windows to track concept drift. *Intelligent data analysis*, *8*(1), 29-59.

Lewis, D. D. (1998, April). Naive (Bayes) at forty: The independence assumption in information retrieval. In *European conference on machine learning* (pp. 4-15). Berlin, Heidelberg: Springer Berlin Heidelberg.

Ngo-Ho, A.-K., Bui, H.-B., Nguyen, K.-G., Trinh, T.-L. (2019b). Ancient glass stone in history. *International Conference on philosophy therapy and lithotherapy* (ICPTL 2019). HCM city.

Ngo-Ho, A. K., Eglin, V., Ragot, N., & Ramel, J. Y. (2017). A multi-one-class dynamic classifier for adaptive digitization of document streams. *International Journal on Document Analysis and Recognition (IJDAR)*, *20*, 137-154.

Ngo-Ho, A.-K., Nguyen, T.-C., Doan, N.-K., Bui, H.-B., & Trinh, T.-L. (2019a). Distinct gemological properties of Sa Huynh glass ornaments. *International Workshop On Heritage Values of Ly Son - Sa Huynh Geopark*, *Viet Nam*.

Ngo-Ho, A.-K., Vo, V.-P., Pham, V.-T., Le, D.-P., Nguyen, T.-C., Doan, N.-K., & Tran, T.-T. (2019c). Artificial intelligence system for automatic identification of antique Oc Eo Glass jewelry. *National Conference on Oc Eo Runes. An Giang, Viet Nam*.

Ngo-Ho, A.-K., Vo, V.-P., Pham, V.-T., Le, D.-P., Nguyen, T.-C., Doan, N.-K., & Tran, T.-T. (2022). Automatic identification of Oc Eo Glass jewelry. *Journal of Vietnam Gemology* (accepted).

Ngo-Ho, A.-K., Bui, H.-B., & Pham, V.-T. (2023a). Continuous learning for automatic identification of Oc Eo antique glass jewelry. *The 15th IEEE International Conference on Knowledge and Systems Engineering* (KSE 2023). (accepted).

Ngo-Ho, A.-K., Bui, H.-B., & Nguyen, A.-D. (2023b). Identification between the Dong Son and Sa Huynh Antique Glass Jewelry Using the Evolving Learning. *Vietnamese National Conference on Information & Communication Technologies* (VNICT @ 2023). (accepted).

Ngo-Ho, A.-K., Pham, V.-T., Nguyen, T.-C., & Le, D.-P. (2020). Evidence of New Artifacts for The Existence of Glass-Making Furnaces in Óc Eo – Culture Area. *Journal of Vietnam Archeaology*, *Volumn 2 (2020)*.

Powers, D. M. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*.

Rehren, T., & Freestone, I. C. (2015). Ancient glass: from kaleidoscope to crystal ball. *Journal of Archaeological Science*, *56*, 233-241.

Sayre, E. V., & Smith, R. W. (1961). Compositional categories of ancient glass. *Science*, *133*(3467), 1824-1826.

Tait, H. (1991). *Five thousand years of glass*. University of Pennsylvania Press.

Tran, B. V., Vo, N. D., & Nguyen, K. (2022). An object detection method for aerial hazy images. *Can Tho University Journal of Science*, *14*(1), 91-98.

Truong, Q.-B., Dao, T.-X., & Vo, V.-P. (2014). New features set for signature recognition using neural network. *Vietnam Conference on Mechatronics (VCM'14)*, 2014.

Wedepohl, K. H., & Baumann, A. (2000). The use of marine molluskan shells for Roman glass and local raw glass production in the Eifel area (Western Germany). *Naturwissenschaften*, *87*, 129-132.

Widmer, G., & Kubat, M. (1996). Learning in the presence of concept drift and hidden contexts. *Machine learning*, *23*, 69-101.