Journal of Science and Development Economics
Nam Can Tho University

Website: jsde.nctu.edu.vn

Hardware design for jewelry inspection devices of Oc Eo and Sa Huynh cultures' glass

Vo Khuong Duy[1], Nguyen Anh Duy[1*], Ngo Ho Anh Khoi[2], Ngo Ho Anh Khoa[3]

[1]Adhightech Co., Ltd., Vietnam

[2]Faculty of Information Technology, Nam Can Tho University, Can Tho City, Vietnam

[3]Quantmetry Technology Co., Ltd., France

[*]Corresponding author: Nguyen Anh Duy (email: nguyenanhduy@adhigtechn.com)

Received: 10/1/2025

Revised: 3/2/2025

Accepted: 20/2/2025

Keywords: *archaeology, ancient glass classification, machine learning, scanning electron microscopy (SEM)*

Từ khóa: *kỹ thuật quét điện tử, khảo cổ, máy học, máy học tăng tiến SEM, phân loại thuỷ tinh cổ*

ABSTRACT

This study presents the RAS-OESHG device, which utilizes advanced machine learning algorithms to automatically differentiate between glass bead jewelry from the Oc Eo and Sa Huynh cultures through SEM gemological analysis. This device is specifically designed for experts, archaeologists, and individuals in the field of archaeology. By providing high accuracy in distinguishing between the two types of jewelry, the RAS-OESHG device is a valuable and innovative tool for the archaeology industry in Vietnam. It not only enhances the precision and efficiency of studying the Oc Eo and Sa Huynh cultures, but also promotes the use of cutting-edge technology in archaeology, opening up new avenues for research. The device is currently in the process of being patented.

TÓM TẮT

Nghiên cứu này giới thiệu thiết bị RAS-OESHG phân biệt tự động trang sức thủy tinh Oc Eo và Sa Huynh dựa trên phân tích ngọc học SEM. Thiết bị RAS-OESHG sử dụng thuật toán học máy tiên tiến để phân tích hình ảnh và dữ liệu SEM của trang sức, giúp phân biệt hai loại trang sức với độ chính xác cao. Thiết bị được cung cấp cho các chuyên gia, nhà khảo cổ và những người hoạt động trong ngành khảo cổ. Thiết bị RAS-OESHG là một công cụ mới mẻ và hiệu quả cho ngành khảo cổ học Việt Nam, giúp nâng cao độ chính xác và hiệu quả trong nghiên cứu văn hóa Oc Eo và Sa Huynh. Thiết bị cũng thúc đẩy ứng dụng công nghệ tiên tiến vào lĩnh vực khảo cổ học, mở ra hướng nghiên cứu mới đầy tiềm năng, thiết bị đang trong quá *trình xin đăng ký bằng sáng chế.*

1. INTRODUCTION

During the commercialization of ancient glass artifacts from the Dong Son, Sa Huynh, and Oc Eo cultures, these items have been widely distributed across various regions in Vietnam and globally. However, in the private antiquities collecting field, classifying and identifying these ancient glass jewelry pieces often face numerous challenges, particularly common misconceptions and misunderstandings. Specifically, in the case of glass jewelry, they are frequently referred to by different names or mistakenly identified as products from various cultures such as Dong Son, Sa Huynh, and Oc Eo.

In this study, we address the issue of distinguishing between Oc Eo and Sa Huynh glass jewelry using gemological methods, combined with artificial intelligence devices. The aim is to develop a useful tool to automatically differentiate between Oc Eo and Sa Huynh glass, as well as to progress toward classifying other ancient glass types in the region, including modern glass products (ancient replicas, new items, etc.). Our research is presented through four sections. First, we provide an overview of the history of research in this field. Next, we describe in detail the process, equipment, methods used, and the theoretical foundation for building a database of Oc Eo - Sa Huynh ancient glass. We then summarize the experimental procedure of the artificial intelligence device and evaluate the results obtained. Finally, we conclude with the findings and results of the study. Additionally, to enhance the efficiency of the classification and identification process, we also integrate a hardware system into our research, improving on-site usability as well as the security of the artifact information.

In the field of ancient glass research, gemological methods have become an essential tool since the 1990s, initiated by a series of studies by Henderson (1991) [1] on ancient Roman glass in Britain. The differentiation between ancient glass types through gemological methods is often based on structural characteristics and compositional content. There are two main groups of ancient glass studied, based on the glass production process: lead-glass and natron-glass. While the lead-glass group, also known as ancient Eastern glass, originated from the Mesopotamians and became Han glass, the natron-glass group, also known as ancient Roman glass (Klinkenberg, 2004) [17], was invented by the Romans and popularized in cultures adjacent to the Mediterranean (Tait, 1991) [2]. However, in Asia, research on this type of glass is rare, with a few studies from Thailand, China, and some isolated French studies in Vietnam. Research typically focuses on the natron-glass group, especially in Europe, and has developed a comprehensive research system over the past 30 years.

In recent times, the development of artificial intelligence has opened new opportunities to improve the classification and evaluation of ancient glass types through the integration of evolutionary learning methods. This is particularly important in archaeology, where unique challenges like data scarcity and conceptual variability exist. Evolutionary learning methods can provide a more flexible and adaptive approach to handle the continuous changes in archaeological data. We will compare and evaluate the results of these methods and apply Klinkenberg's ideas to classical methods to create new evolutionary methods. In this study, we propose using evolutionary learning methods to enhance the classification system of ancient glass, using archaeological data and integrating the solution into a unified hardware device. This will allow us to control information and optimize the performance of the ancient glass classification system in the archaeological environment.

2. RESEARCH METHODS

2.1 Building a database of ancient Vietnamese glass

Specimens from the Oc Eo and Sa Huỳnh periods have been used in building a database of ancient Vietnamese glass materials, displayed at various national exhibitions and approved by the National Appraisal Council. These specimens were collected by Mr. Nguyen Trong Co and Mr. Ngo Ho Anh Khoi and are important artifacts belonging to the Sa Huynh and Oc Eo cultures. The SEM-EDS technique has been detailed in numerous previous studies (Khoi et al., 2019,2015,2023) [3],[4],[5]. It is basically described as follows:

2.2 Backscattered electrons

Based on the research concept of Jones et al.'s systematic classification of ancient plants (Jones & Bickler, 2019) [11]. A scanning electron microscope (SEM) is combined with EDS (Model: Quanta 450; Manufacturer: FEI-USA). This electron microscope can generate high-resolution images of the sample's surface by using a narrow beam of electrons that scans the surface. The imaging process is carried out by detecting and analyzing the radiation emitted from the interaction of the electron beam with the sample surface. Backscattered electrons are electrons in the initial beam that interact with the sample surface and are reflected back, typically with high energy. This scattering is highly dependent on the chemical composition of the sample surface, making backscattered electron images useful for chemical characterization (EDS). Additionally, backscattered electrons can also be used to capture diffraction images from the backscattered electrons, aiding in crystal structure analysis (electron polarization mode).

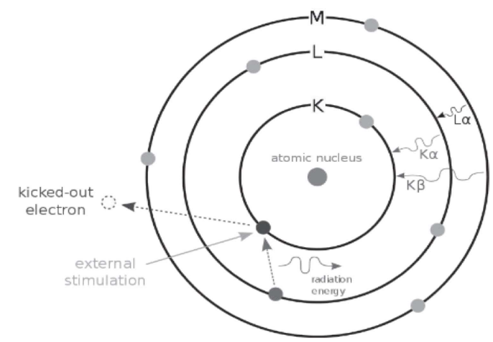

Figure 1. Illustrates the technique described in terms of electron interactions

2.3 Energy dispersive X-ray spectroscopy - EDXS

Energy Dispersive X-ray Spectroscopy (EDXS) or Wavelength Dispersive X-ray Spectroscopy (WDXS) is a technique used to analyze the chemical composition of solid materials by recording the X-ray spectra emitted from the solid material as a result of its interaction with radiation, primarily a high-energy electron beam in electron microscopes.

The sample is analyzed in a vacuum chamber environment, with high or low vacuum pressure applied depending on the type of sample (e.g., low vacuum pressure is applied for biological samples). High-resolution SEM imaging of the sample surface is conducted using an electron beam emitted from the electron gun (which may be thermionic or field emission...), which is then accelerated. The emitted electrons are accelerated and focused into a narrow electron beam through a magnetic lens system, then scanned across the sample surface using electrostatic scanning coils. The SEM resolution is determined by the size of the focused electron beam and depends on the interaction between the surface material of the sample and the electrons.

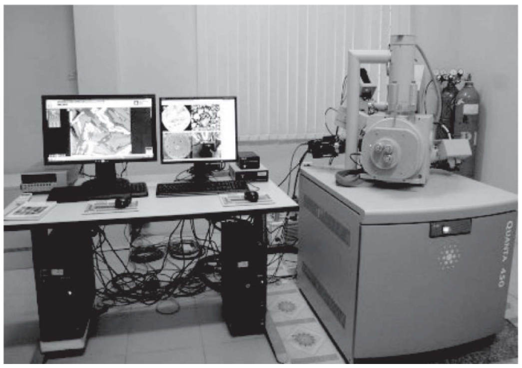

Figure 2. Shows the experimental apparatus used for gemological SEM-EDS analysis

2.4 Description of the process

The sample analyzed by SEM is solid, and its size must fit the sample chamber (smaller than 3 cm^3). Depending on the characteristics of the sample, it will be processed to ensure the best SEM image quality. For non-conductive samples, they are coated with Carbon, Platinum, or Gold to prevent surface charging. The sample is mounted on an alloy sample holder using specialized carbon adhesive tape and then placed in the sample chamber.

The location and number of analysis treatments are important in making judgments, and because each analysis shot at a location is costly, it is necessary to analyze only the essential locations to minimize redundant data. Therefore, each sample is analyzed five times at different locations on the specimen to achieve diversity and high reliability. Semi-quantitative elemental composition is determined by energy dispersive X-ray spectroscopy, where the electron beam scans the sample surface and interacts with the electron clouds of the elements. The effective reflection intensity and other values are then determined to calibrate the elemental composition.

2.5 Analysis parameters

Semi-quantitative elemental composition: The electron beam scans the sample surface and interacts with the electron clouds of the elements (M, L, K electron layers,...). The reflected components include the energy dispersive X-ray spectrum, which is detected to determine the elemental composition. The effective reflection intensity (Net Int) represents the reflection level of the electron beam, the Kratio value is the density of reflected electrons, and R is the resolution in microns. Once the density of reflected electrons is determined, the elemental composition is calibrated by comparing the spectrum of the standard sample with that of the analyzed sample. The Z, R, A, F values are calibration values, with: Z being atomic calibration when comparing the standard sample to the analyzed sample; A: Absorption calibration (adjustment for X-rays that are released but remain within the sample); F: Fluorescence calibration (adjustment for the amount of X-rays released with higher energy source levels) [1].

Table 1. SEM characteristics table

Element[1]	Weight %[2]	Atomic %[3]	Net Int.[4]	Error %[5]	Kratio[6]	Z[7]	R[8]	A[9]	F[10]
[1] Name of the element based on its electron sublayer	[2] Weight (%) in the sample	[3] Atomic (%) in the sample	[4] Effective intensity detected in the sample	[5] Error percentage	[6] Density of reflected electrons in the sample	[7] Atomic calibration	[8] Resolution in microns	[9] Absorption correction	[10] Fluorescence calibration

All results were based on balanced accuracy, which is a suitable metric for evaluating the performance of binary classifiers, especially in cases of class imbalance. Balanced accuracy considers the true negative rate and true positive rate, overcoming the limitation of regular accuracy calculations in imbalanced datasets. By considering both precision (PRE) and recall (REC), the balanced accuracy formula provides a more accurate and optimal ratio, particularly for datasets with significant imbalance, where synthetic data generation methods are not feasible or limited, as is often the case with archaeological data. The formula for balanced accuracy (BA) is as follows: B*A* = ½ (PRE+ REC).

3. RESULTS AND DISCUSSION

In this study, we conducted experiments on nine different classical machine learning methods, combined with Klinkenberg's idea to create evolving learning methods, to evaluate the effectiveness of the optimal window size when using an archaeological dataset. The experiment was performed on a total of 9 machine learning algorithms as follows:

AdaBoost Algorithm (Zhu et al., 2009) [7]: A boosting ensemble estimator that adjusts the classifier on the original dataset and further adjusts copies to focus on difficult cases. Gaussian Naïve Bayes Algorithm (Lewis, 1998a) [6]: A variant of Naïve Bayes following a Gaussian normal distribution and supporting continuous data. Decision Tree Algorithm (Freund & Schapire, 1995) [8]: A tree-like structure that uses decision rules to split data based on attribute values. K-Neighbors Algorithm (Goldberger et al., 2005) [10]: A method that finds a predetermined number of nearest training samples based on distance to a new point and predicts the label from these neighbors. Extra Tree Algorithm (Geurts et al., 2006) [15]: An ensemble estimator based on making random decisions on subsets of the dataset to improve prediction accuracy. Bagging Algorithm (Breiman, 1996) [12]: A meta-ensemble estimator that creates base classifiers on random subsets of the dataset and combines their predictions. MLP Algorithm (Lazarescu et al., 2003) [14]: Based on a neural network to perform classification, optimizing the log loss function. Random Forest Algorithm (Breiman, 2001) [13]: A meta-ensemble estimator that creates multiple decision tree classifiers on different subsets of the dataset and averages their predictions to improve accuracy. Bernoulli Naive

Bayes Algorithm (Lewis, 1998b) [18]: A variant of Naive Bayes that works well with binary-transformed data. These machine learning methods were combined with the "sliding windows" approach and incorporated Klinkenberg's idea to transform these machine learning methods into evolving learning methods, suitable for the specific problems of archaeological data.

3.1 Experimental methods on the database

The algorithms used in this experiment were implemented using version 0.24.2 of the scikit-learn library, with the sliding windows technique added using Python. The data was processed in a streaming manner using small sliding windows that considered the most recent samples with size n. The study started with n = 1, corresponding to classical online learning, then gradually increased n > 1, representing batch learning. For each method, only the best results of n were selected to compare with other machine learning methods. The archaeological dataset comprised a total of 108 samples, of which 25 belonged to Oc Eo samples, and the remaining samples belonged to Sa Huỳnh. Each sample was characterized by 40 features, representing 8 chemical elements with 5 indicators for each element.

The dataset was split into a training set (70%) and a test set (30%) with random cross-validation. The data positions in the training and test sets were randomly assigned for each experiment, repeated 10 times (10 times separation of train/test set randomly). Additionally, the training dataset was shuffled randomly 10 times to create different streaming learning scenarios, reflecting the model's adaptability to evolving data. This means that each learning method, with each parameter variant, underwent 100 random experiments before obtaining the final comparison data. The results were averaged to ensure the reliability and generalizability of the experiments.

3.2 Comparison of the algorithm experiments

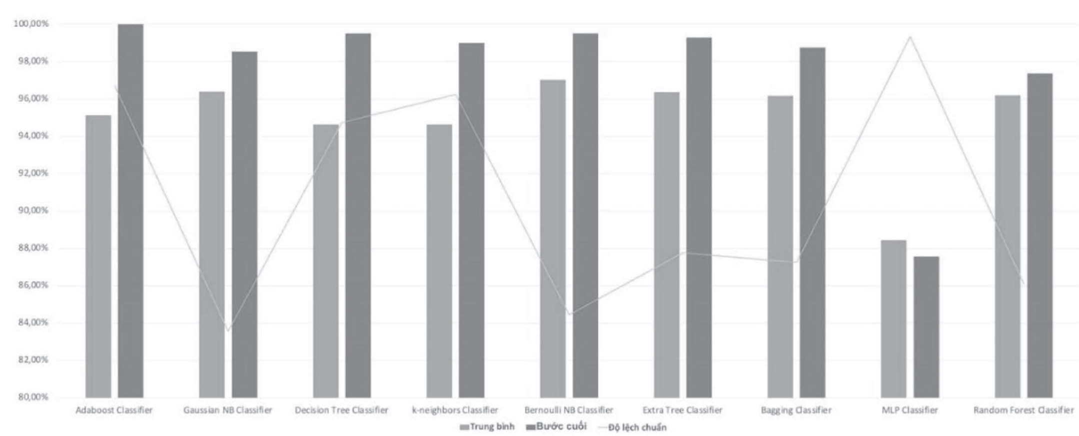

Figure 3. Correlation chart of results among 9 algorithms

When examining the average balanced accuracy (BA) results, we observe that all methods achieved high performance, with most algorithms reaching over 90%. However, the ranking of the algorithms is as follows: Bernoulli NB Classifier (97,03%), Gaussian NB Classifier (96,38%), Extra Tree Classifier (96,36%), Random Forest Classifier (96,19%), Bagging Classifier (96,18%), Adaboost Classifier (95,13%), k-neighbors Classifier (94,64%), Decision Tree Classifier (94,62%), MLP Classifier (88,44%). It is clear that the Naive Bayes learning group (Bernoulli and Gaussian NB) achieved the highest results, followed by the Ensemble learning group (Extra Tree Classifier, Random Forest Classifier, Bagging Classifier, Adaboost Classifier), and finally the other algorithm group (k-neighbors Classifier, Decision Tree Classifier, MLP Classifier). Both Naive Bayes learning methods (Bernoulli and Gaussian NB) demonstrated low standard deviation, while the other two groups had higher deviations, with the MLP Classifier exceeding 5%.

The lower performance of the MLP Classifier aligns with theoretical expectations, as this method is sensitive to limited data. Accurate adjustments only occur with a large amount of data, and with fewer data points, the accuracy per class decreases, leading to significant variations in deviation. Ensemble learning methods typically yield good results, and this problem is no exception. The performance gap between this group and the Naive Bayes learning group (Bernoulli and Gaussian NB) is minimal (mostly below 1%). The effectiveness of the Naive Bayes learning group is based on probability theory, indicating that the separation of data between the Sa Huỳnh and Oc Eo classes is distinct and highly probabilistic. This suggests a clear separation of these two classes in the data space.

These results also show that although the data is limited, it still reflects relatively evenly in the conceptual space of the classes. If there were significant imbalance, the standard deviation within the Ensemble learning group would be much higher compared to that of the Naive Bayes learning group. Through this approach, the high effectiveness of the k-neighbors Classifier and Decision Tree Classifier is mainly due to the clear division in the conceptual space. Both methods performed similarly well, and this can be easily explained by the distinct separation in the conceptual space of both classes (in Table 2).

Table 2. Results of average Balanced Accuracy (BA), final step BA, and standard deviation

Algorithms	BA average	BA final step	Standard deviation
Bernoulli NB Classifier	97.03%	99.50%	1.21%
Gaussian NB Classifier	96.38%	98.54%	0.97%
Extra Tree Classifier	96.36%	99.29%	2.12%
Random Forest Classifier	96.19%	97.37%	1.67%
Bagging Classifier	96.18%	98.75%	1.98%
Adaboost Classifier	95.13%	100%	4.56%
k-neighbors Classifier	94.64%	99.00%	4.43%
Decision Tree Classifier	94.62%	99.50%	4.02%
MLP Classifier	88.44%	87.56%	5.28%

However, there is a small discrepancy between the average balanced accuracy and the balanced accuracy at the final step. At the final step, the ranking slightly changed: most methods achieved relatively uniform results, with no significant differences (most above 98%), except for the relatively poor performance of the MLP Classifier (~87%). This indicates that the final results of the algorithms were all successful due to the relatively easy-to-separate dataset. The only issue with the dataset is its small size (which cannot be remedied), affecting the MLP Classifier's results.

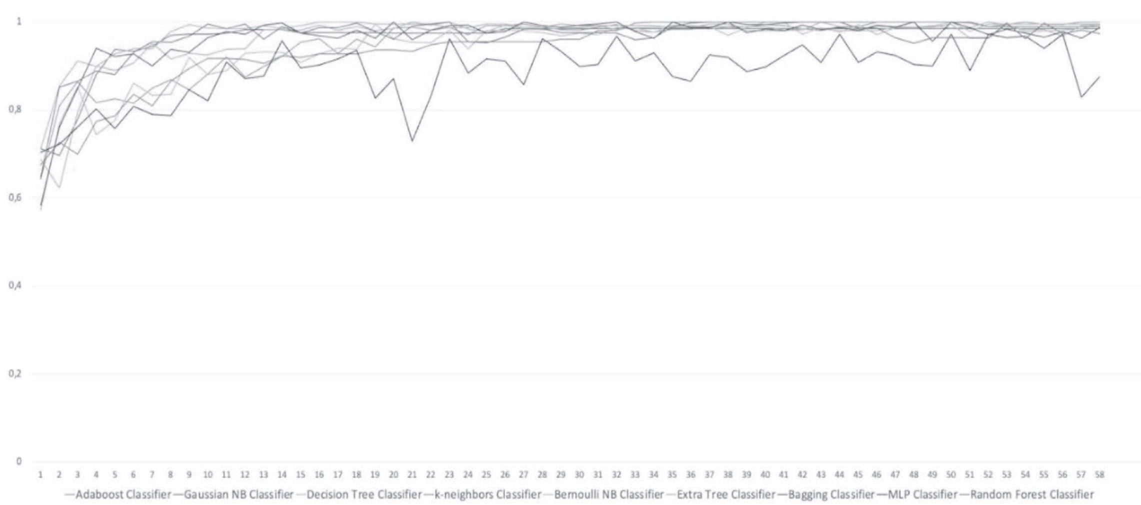

Figure 4. Chart showing the learning progress of 9 algorithms

From the perspective of tracking the learning progress of the algorithms (see Figure 4), it is easy to observe the instability of the MLP Classifier with relatively large fluctuations. The other algorithms show good results in the final steps with no significant differences. The main distinction between the algorithms (except for the MLP Classifier) lies in the early phase (before step 20), when the algorithms are trying to find the best model through data accumulation and selection. Most algorithms begin to stabilize around step 20-25, and stabilize further between steps 25-30. The large differences in the average balanced accuracy (BA) are primarily impacted by this initial unstable phase as the algorithms search for the best model. During this phase, two small groups can be easily seen: one group stabilizes early (the Ensemble learning algorithms and Naive Bayes learning algorithms), reaching stability by step 5 and gradually improving performance; the other group stabilizes later (Decision Tree Classifier, k-neighbors Classifier, and MLP Classifier), with a relatively slow increase in performance, only stabilizing by steps 15-20. This explains the discrepancy between the average BA results and the BA results at the final step.

The final outcome of this problem showed that the Bernoulli NB Classifier performed the best, both in terms of high balanced accuracy and low standard deviation. This algorithm was implemented into the official device for practical use in this study. However, the research team also noted that the Extra Tree Classifier showed promising results, especially if a larger dataset becomes available in the future.

3.3 Hardware system design

The RAS-OESHG device was designed with portability in mind, featuring a compact size that is easy to carry. Starting with the interface, this device is designed to provide an easy and convenient user experience. With its compact size, users can easily take the device anywhere without requiring much space. It includes 4 USB ports, offering flexibility in connectivity and data transmission, allowing for multiple input and output devices to be connected simultaneously. This optimizes workflow and enhances system performance, reflecting the emphasis on convenience and efficiency in the design of RAS-OESHG.

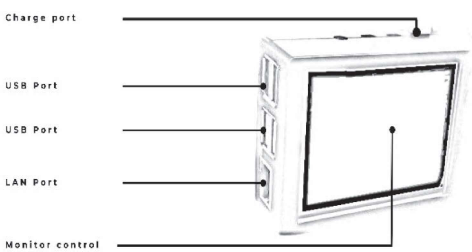

Figure 5. Design of the RAS-OESHG device interface

Looking inside the device, it is equipped with a powerful processing circuit that ensures good processing capabilities and stable performance. It uses a control board with 2GB of RAM and a Broadcom BCM2835 SoC processor, providing strong multitasking capabilities. This SoC includes an integrated CPU along with RAM, a microSD card slot for storage expansion, Wi-Fi and Bluetooth for flexible wireless connectivity, and 4 USB 2.0 ports to connect to peripheral devices.

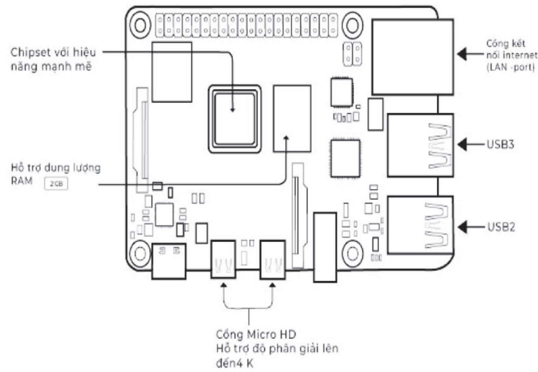

Figure 6. Internal structure diagram of the RAS-OESHG system

Specifications:

- CPU: Quad-core Cortex-A72 (64-bit) running at 1.5GHz.
- GPU: Supports H264 (1080p60 decode, 1080p30 encode), OpenGL ES 3.0 graphics, H.265 (4kp60 decode).
- RAM: Supports up to 4GB.
- Operating voltage: 5V with a minimum current of 3A.
- GPIO: 28 I/O pins.
- LAN: Yes.
- PoE: Supported.
- WIFI: Yes.
- Bluetooth: Version 5.0.
- HDMI: 2 HDMI ports supporting 4k display (mini-HDMI).
- Power supply: Can use either a DC power jack or mini USB-C port.
- Expansion connectivity: 40 pins (supports SPI, I2C, LCD, UART, PWM, SDIO).
- USB: 2 USB 2.0 ports and 2 USB 3.0 ports.
- Camera: Supported via CSI interface.
- Display: XPT2046 Touch Controller.
- Operating temperature: From 0 to 50 degrees Celsius.

Additionally, the device is connected to an interactive touchscreen circuit, with a resolution of 480x320px, controlled by the XPT2046 Touch Controller. This screen not only provides clear display but also allows interaction, enhancing user experience and system interactivity. Overall, with the combination of these components, the RAS-OESHG device ensures not only strong performance but also offers a flexible and convenient user experience

Table 3. Processing speed and variance analysis

Time	% accuracy	Processing speed (s)	Average	Standard deviation	Variance	Result
1	96.88	6.91	51.895	44.98	2023.65	Oc Eo
2	97.3	5.71	51.505	45.795	2097.18	Sa Huynh
3	90.88	6.026	48.453	42.427	1800.05	Oc Eo
4	94.22	5.441	49.8305	44.389	1.970	Oc Eo
5	94.22	5.457	49.838	44.381	1969.71	Oc Eo
6	74.54	5.999	40.269	34.27	1174.46	Sa Huynh
7	74.64	5.501	40.07	34.569	1195.05	Sa Huynh
8	93.94	6.409	50.174	43.765	1915.41	Sa Huynh
9	94.22	6.727	50.47	43.746	1913.75	Oc Eo
10	96.52	5.997	51.258	45.261	2048.6	Sa Huynh

Through multiple measurements and tests, the average time for the device to process and provide results is about 3-6 seconds. This is a groundbreaking time, fully capable of meeting practical demands.

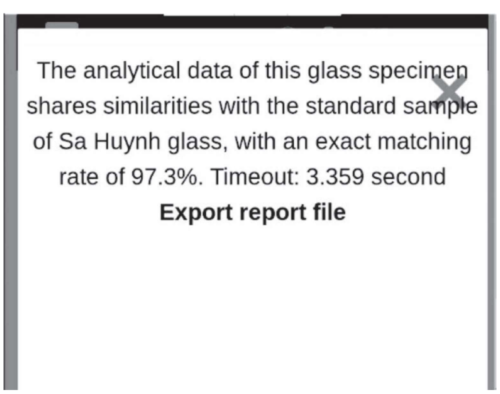

Figure 7. Average processing speed time

The analysis results will be exported via USB and displayed as a radar chart with detailed analyses.

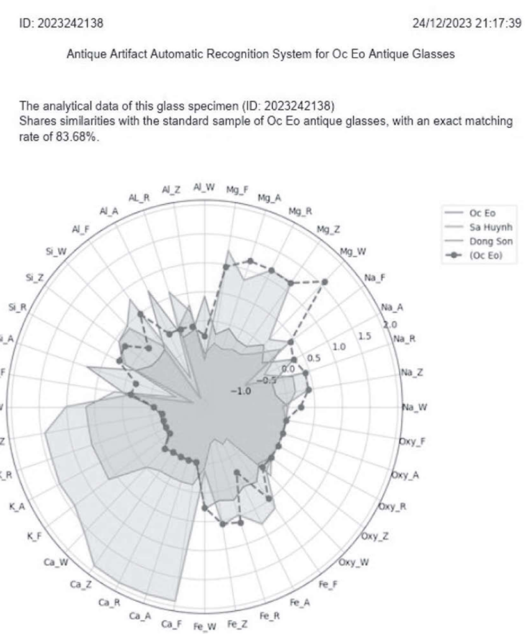

Figure 8. Report on automatic glass examination results

The above results stem from a process of refinement and innovation aimed at promoting automation in archaeological research. In archaeology, artifacts initially classified into a specific group belonging to a particular culture may later be reclassified based on additional data and new discoveries. The classification of objects or groups into specific categories is based on multidisciplinary factors. This complexity makes classification research challenging, as previously conducted studies may become outdated and lose relevance over time. As a result, these circumstances present challenges for data science in archaeology, particularly concerning the phenomenon of "concept drift" in data science (Widmer & Kubat, 1996) [9], which is a crucial factor driving the adoption of evolutionary learning methods in archaeological datasets (Klinkenberg, 2004; Wedepohl & Baumann, 2000) [16],[17].

4. CONCLUSION

This research was conducted in the specific context of archaeology, where the scarcity and irregular flow of data create significant challenges in building classification devices. This data scarcity arises from three main factors: the difficulty of extracting data from artifacts in real-world contexts, the high costs associated with conducting excavations and extracting features from these objects, and the irregular nature of data collection over time. This paper introduces an artificial intelligence system that enables the automatic recognition of Oc Eo and Sa Huynh glass jewelry through SEM gemological parameters. The system represents an effort to apply advanced technology in artifact evaluation and recognition, a technological step forward compared to traditional methods, keeping pace with recent significant developments in archaeological centers worldwide. The integration of artificial intelligence technology into archaeological activities is an inevitable trend in the future of archaeology. The results of the research have been integrated into a system—the Oc Eo Sa Huynh Glass Jewelry Automatic Recognition System—which is currently available for free to archaeological experts on the system's website. The RAS-OCSHG device (Oc Eo Sa Huynh Glass Jewelry Automatic Recognition) is currently offered for registered use by archaeological experts and researchers. Although the system performs well in recognition tasks, it still requires further support to expand and improve the dataset in terms of volume and methods to enhance the results. The RAS-OCSHG device is currently in the process of being patented.

REFERENCES

[1] Henderson, J. (1991). Chemical characterisation of Roman glass vessels, enamels and tesserae, in Materials Issues in Art and Archaeology II, P. B. Vandiver, J. R. Druzik, and G. S. Wheeler, Eds. Pittsburgh: *Materials Research Society.*

[2] Tait, H. (1991). *Five thousand years of glass.* Philadelphia: University of Pennsylvania Press.

[3] Khoi, N. H. A., Pham, V. P., Pham, V. T., Le, D. P., Nguyen, T. C., Doan, N. K., & Tran, T. T. (2019b). Artificial intelligence system for automatic identification of antique Oc Eo Glass jewelry, in *National Conference on Oc Eo Runes: Ba The, Nen Chua - Excavation, Research, Conservation and Promotion of Values, An Giang, Vietnam.*

[4] Khoi, N. H. A. (2015). *Méthodes de classifications dynamiques et incrémentales: application àla numérisation cognitive d'images de documents* (Doctoral dissertation). Université de Tours.

[5] Khoi, N. H. A., Bui, H. B, & Pham, V. T. (2023). Automatic identification of Dong Son antique glass artifacts using evolving learning, in *The International Conference on Intelligent Systems and Data Science (ISDS 2023),* accepted for publication.

[6] Lewis, D. D. (1998a). Naive (Bayes) at forty: The independence assumption in information retrieval, in Machine Learning: ECML-98, *Lecture Notes in Computer Science*, vol. 1398.

[7] Zhu, H., Zou, S., Rosset, T., & Hastie, T. (2009). *Multi-class AdaBoost*.

[8] Freund, Y. & Schapire, R. (1995). *A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting.*

[9] Widmer, G., & Kubat, M. (1996). Learning in the presence of concept drift and hidden contexts. *Machine Learning*, vol. 23, no.1.

[10] Goldberger, J., Roweis, S., Hinton, G., & Salakhutdinov, R. (2005). Neighbourhood Components Analysis, in *Advances in Neural Information Processing Systems*, vol. 17.

[11] Jones, B., & Bickler, S. H. (2019). Multi-Scalar and Semi-Automatic Approaches to Detect Archaeological Features in NZ Using Airborne LiDAR Data. *Archaeology in New Zealand*.

[12] Breiman, L. (1996). Bagging predictors. *Machine Learning*, vol. 24, no.2.

[13] Breiman, L. (2001). Random Forests. *Machine Learning*, vol. 45, no.1.

[14] Lazarescu, M. Venkatesh, S., & Bui, H. (2003). Using multiple windows to track concept drift. *Intelligent Data Analysis*, vol. 8, no.1.

[15] Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, vol. 63, no.1.

[16] Klinkenberg, R. (2004) Learning drifting concepts: Example selection example weighting, Intelligent Data Analysis. *Special Issue on Incremental Learning Systems Capable of Dealing With Concept Drift*, vol. 8, no.3.

[17] 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*, vol.87.

[18] Lewis, D. D. (1998b). Naive (Bayes) at forty: The independence assumption in information retrieval. In European conference on machine learning. *Berlin, Heidelberg: Springer Berlin Heidelberg.*