ACTA IMEKO
ISSN: 2221-870X
September 2026, Volume 15, Number 3, 1 – 7

Comparative evaluation of CNN and ViT architectures for the petrographic classification of Levantine ceramic micrographs

Sara Capriotti[1], Donatella Genovese[2], Alessio Devoto[2], Kamal Badreshany[3], Dennis Braekmans[4], Silvano Mignardi[1], Simone Scardapane[5], Laura Medeghini[1]

[1] *Department of Earth Sciences, Sapienza University of Rome, P.le Aldo Moro 5, 00185 Rome, Italy*
[2] *Department of Computer, Control and Management Engineering, Sapienza University of Rome, Via Ariosto 25, 00185 Rome, Italy*
[3] *Department of Archaeology, University of Durham, Stockton Road, DH13LE, Durham, United Kingdom*
[4] *Department of Archaeological Sciences, Faculty of Archaeology, Leiden University, Einsteinweg 2, 2333 CC Leiden, The Netherlands*
[5] *Department of Information Engineering, Electronics and Telecommunications, Sapienza University of Rome, Via Eudossiana 18, 00184 Rome, Italy*

ABSTRACT
This study explores the application of artificial intelligence (AI) techniques for the classification of Levantine ceramic thin sections based on their petrographic *fabrics*. The Levant is a key archaeological region, particularly with regard to the transition from the Late Chalcolithic to the Early Bronze Age, marked by urbanization, craft specialization, and expanding interregional exchange networks. Ceramic production therefore provides crucial evidence for understanding technological practices and cultural interactions. We apply two deep learning architectures, convolutional neural networks (CNNs) and vision transformers (ViTs), to a large and heterogeneous dataset of ceramic thin-section micrographs dating from the Uruk period through the Bronze and Iron Ages. The dataset includes samples from 14 archaeological sites across the Levant, representing 20 different petrographic *fabrics*. To evaluate model robustness and generalization, we performed additional experiments by excluding samples from selected sites and using them as an unseen test domain, simulating a real-world classification scenario under domain shift and class imbalance. Model performance was evaluated using macro-, micro-, and weighted-F1 scores. To improve model transparency, visual explainable AI approaches, such as Guided Grad-CAM and transformer attention maps, are applied to identify the most informative features driving classification decisions. The results demonstrate high classification performance, reaching an accuracy of 88.44 % for ResNet18 and 90.52 % for DeiT-small. Visual explanations further indicate that the models rely on mineralogical and textural features consistent with petrographic criteria used by human experts, supporting the archaeological interpretability of the proposed approach.

Section: RESEARCH PAPER

Keywords: Levantine pottery; ceramic petrography; neural networks; explainable AI

Citation: S. Capriotti, D. Genovese, A. Devoto, K. Badreshany, D. Braekmans, S. Mignardi, S. Scardapane, L. Medeghini, Comparative evaluation of CNN and ViT architectures for the petrographic classification of Levantine ceramic micrographs, Acta IMEKO, vol. 15 (2026) no. 3, pp. 1–7. DOI: 10.21014/actaimeko.v15i3.2325

Section Editor: Fabio Leccese, Università Degli Studi Roma Tre, Rome, Italy

Received: Februar 17, 2026; **In Final Form:** September 15, 2026; **Published:** September, 2026.

Funding: No funding specified.

Corresponding Author: Sara Capriotti, e-mail: sara.capriotti@uniroma1.it

1. INTRODUCTION

This study applies artificial intelligence (AI) methodologies to the grouping and classification of Levantine archaeological ceramics based on their petrographic *fabrics*. The Bronze Age represents a critical phase in the ancient history of the Levant, characterized by increasing complexity in settlement patterns, production practices, and trade networks. These developments are particularly evident in pottery production, where ceramics constitute key indicators for understanding technological practices and patterns of cultural interaction across time and space [1]. The Levantine region played a central role in these processes, acting as a hub of ceramic innovation and craft specialization. From the Early Bronze Age onward, the consolidation of urban life, the expansion of interregional exchange, and the emergence of political centralization transformed production systems. These changes reflect a path toward centralized social organization, together with the simplification of material culture and the restructuring of ceramic manufacture [2], [3]. Despite growing scholarly interest in the region, the study of ancient exchange systems and the reconstruction of trade networks remain major challenges.

In particular, the lack of sufficient comparative large ceramic assemblages limits the ability to study long-distance interactions and trade connections [4].

Artificial intelligence is a branch of computer science that is increasingly being applied in archaeometry and archaeological research [5]–[9]. In ceramic studies, machine learning (ML) and deep learning (DL) models have been used to classify pottery based on typological, decorative, and morphological characteristics, as well as for provenance analyses [10]–[13]. Recently, convolutional neural network architectures have been applied to the classification of ceramic thin sections according to their petrographic *fabrics* [14], [15]. Yang et al. (2025) proposed an enhanced ResNet50 model integrating Squeeze-and-Excitation blocks and a self-attention module to improve ceramic classification [16]. Despite this step toward transformer-based architectures, there is still little to no direct evidence of the application of more recent vision transformers, which are instead discussed as a potential future direction [17]. This limitation may be related to the fact that ViTs typically achieve their best performance when trained on very large datasets [18], a condition often difficult to meet in archaeological research due to the scarcity and fragmentation of material evidence.

Building on this progress and on our previous research [19], which successfully implemented deep learning models for the classification of ceramic petrographic *fabrics*, the present study extends this approach by incorporating a significantly expanded dataset and evaluation procedure. Although the same architectures of CNNs and ViTs are adopted, this study introduces new ceramic samples from additional archaeological sites and includes new petrographic classes. This broader dataset allows for a more comprehensive and representative analysis of the variability in Levantine pottery production. In addition, this study includes dedicated experiments aimed at evaluating the ability of the implemented models to generalize to new data domains. To this end, selected ceramic samples were excluded from the original dataset and used as a separate, unseen test domain to simulate a realistic classification scenario. Finally, the study integrates visual explainability techniques to improve transparency in model behavior. In fact, although deep learning methods are known for their high predictive accuracy, their internal decision-making mechanisms are often difficult to understand, leading them to be described as "black boxes" [20], [21]. To address this limitation, explainable artificial intelligence (XAI) approaches were employed to investigate model interpretability and identify the most informative features for the classification process. Specifically, Guided Grad-CAM [22] was used to generate high-resolution, class-discriminative visualizations for the ResNet18 model, while transformer attention maps were applied to highlight the features attended by ViT models [23], [24], supporting a more transparent and informed application of AI methodologies in archaeological research.

2. MATERIALS AND METHODS

The dataset was built by acquiring 4,740 images from 593 ceramic thin sections, selected from a broad range of archaeological sites across the Levantine region. As shown in Figure 1, these include Bethlehem, Jericho, Tell el-Far'ah (North), Al Jib, Tell Balata and, Jerusalem (West Bank); Khirbat Iskandar, Khirbat al-Batrawy, and Deir Alla' (Jordan); Tell Qasile (Israel), Ebla, Tell Nebi Mend, Jebel Aruda, and Tell Hadidi (Syria). The geographic and chronological diversity of these sites contributes to the representativeness and variability of the dataset (see Table 1 for more specific information about the sites).

Figure 1. Geographical map of the archaeological sites.

Table 1. Summary of the archaeological sites and chronological periods, with the corresponding number of selected samples.

Archaeological sites	Age	N. of Samples
Tell el-Far'ah (North)	Early Bronze	58
Khirbat al-Batrawy	Early Bronze	42
Khirbat Iskandar	Early Bronze	31
Bethlehem	Bronze and Iron Ages	24
Ebla	Early and Middle Bronze	54
Jericho	Early Bronze	10
Tell Balata	Iron Age	6
Deir Alla'	Bronze and Iron Ages	190
Al Jib	Iron Age	6
Tell Hadidi	Early Bronze	13
Tell Qasile	Iron Age	6
Tell Nebi Mend	Middle Bronze	79
Jerusalem	Bronze and Iron Ages	48
Jebel Aruda	Late Uruk	26

To reflect realistic acquisition conditions, the images were intentionally captured using different microscopes and camera systems. We used a ZEISS D-7082 Oberkochen petrographic microscope coupled with an AXIOCAM 208 ZEISS camera, as well as two Leica petrographic microscopes (Leica 2700P coupled with a Leica MC170 HD camera and Leitz Labrolux 12 pol with a Leica EC3). This choice introduces natural variability and noise in terms of resolution, colour calibration, and lighting. While adding this heterogeneity helps reduce the risk of overfitting, it also improves the realism and generalizability of the trained models, making them better suited for real-world archaeological applications where imaging conditions may vary. For each section, multiple images were taken under both plane-polarized light (PPL) and cross-polarized light (XPL), and different magnifications were used. All images were grouped into 20 petrographic *fabrics* according to the Whitbread approach, based on the analysis of matrix, inclusions, and voids [25], [26]. Unlike the previous study [19], which used a smaller dataset and applied only a bin-

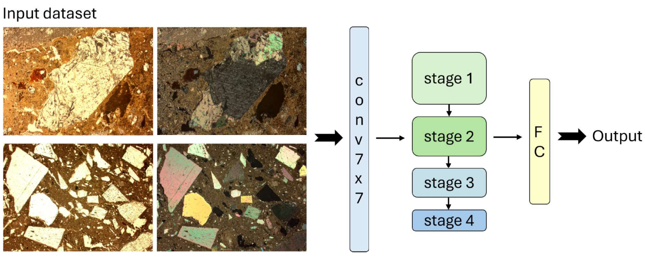

Figure 2. Scheme of the ResNet18 architecture.

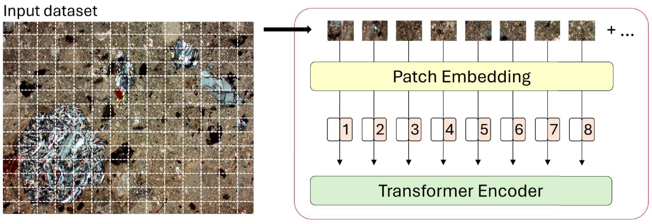

Figure 3. Scheme of the Deit-small architecture.

ary split (training/test with a ratio of 80:20), the present work benefits from a significantly expanded dataset and adopts a more robust three-way split: 70% for training, 15% for validation, and 15% for testing. This configuration not only supports a more rigorous model evaluation of the models, but also allows hyperparameter selection and fine-tuning through the validation set. Two types of deep learning models were selected for this study: a CNN using the ResNet18 [27] architecture and the DeiT-small vision transformer [28]. CNNs process images through multiple convolutional layers that detect increasingly complex features. Specifically, the ResNet18 architecture consists of 18 layers organized into an initial convolutional block, followed by four residual stages, each composed of multiple residual blocks with skip connections, including convolutional layers, batch normalization, and ReLU activations. The network ends with a global average pooling layer and a fully connected layer for classification (Figure 2). In contrast, vision transformers work with a different approach. Input images are divided into fixed-size patches (196 patches of 16×16 pixels in our case, Figure 3), which are linearly embedded with positional information, and then processed through a series of transformer encoder blocks based on multi-head self-attention mechanisms [29]. Our ViT implementation employs 12 transformer encoder blocks, each consisting of 6 attention heads and a hidden dimension of 384.

Both models were trained to classify ceramic samples into their respective petrographic *fabrics* by minimizing a cross-entropy loss function. Transfer learning was applied by fine-tuning models pre-trained on ImageNet, a large dataset containing more than one million images across 1,000 classes [30], retaining the backbone weights and retraining only the final classification layers. Training was performed for 25 epochs using the AdamW optimizer. For ResNet18, a learning rate of $1 \cdot 10^{-4}$, a weight decay of $1 \cdot 10^{-4}$, and a batch size of 32 were used, while for ViT, a learning rate of $1 \cdot 10^{-4}$, a weight decay of $3 \cdot 10^{-4}$, and a batch size of 64 were applied. To improve model generalization and artificially increase the training data size, data augmentation techniques, including random axis flipping, color jittering, random cropping and Gaussian blur, were applied consistently during the training phase to reduce overfitting [31]. The performance of the two deep learning models was compared in terms of both classification accuracy and generalization capabilities. Their predictive performance was evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. In addition, a confusion matrix was generated for each model to visualize the performance and identify misclassification patterns. To further assess generalization ability, we conducted additional experiments in which data from specific archaeological sites (Jericho, Al Jib, Tell Hadidi, and several Bronze Age samples from Jerusalem) were entirely excluded from the training, validation, and initial test sets, and used solely as independent and out-of-domain new test set. The excluded sites and samples were selected based on data availability and class representation, ensuring that their removal did not negatively affect the balance or size of the training dataset. This strategy allowed us to exclude both entire sites and selected samples from a larger site while preserving sufficient training data. This approach simulates realistic scenarios in which the models are required to analyze previously unseen data from new archaeological contexts.

3. RESULTS AND DISCUSSION

This work extends the work of Capriotti et al. (2025) [19]. While the earlier study was based on a dataset of 1,424 images across 10 classes, the present work introduces a significantly larger dataset (4,740 images) and 20 petrographic classes, resulting in a more challenging classification task. Using this improved dataset, we revisited the same architectures (ResNet18 and DeiT-small), and trained them on the updated ceramic collection.

The results were promising, with both models achieving strong performance across all evaluation metrics. Figures 4 and 5 show the confusion matrices for ResNet18 and DeiT-small, respectively, illustrating the distribution of correct and incorrect predictions relative to the true labels.

Accuracy, precision, recall, and F1-score values on the full in-domain dataset are reported in Table 2, confirming the robustness of both models. Specifically, ResNet18 achieved an accuracy of 88.44 %, while DeiT-small shows an accuracy of 90.52 %. A direct comparison with the results reported in [19] highlights different behaviors of the two models. ResNet18 shows a slight decrease in performance, exemplified by a reduction in accuracy from 92.11 % to 88.44 %, consistent with the increased task complexity introduced by the expanded dataset. In contrast, DeiT-small exhibits an improvement in performance, with accuracy increasing from

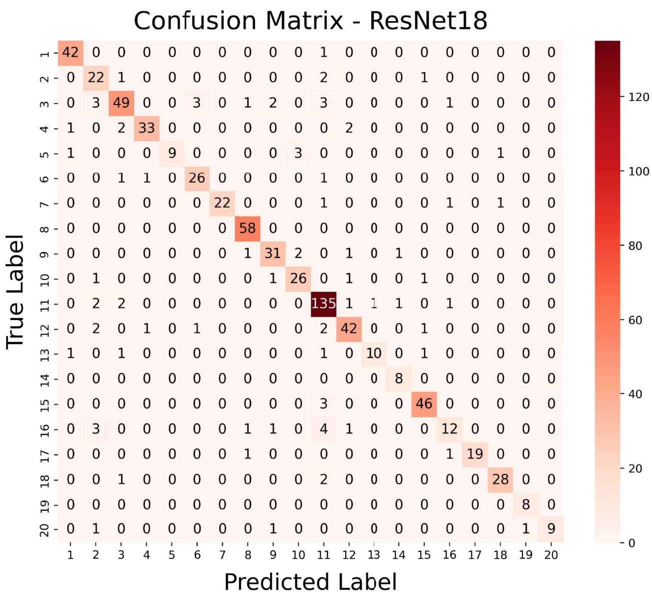

Figure 4. Confusion matrix showing the classification performance of ResNet18.

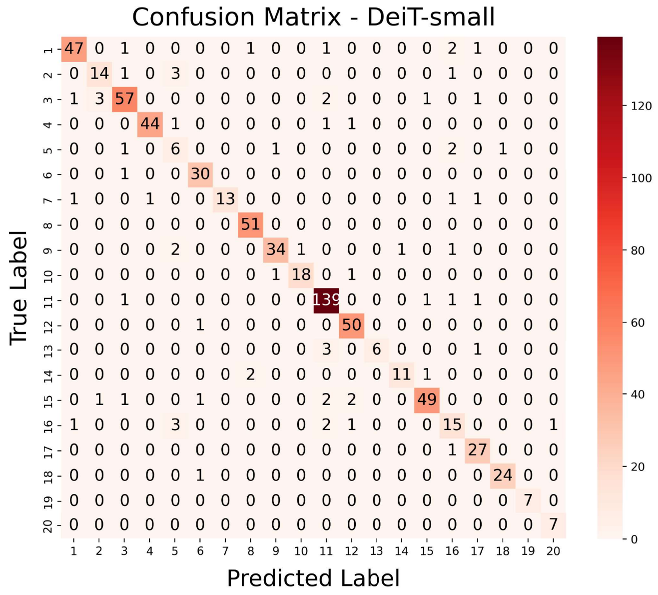

Figure 5. Confusion matrix showing the classification performance of DeiT-small.

Table 2. In-domain classification performance of ResNet18 and DeiT-small on the full dataset and on the reduced in-domain dataset.

	ResNet18		DeiT-small	
Metric	Full	Reduced	Full	Reduced
Accuracy	88.44 %	86.12 %	90.52 %	87.72 %
Precision	89.16 %	83.14 %	88.72 %	88.09 %
Recall	85.93 %	83.46 %	86.38 %	85.31 %
F1-score	86.92 %	82.53 %	87.05 %	86.03 %

88.34 % to 90.52 %. This result suggests that ViT benefits from the increased dataset size, even in the presence of a larger number of classes. Similar trends are observed across all evaluation metrics for both architectures.

To evaluate the generalization ability of ResNet18 and DeiT-small, all samples from the Jericho, Al Jib, and Tell Hadidi sites, as well as a subset of samples from Jerusalem, were excluded from the original dataset and used to construct a separate out-of-domain test set. It is important to note that this test set comprises fewer classes than the original collection, making the evaluation more specific to those reduced classes. Both models were first trained, validated, and tested on the remaining in-domain dataset. As also reported in Table 2, training on the reduced data resulted in a performance comparable to that obtained on the full set.

The models were subsequently evaluated on the unseen data to simulate a "real-world" scenario. The results obtained are reported in Table 3. ResNet18 achieved a macro-F1 score of 75.25 %, with micro- and weighted-F1 scores of 82.87 % and 83.33 %, respectively, and an accuracy 95 % confidence interval of 74.24 %–83.72 %. DeiT-small exhibited a very similar performance, reaching a macro-F1 score of 73.92 %, a micro-F1 of 81.03 %, a weighted-F1 of 82.53 %, and an accuracy 95 % confidence interval of 72.35 %–82.58 %. For clarity, we primarily report macro-, micro-, and weighted-F1 scores, as these metrics are particularly informative for measuring how well the classification models handle minority classes. Macro-F1 assigns equal importance to all classes and is therefore especially suitable for assessing performance on underrepresented classes. In contrast, micro- and weighted-F1 reflect the overall classification performance by accounting for class frequency. As a result, these metrics provide a more reliable evaluation than accuracy alone, particularly in scenarios involving imbalanced data and reduced label spaces, such as the out-of-domain setting considered here

Table 3. Macro-F1, micro-F1, weighted-F1, and accuracy (95 % confidence interval) for ResNet18 and DeiT-small evaluated on the out-of-domain dataset.

Model Metrics	ResNet18	DeiT-small
Macro-F1	75.25 %	73.92 %
Micro-F1	82.87 %	81.03 %
Weighted-F1	83.33 %	82.53 %
Accuracy 95 % CI	74.24–83.72 %	72.35–82.58 %

[32], [33].

The slightly larger performance drop observed for the DeiT-small model in the out-of-domain evaluation, compared to its in-domain accuracy of approximately 90 %, may partly be explained by architectural differences between convolutional neural networks and vision transformers. Although DeiT-small is pre-trained on ImageNet and benefits from strong generic visual representations, transformer-based architectures inherently lack the strong spatial inductive biases that characterize CNNs. While ViTs are effective at modeling long-range dependencies across image patches, they tend to rely more heavily on the availability of different and representative training data during fine-tuning [34]. As a result, they may exhibit increased sensitivity to distribution shifts when adapted to smaller, class-imbalanced, and domain-specific datasets. In contrast, convolutional architectures, such as ResNet18, leverage locality and translation-equivariant feature extraction, which promotes the learning of stable low-level and mid-level representations that are less sensitive to changes in data distribution [35]. This effect becomes particularly evident in our case, where the unseen dataset is smaller, exhibits class imbalance, and only partially overlaps with the original label space of the training data. Under these conditions, DeiT-small shows slightly reduced robustness, while ResNet18 maintains more consistent generalization performance.

In our previous study [19], we explored in-domain visual explainability methods by applying Guided Grad-CAM and transformer attention maps to ResNet18 and DeiT-small, respectively, which highlighted petrographically meaningful characteristics of the micrographs. Building on these results, in the present work, we extend these explainability analyses to the out-of-domain test set, comprising samples from Jericho, Al Jib, Tell Hadidi, and Jerusalem. Figure 6 shows the Guided Grad-CAM visualizations for samples J-EB-11, GE-2, AJ-8, and HT-8, belonging to class 1 (*Calcite-rich fabric*), class 4 (*Dolomite-rich A fabric*), class 12 (*Microfossil-rich fabric*), and class 15 (*Volcanic fabric*), respectively. Figure 7 presents the DeiT-small attention maps for samples J-EB-12, GE-6, GE-50, and HT-12, corresponding to class 1 (*Calcite-rich fabric*), class 8 (*Basalt-rich A fabric*), class 11 (*Carbonate-quartz-rich fabric*), and again class 15 (*Volcanic fabric*).

The visual explainability maps obtained on the out-of-domain test set are consistent with those observed in the in-domain setting reported in [19]. In both cases, the most informative regions correspond to the dominant inclusions and mineralogical components that characterize each petrographic class, while the matrix and voids appear to be significantly less influential in the models' decision-making process. Guided Grad-CAM provides pixel-level explanations by emphasizing local structures such as edges and textural variations associated with specific mineral phases and inclusions, including the angular shape of calcite crystals (class 1), the characteristic rhombohedral shape of dolomite (class 4), the rounded morphology of microfossils, mainly identified as foraminifera (class 12), and the volcanic lithic fragments of class 15, such as the biotite crystal present in the micrograph in Figure 6. Similarly, the attention maps produced by DeiT-small identify diagnostic features by highlighting broader, patch-based regions

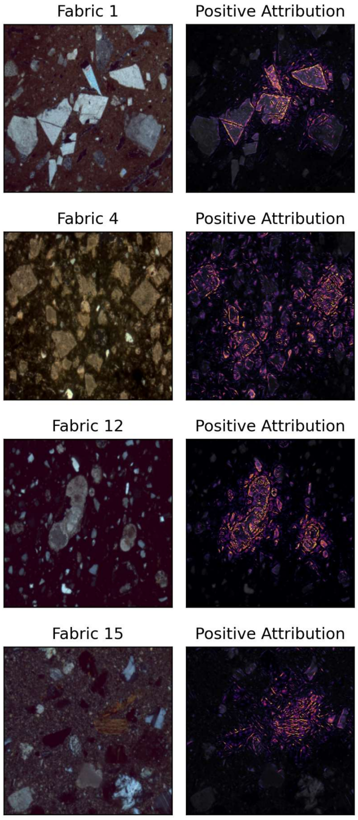

Figure 6. Guided Grad-CAM results for samples J-EB-11, GE-2, AJ-8, and HT-8 belonging to class 1 (*Calcite-rich fabric*), class 4 (*Dolomite-rich A fabric*), class 12 (*Microfossils-rich fabric*), and class 15 (*Volcanic fabric*).

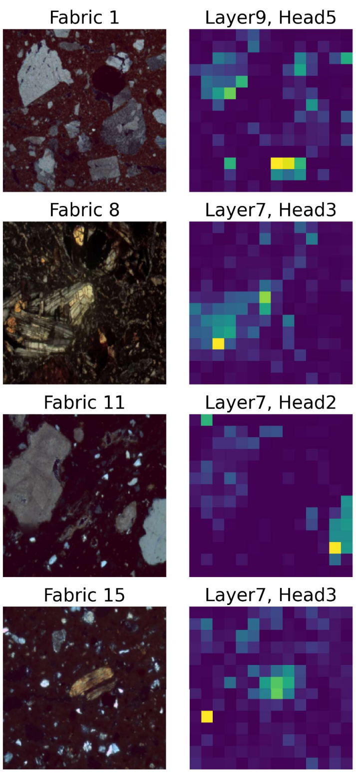

Figure 7. Attention maps results for samples J-EB-12, GE-6, GE-50, and HT-12 belonging to class 1 (*Calcite-rich fabric*), class 8 (*Basalt-rich A fabric*), class 11 (*Carbonate-Quartz-rich fabric*), and class 15 (*Volcanic fabric*).

of interest, capturing more global spatial patterns of petrographic features across the micrographs.

The heatmaps indicate that the models mainly rely on visually distinctive inclusions and mineral phases for classification. In contrast, features such as the matrix and voids, which play a central role in the petrographic definition of *fabrics*, are less influential for the models. This suggests that the learned decision strategy does not fully replicate traditional petrographic criteria, but rather prioritizes features that are more discriminative from a visual and statistical perspective.

This study demonstrates the practical applicability of deep learning as a supportive tool for the rapid screening of large ceramic assemblages, providing an objective and reproducible framework for ceramic classification.

4. CONCLUSIONS

Both ResNet18 and DeiT-small demonstrated comparable performance in the out-of-domain evaluation. However, while DeiT-small performed better than ResNet18 in the in-domain setting, it exhibited a more pronounced relative drop in performance when exposed to unseen data. This suggests that ResNet18, despite its lower performance, may exhibit greater robustness across domains, particularly in the presence of class imbalance and limited dataset size. These results indicate that both models can generalize to previously unseen ceramic samples, although targeted strategies aimed at improving predictions on less represented classes could further improve performance. More broadly, the findings highlight the importance of training on datasets with balanced class distributions and sufficient variability, as well as the need for larger datasets containing an adequate number of samples per class to achieve more robust and stable learning. The results also underscore the potential of deep learning approaches for the automated classification of ceramic thin sections. In this context, the integration of explainable artificial intelligence techniques provides insights into model behavior, improves transparency, and supports the real-world applicability of these methods in archaeometric research. Future work will focus on further optimizing the dataset as a fundamental step, exploring advanced architectures such as dynamic vision transformers, and extending the experiments to additional out-of-domain datasets. Particular attention will also be given to out-of-distribution detection strategies to identify samples that do not belong to any known class, allowing the recognition of potential outliers.

AUTHORS' CONTRIBUTION

Sara Capriotti: Conceptualization, Data curation, Formal analysis, Investigation, Software, Visualization, Writing – original draft. Donatella Genovese: Formal analysis, Methodology, Software, Writing – original draft. Alessio Devoto: Formal analysis, Methodology, Software, Writing – original draft. Kamal Badreshany: Resources, Writing – review & editing. Dennis Braekmans: Resources, Writing – review & editing. Silvano Mignardi: Supervision, Writing – review & editing. Simone Scardapane: Conceptualization, Software, Supervision, Writing – review & editing. Laura Medeghini: Conceptualization, Resources, Supervision, Writing – original draft.

5. ACKNOWLEDGMENTS

The authors acknowledge The Levantine Ceramic Project group for providing access to their samples from the Tell Qasile archaeological site. DG acknowledges support from the PNRR MUR project PE0000013-FAIR.

The pottery from Tell el-Far'ah (North) was collected during nine excavation seasons carried out at the site by Roland de Vaux between 1946 and 1960. The Italian-Palestinian expeditions to Bethlehem (2014–2015) and to Tell es-Sultan (1997–2022) were conducted by the joint team of Sapienza University of Rome and the Department of Archaeology and Cultural Heritage of the Palestinian Ministry of Tourism and Antiquities (MoTA-DACH). The Tell es-Sultan project has also been supported by the Italian Ministry of Foreign Affairs and International Cooperation (MAECI). Both missions were carried out under the guidance of Lorenzo Nigro. Additional thin section materials belong to the collections of Leiden University and form part of a legacy research collection. Some of these materials belong to Professor Henk Franken's collection and derive from the excavations of Kathleen Kenyon at Tell es-Sultan (Jericho) (1955–1958) and Jerusalem (1961–1967).

REFERENCES

[1] R. Greenberg, The Archaeology of the Bronze Age Levant, Cambridge University Press, 2019. DOI: 10.1017/9781316275993

[2] M. Steiner, A. E. Killebrew, The Oxford Handbook of the Archaeology of the Levant: c. 8000–332 BCE, Oxford University Press, 2014.

[3] R. Greenberg, Traveling in (world) time: transformation, commoditization, and the beginnings of urbanism in the Southern Levant, Interweaving worlds : systemic interactions in Eurasia, 7th to 1st millennia BC, Oxbow Books, 2011, pp. 231–242.

[4] K. Badreshany, G. Philip, Ceramic studies and petrographic analysis in Levantine archaeology, the limitations of current approaches, Levant, vol. 52, 2020, no. 1-2, pp. 5–14. DOI: 10.1080/00758914.2020.1786270

[5] S. H. Bickler, Machine learning arrives in archaeology. Advances in Archaeological Practice, vol. 9, 2021, no. 2, pp. 186–191. DOI: 10.1017/aap.2021.6

[6] A. Guyot, M. Lennon, T. Lorho, L. Hubert-Moy, Combined detection and segmentation of archaeological structures from lidar data using a deep learning approach, Journal of Computer Applications in Archaeology, vol. 4, 2021, no. 1. DOI: 10.5334/jcaa.64

[7] Ø. D. Trier, D. C. Cowley, A. U. Waldeland, Using Deep Neural Networks on Airborne Laser Scanning Data: results from a Case Study of Semi-Automatic Mapping of Archaeological Topography on Arran, Scotland, Archaeological Prospection, vol. 26, 2019, no. 2, pp. 165–175. DOI: 10.1002/arp.1731

[8] M. Troiano, E. Nobile, F. Mangini, M. Mastrogiuseppe, C. C. Barbaro, F. Frezza, A comparative analysis of the bayesian regularization and levenberg–marquardt training algorithms in neural networks for small datasets: a metrics prediction of Neolithic laminar artefacts, Information, vol. 15, 2024, no. 5. DOI: 10.3390/info15050270

[9] R. Zhang, Y. Cheng, J. Huang, Y. Zhang, H. Yan, Unsupervised weathering identification of grottoes sandstone via statistical features of acoustic emission signals and graph neural network, Heritage Science, vol. 12, 2024, no. 1, p. 323. DOI: 10.1186/s40494-024-01432-w

[10] P. Navarro, C. Cintas, M. Lucena, J. M. Fuertes, C. Delrieux, M. Molinos, Learning feature representation of Iberian ceramics with automatic classification models, Journal of Cultural Heritage, vol. 48, 2021, pp. 65–73. DOI: 10.1016/j.culher.2021.01.003

[11] G. Ruschioni, D. Malchiodi, A. M. Zanaboni, L. Bonizzoni, Supervised learning algorithms as a tool for archaeology: classification of ceramic samples described by chemical element concentrations, Journal of Archaeological Science, vol. 49, 2023. DOI: 10.1016/j.jasrep.2023.103995

[12] A. Anglisano, L. Casas, I. Queralt, R. D. Febo, Supervised machine learning algorithms to predict provenance of archaeological pottery fragments, Sustainability, vol. 14, 2022. DOI: 10.3390/su141811214

[13] G. Barone, P. Mazzoleni, G. V. Spagnolo, S. Raneri, Artificial neural network for the provenance study of archaeological ceramics using clay sediment database, Journal of Cultural Heritage, vol. 38, 2019, pp. 147–157. DOI: 10.1016/j.culher.2019.02.004

[14] M. Lyons, Ceramic fabric classification of petrographic thin sections with deep learning, Journal of Computer Applications in Archaeology, vol. 4, 2021, no. 1, pp. 188–201. DOI: 10.5334/jcaa.75

[15] M. Lyons, F. Fecher, M. Reindel, From lidar to deep learning: a case study of computer-assisted approaches to the archaeology of Guadalupe and northeast Honduras, it – Information Technology, vol. 64, 2022, no. 6, pp. 233–246. DOI: 10.1515/itit-2022-0004

[16] L. Yang, W. Zhou, W. Qiu, Chronological classification of Ming and Qing dynasty ceramics images based on an enhanced resnet50 model, STAR: Science & Technology of Archaeological Research, vol. 11, 2025, no. 1, p. e2498260. DOI: 10.1080/20548923.2025.2498260

[17] Z. Ling, G. Delnevo, P. Salomoni, S. Mirri, Findings on machine learning for identification of archaeological ceramics: a systematic literature review, IEEE Access, vol. 12, 2024, pp. 100 167–100 185. DOI: 10.1109/ACCESS.2024.3429623

[18] M. Raghu, T. Unterthiner, S. Kornblith, C. Zhang, A. Dosovitskiy, Do vision transformers see like convolutional neural networks? Advances in Neural Information Processing Systems, Curran Associates, Inc., 2021, vol. 34, pp. 12 116–12 128. Online. [Accessed: 2026-09-19]. https://proceedings.neurips.cc/paper_files/paper/2021/file/652cf38361a209088302ba2b8b7f51e0-Paper.pdf

[19] S. Capriotti, A. Devoto, S. Scardapane, S. Mignardi, L. Medeghini, Interpretable classification of Levantine ceramic thin sections via neural networks, Machine Learning: Science and Technology, vol. 6, 2025, no. 2. DOI: 10.1088/2632-2153/ade6c4

[20] D. Castelvecchi, Can we open the black box of AI? Nature, vol. 538, 2016, no. 7623, pp. 20–23. DOI: 10.1038/538020a

[21] R. Kashefi, L. Barekatain, M. Sabokrou, F. Aghaeipoor, Explainability of vision transformers: a comprehensive review and new perspectives, 2023. Online. [Accessed: 2026-09-19]. https://arxiv.org/abs/2311.06786

[22] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017. DOI: 10.1109/ICCV.2017.74

[23] J.-B. Cordonnier, A. Loukas, M. Jaggi, On the relationship between self-attention and convolutional layers, 2020. Online. [Accessed: 2026-09-19]. https://arxiv.org/abs/1911.03584

[24] T. Shiraishi, D. Miwa, T. Katsuoka, V. N. L. Duy, K. Taji, I. Takeuchi, Statistical test for attention map in vision transformer, 2024. Online. [Accessed: 2026-09-19]. https://arxiv.org/abs/2401.08169

[25] I. K. Whitbrad, The characterisation of argillaceous inclusions in ceramic thin sections. Archaeometry, vol. 28, 1986, no. 1, pp. 79–88. DOI: 10.1111/j.1475-4754.1986.tb00376.x

[26] I. K. Whitebread, Greek Transport Amphorae. A Petrological and Archaeological Study., British School of Athens., 1995.

[27] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90

[28] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An image is worth 16x16 words: transformers for image recognition at scale, 2021. Online. [Accessed: 2026-09-19].
https://openreview.net/forum?id=YicbFdNTTy

[29] S. Alijani, J. Fayyad, H. Najjaran, Vision transformers in domain adaptation and domain generalization: a study of robustness, Neural Computing and Applications, vol. 36, 2024, no. 29, pp. 17 979–18 007.
DOI: 10.1007/s00521-024-10353-5

[30] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, L. Fei-Fei, Imagenet: a large-scale hierarchical image database, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255.
DOI: 10.1109/CVPR.2009.5206848

[31] S. Yang, W. Xiao, M. Zhang, S. Guo, J. Zhao, F. Shen, Image data augmentation for deep learning: a survey, 2023. Online. [Accessed: 2026-09-19].
https://arxiv.org/abs/2204.08610

[32] S. M. A. Elrahman, A. Abraham, A review of class imbalance problem, Journal of Network and Innovative Computing, vol. 1, 2013, pp. 9–9. Online. [Accessed: 2026-09-19].
https://api.semanticscholar.org/CorpusID:10781880

[33] S. Bej, N. Davtyan, M. Wolfien, M. Nassar, O. Wolkenhauer, LoRAS: an oversampling approach for imbalanced datasets, Machine Learning, vol. 110, 2021, no. 2, pp. 279–301.
DOI: 10.1007/s10994-020-05913-4

[34] H. Zhu, B. Chen, C. Yang, Understanding why vit trains badly on small datasets: an intuitive perspective, 2023. Online. [Accessed: 2026-09-19].
https://arxiv.org/abs/2302.03751

[35] R. Geirhos, J.-H. Jacobsen, C. Michaelis, R. Zemel, W. Brendel, M. Bethge, F. A. Wichmann, Shortcut learning in deep neural networks, Nature Machine Intelligence, vol. 2, 2020, no. 11, pp. 665–673.
DOI: 10.1038/s42256-020-00257-z