

The Eurasia Proceedings of Science, Technology, Engineering and Mathematics (EPSTEM), 2026

Volume 42, Pages 229-236

ICBASET 2026: International Conference on Basic Sciences, Engineering and Technology

Deep Learning-Based Classification of Archaeological Artifacts Using Image Processing

Fatma Ozge Ozkok
Erciyes University

Abstract: The classification of archaeological artifacts according to their material types is important for archaeological documentation and analysis. This study evaluates five convolutional neural network (CNN) architectures for the classification of archaeological artifact images. A balanced dataset containing 1,000 images from four material classes, namely ceramic, glass, metal, and stone, was used. Five CNN architectures, SqueezeNet1.1, MNASNet0.5, RegNetX-200MF, FasterNet-T0, and InceptionNeXt-Tiny, were evaluated using stratified five-fold cross-validation. Model performance was assessed using accuracy, precision, recall, F1-score, and confusion matrices. The experimental results showed that InceptionNeXt-Tiny achieved the highest average accuracy of $90.70 \pm 0.76\%$ and F1-score of $90.66 \pm 0.76\%$. RegNetX-200MF and FasterNet-T0 achieved average accuracies of 87.50% and 86.20%, respectively. Confusion matrix analysis showed that classification errors varied across material classes, particularly between glass and metal and between ceramic and stone. The results demonstrate the potential of CNN architectures for automated material classification of archaeological artifacts.

Keywords: Archaeological artifacts, Material classification, Convolutional neural networks, Deep learning, Image classification.

Introduction

Archaeological artifacts provide important information about past civilisations, including details of their cultural, technological and artistic developments. They are commonly examined according to their material, shape, production technique and historical period (Kowalski et al., 1972; Rouse, 1960). Ceramics, metals, stones, and glasses are among the materials frequently encountered in archaeological collections. Classifying these objects is important for archaeological research, museum documentation and organising digital collections (Mikhailova et al., 2019). The growing accessibility of digitised museum collections has given rise to opportunities for the application of image processing and machine learning techniques (Bertacchini & Morando, 2013; Li et al., 2012). Traditional classification approaches typically necessitate manual examination and expert knowledge. However, the sheer quantity of objects in museum collections means manual classification is extremely time-consuming. Automated image classification methods can support the identification and organisation of archaeological artifacts based on their visual characteristics (Mnasri & D'Andrea, 2025).

Deep learning methods, particularly convolutional neural networks (CNNs), have been widely used in image classification tasks. CNN architectures can automatically extract visual features such as edges, textures, shapes, and spatial patterns from images (Nani et al., 2025). Different CNN architectures have been developed to improve classification performance while reducing model parameters and computational requirements. Lightweight architectures provide different approaches for image classification tasks (Chen et al., 2024; Zhou et al., 2020).

In this study, five CNN architectures, including SqueezeNet1.1, MNASNet0.5, RegNetX-200MF, FasterNet-T0, and InceptionNeXt-Tiny, were employed for material-based classification of archaeological artifacts. A dataset containing 1,000 images was prepared from the publicly available iMet Open Access dataset. The artifacts were

- This is an Open Access article distributed under the terms of the Creative Commons Attribution-Noncommercial 4.0 Unported License, permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

- Selection and peer-review under responsibility of the Organizing Committee of the Conference

© 2026 Published by ISRES Publishing: www.isres.org

divided into four classes, including ceramic, metal, stone, and glass, with 250 images in each class. Stratified five-fold cross-validation was used to evaluate the models. Classification performance was compared using Accuracy, Precision, Recall, and F1-score.

Related Works

Artificial intelligence and deep learning have been used in archaeology to identify, classify, and digitally document artifacts. Gattiglia (2025) reviewed the applications of artificial intelligence in archaeology and discussed the problems associated with archaeological data. The main issues included data quality, classification methods, and the limitations of AI-based approaches.

Gualandi et al. (2021) developed an open system for identifying archaeological pottery using neural network algorithms. Their work focused on the shape and decoration of pottery and included two approaches for identifying archaeological ceramics. Cintas et al. (2020) used deep convolutional networks for feature extraction and classification of Iberian ceramics. Their experiments were based on profile images of wheel-made pottery vessels.

Resler et al. (2021) combined an ImageNet-pretrained EfficientNetB3 model with metric learning to classify archaeological artifacts and identify archaeological communities. Artifacts from different archaeological sites and historical periods were included in their analysis. For rock art identification, Jalandoni et al. (2022) investigated machine learning methods and developed an approach for identifying photographs containing painted rock art. Tsigkas et al. (2020) addressed a related task involving the automatic detection of ancient rock carvings using image-based recognition methods.

Grove and Blinkhorn (2020) investigated the differences between Middle and Later Stone Age lithic assemblages in eastern Africa. They used artificial neural networks for classification. Their classification was based on the presence or absence of 16 technological characteristics, and more than 94% of the assemblages were correctly classified. Mnasri and D'Andrea (2025) investigated the use of object detection and deep learning for the automatic inventory of archaeological artifacts. Their work included object identification and material-based classification. They also compared their approach with transfer learning models such as SqueezeNet and GoogLeNet.

Previous studies have investigated different applications of deep learning in archaeology, including pottery classification, rock art detection, and artifact identification. In this study, we classified archaeological artifacts into four groups based on their materials: ceramic, metal, stone, and glass. Five CNN models with different convolutional structures were evaluated using the same dataset and experimental procedure. We compared their classification results using accuracy, precision, recall, F1-score, and confusion matrices.

Method

This section describes the dataset, CNN architectures, experimental setup, and performance evaluation methods used in this study.

Dataset

The dataset used in this study was obtained from the publicly available iMet Open Access dataset, which is based on The Metropolitan Museum of Art's Open Access collection. The dataset is related to the iMet Collection 2019 Challenge Dataset described by Zhang et al. (2019). Archaeological artifacts were selected from the Ancient West Asian Art, Egyptian Art, and Greek and Roman Art departments. The first seven training files were examined, containing 4,477 records from these departments. Objects with missing end dates or end dates later than 1500 CE were excluded.

The artifacts were put into four groups based on the material they were made of, including ceramic, metal, stone, and glass, using the material information provided in the metadata. The Ceramic class included things made of terracotta, pottery, ceramic and faience. The Metal class included bronze, gold, silver, copper, iron, and lead objects. The Stone class included limestone, marble, granite, sandstone, and alabaster objects. The Glass class included objects made of glass.

After filtering, 3,293 artifacts remained, including 2,273 ceramic, 468 metal, 277 stone, and 275 glass objects. From these records, 250 artifacts were randomly selected for each class using a random seed of 42. The final dataset consisted of 1,000 images, with 250 images in each class.

Table 1. Distribution of archaeological artifacts by material

Class	Eligible artifacts	Selected artifacts
Ceramic	2,273	250
Metal	468	250
Stone	277	250
Glass	275	250
Total	3,293	1,000

The distribution of the selected artifacts is presented in Table 1. Representative images from each material class are shown in Figure 1.

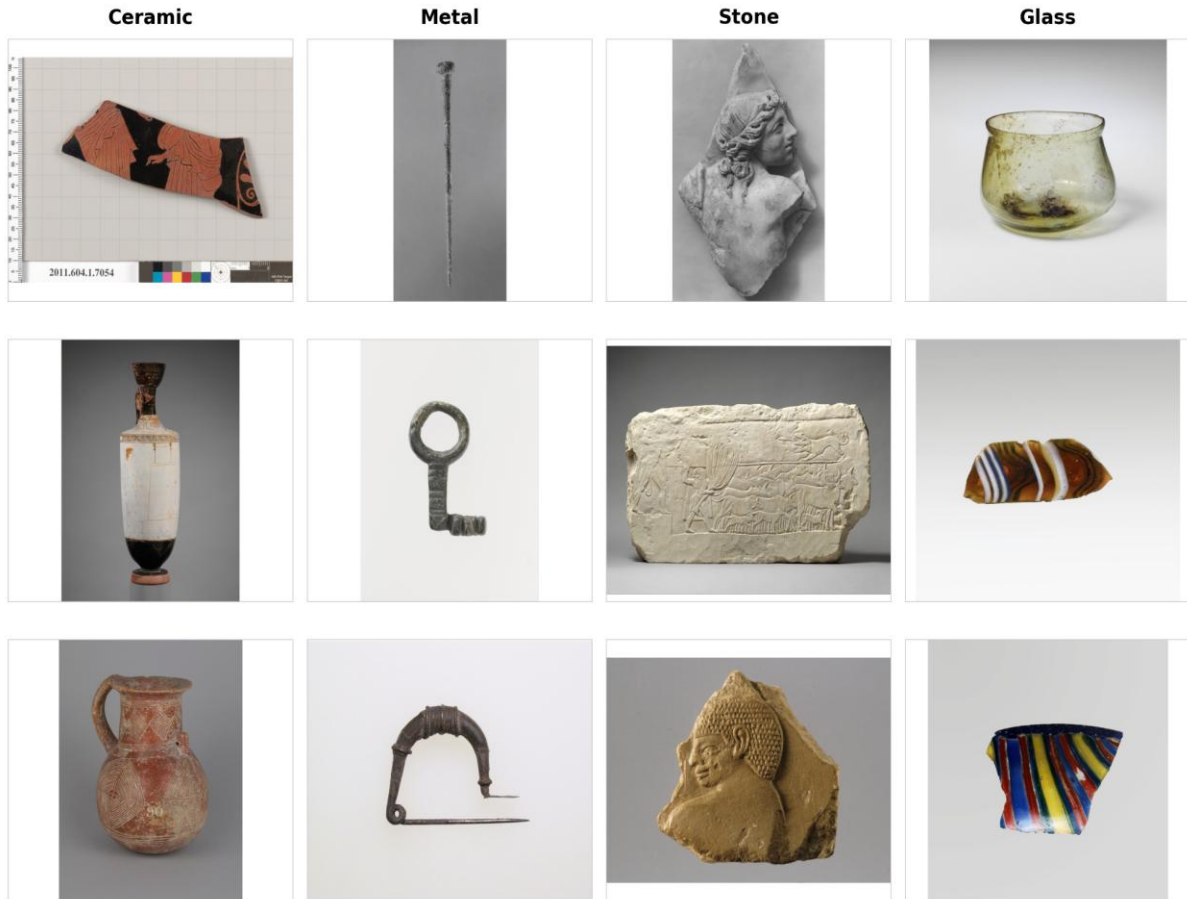


Figure 1. Sample images of archaeological artifacts from the four material classes.

CNN Architectures

Five CNN architectures were used to classify archaeological artifacts. These were SqueezeNet1.1, MNASNet0.5, RegNetX-200MF, FasterNet-T0, and InceptionNeXt-Tiny. The models have different convolutional structures, and their classification results were compared under the same experimental conditions. SqueezeNet1.1 uses Fire modules to extract image features with fewer parameters. Each Fire module contains a squeeze layer with 1×1 convolutions and an expand layer with both 1×1 and 3×3 convolutions (Iandola et al., 2016).

MNASNet0.5 was developed using neural architecture search. The model uses mobile inverted bottleneck blocks and depthwise separable convolutions. Compared with the standard MNASNet architecture, the 0.5 width multiplier reduces the number of channels (Tan et al., 2019). RegNetX-200MF uses grouped convolutions and a regular network structure in which channel widths increase across stages. Its network configuration is determined

using a parameterized design. The 200MF version was developed for a computational budget of approximately 200 million floating-point operations (Radosavovic et al., 2020).

FasterNet-T0 applies partial convolutions to only some of the input channels. The other channels are not processed by spatial convolutions and are later combined with the processed channels. This design aims to reduce unnecessary computations during feature extraction (Chen et al., 2023).

InceptionNeXt-Tiny uses depthwise convolutions together with an Inception-based structure. Different convolutional kernels are applied in separate branches to extract spatial information. This structure is designed to reduce computational complexity (Yu et al., 2024).

All models were initialized with ImageNet-pretrained weights. The original classification layers were replaced to classify the four material classes. The same dataset partitions, training settings, and evaluation procedure were used for all models.

Experimental Setup

The experiments were conducted using Python and PyTorch in Google Colab with an NVIDIA A100 GPU. Stratified five-fold cross-validation was used to evaluate the classification performance of the five CNN architectures. In each fold, 80% of the dataset was allocated to training and validation, while the remaining 20% was used for testing. The training and validation data were further divided into 80% training and 20% validation sets. Accordingly, each fold contained 640 training, 160 validation, and 200 test images. The same data partitions were used for all models. All images were resized to 224×224 pixels and normalized using ImageNet mean and standard deviation values. Random horizontal flipping, rotation, and color jittering were applied to the training images. No data augmentation was applied to the validation and test images.

All models were trained using the AdamW optimizer with an initial learning rate of 0.0001, weight decay of 0.0001, and batch size of 32. The maximum number of training epochs was set to 20. A cosine annealing learning rate scheduler was used during training. Early stopping was applied with a patience of five epochs based on validation Macro F1-score. The model weights corresponding to the highest validation Macro F1-score were selected for testing. A fixed random seed of 42 was used for data partitioning.

Performance Evaluation

The classification performance of the models was evaluated using Accuracy, Precision, Recall, and F1-score. Macro averaging was applied to Precision, Recall, and F1-score to give equal importance to the four material classes. The evaluation metrics were calculated separately for each test fold, and the results were reported as mean and standard deviation across five folds. Confusion matrices were generated using the combined test predictions from all five folds to examine the classification performance of individual material classes.

Results and Discussion

The classification performance of five CNN architectures was evaluated using stratified five-fold cross-validation. The experiments were conducted on a balanced dataset containing 1,000 archaeological artifact images from four material classes. Model performance was evaluated using accuracy, precision, recall, and F1-score. The results are presented as mean $\pm$ standard deviation across five folds. In addition, confusion matrices were generated to examine class-specific classification performance and misclassification patterns.

Table 2. Classification performance of five CNN architectures

Model	Accuracy	Precision	Recall	F1-score
SqueezeNet1.1	82.50 ± 4.02	83.13 ± 3.95	82.50 ± 4.02	82.45 ± 4.05
MNASNet0.5	73.00 ± 3.18	75.20 ± 3.52	73.00 ± 3.18	72.94 ± 3.16
RegNetX-200MF	87.50 ± 2.81	87.94 ± 2.64	87.50 ± 2.81	87.46 ± 2.81
FasterNet-T0	86.20 ± 1.92	86.52 ± 1.71	86.20 ± 1.92	86.19 ± 1.88
InceptionNeXt-Tiny	90.70 ± 0.76	91.08 ± 0.76	90.70 ± 0.76	90.66 ± 0.76

Table 2 presents the classification results of the five CNN architectures. InceptionNeXt-Tiny achieved the highest average accuracy of $90.70 \pm 0.76\%$ and F1-score of $90.66 \pm 0.76\%$. RegNetX-200MF achieved the second-highest

accuracy of $87.50 \pm 2.81\%$, followed by FasterNet-T0 with $86.20 \pm 1.92\%$. SqueezeNet1.1 achieved an average accuracy of $82.50 \pm 4.02\%$, while MNASNet0.5 showed the lowest accuracy of $73.00 \pm 3.18\%$. InceptionNeXt-Tiny also achieved the highest precision and recall values. The lower standard deviation of this model indicates more consistent classification performance across the five folds.

Figure 2 shows the confusion matrices for the five CNN architectures. Each material class is comprised of 250 images, resulting in 1,000 predictions for each model. The confusion matrices provide a detailed comparison of classification performance across the Ceramic, Glass, Metal, and Stone classes. SqueezeNet1.1 correctly classified 197 ceramic, 207 glass, 210 metal and 211 stone images. The highest level of confusion occurred between the glass and metal classes, with 34 glass images being misclassified as metal and 26 metal images being misclassified as glass. MNASNet0.5 performed relatively poorly in terms of classification, particularly for glass and metal. The model correctly classified 169 images of glass and 161 images of metal. The highest misclassification rate was observed between metal and stone, with 58 metal images being misclassified as stone. Additionally, 46 glass images were misclassified as metal. The RegNetX-200MF model correctly classified 207 ceramic images, 221 glass images, 221 metal images and 226 stone images. The model demonstrated relatively consistent performance across all four categories. The highest misclassification rate was observed between the ceramic and stone classes, with 23 ceramic images being misclassified as stone. The FasterNet-T0 model correctly classified 212 ceramic, 213 glass, 217 metal and 220 stone images. The model performed similarly across the material classes, although 25 glass images were misclassified as metal. The InceptionNeXt-Tiny model correctly classified 212 ceramic, 230 glass, 233 metal and 232 stone images. The model achieved the highest number of correct classifications for glass, metal and stone, while equalling FasterNet-T0's performance for ceramic. The highest misclassification rate

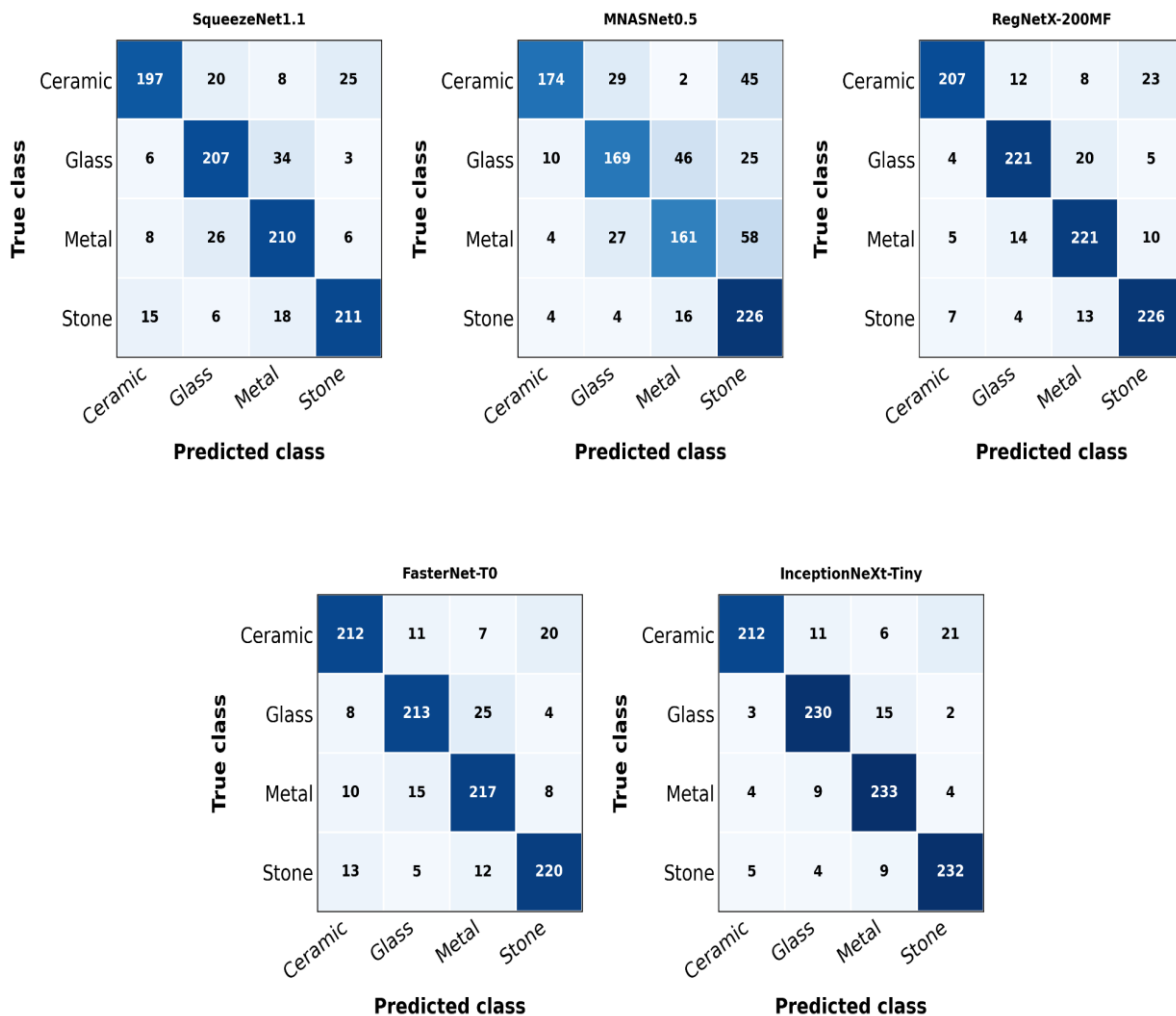


Figure 2. Confusion matrices of five CNN architectures

InceptionNeXt-Tiny achieved the highest number of correct predictions overall, correctly classifying 212 ceramic, 230 glass, 233 metal and 232 stone images. The model achieved the highest number of correct classifications for glass, metal and stone, while equalling FasterNet-T0's performance for ceramic. The highest misclassification rate

was between ceramic and stone, with 21 ceramic images being misclassified as stone. Compared with the other models, InceptionNeXt-Tiny produced fewer misclassifications, particularly for glass and metal.

Overall, InceptionNeXt-Tiny achieved the highest classification performance among the evaluated architectures. RegNetX-200MF and FasterNet-T0 also provided competitive results, while SqueezeNet1.1 and MNASNet0.5 showed lower performance. The confusion matrices revealed that classification errors were mainly observed between glass and metal and between ceramic and stone, although the error patterns differed across models. These results indicate that the evaluated CNN architectures can classify archaeological artifacts according to their material types, with differences in classification accuracy and consistency.

Conclusion

In this study, five CNN architectures were evaluated for the classification of archaeological artifacts into four material classes, namely Ceramic, Glass, Metal, and Stone. A balanced dataset containing 1,000 images was used, and the models were evaluated using stratified five-fold cross-validation. Classification performance was compared using accuracy, precision, recall, F1-score, and confusion matrices.

The experimental results showed that InceptionNeXt-Tiny achieved the highest classification performance, with an average accuracy of $90.70 \pm 0.76\%$ and F1-score of $90.66 \pm 0.76\%$. RegNetX-200MF and FasterNet-T0 also achieved competitive results, with average accuracies of 87.50% and 86.20%, respectively. SqueezeNet1.1 and MNASNet0.5 showed lower classification performance. Confusion matrix analysis revealed that classification errors were mainly observed between Glass and Metal and between Ceramic and Stone, although the misclassification patterns varied across the models.

In summary, the results suggest that CNN architectures can be employed for the automated classification of materials in archaeological artifacts. However, the study was limited to a balanced dataset containing four material classes, so the generalisation performance on independent archaeological image collections remains to be evaluated.

Recommendations

Future studies may investigate transformer-based architectures for classifying archaeological artifacts. Additionally, lightweight CNN architectures could be optimised for deployment on mobile devices, enabling automated material classification in field applications.

Real-time image analysis could also be investigated to support archaeological documentation and material identification. Further experiments using larger datasets, additional material classes and independent image collections could provide a more comprehensive evaluation of classification performance and model generalisation.

Scientific Ethics Declaration

* The author declares that the scientific ethical and legal responsibility of this article published in EPSTEM Journal belongs to the author.

Conflict of Interest

* The author declares that there is no conflict of interest.

Funding

* Financial support was received through COST Action CA22143 – European Materials Informatics Network (EuMiNe), supported by COST (European Cooperation in Science and Technology).

Acknowledgements or Notes

* This article was presented as an oral presentation at the International Conference on Basic Sciences, Engineering and Technology (www.icbaset.net) held in Istanbul/Türkiye on September 17-18, 2026.

* The author thanks the reviewers for their valuable feedback.

References

- Bertacchini, E., & Morando, F. (2013). The future of museums in the digital age: New models for access to and use of digital collections. *International Journal of Arts Management*, 15(2), 60–72.
- Chen, F., Li, S., Han, J., Ren, F., & Yang, Z. (2024). Review of lightweight deep convolutional neural networks. *Archives of Computational Methods in Engineering*, 31(4), 1915–1937.
- Chen, J., Kao, S.-H., He, H., Zhuo, W., Wen, S., Lee, C.-H., & Chan, S.-H. G. (2023). Run, don't walk: Chasing higher FLOPS for faster neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 12021–12031).
- Cintas, C., Lucena, M., Fuertes, J. M., Delrieux, C., Navarro, P., González-José, R., & Molinos, M. (2020). Automatic feature extraction and classification of Iberian ceramics based on deep convolutional networks. *Journal of Cultural Heritage*, 41, 106–112.
- Gattiglia, G. (2025). Managing artificial intelligence in archeology: An overview. *Journal of Cultural Heritage*, 71, 225–233.
- Grove, M., & Blinkhorn, J. (2020). Neural networks differentiate between Middle and Later Stone Age lithic assemblages in eastern Africa. *PLOS ONE*, 15(8), e0237528.
- Gualandi, M. L., Gattiglia, G., & Anichini, F. (2021). An open system for collection and automatic recognition of pottery through neural network algorithms. *Heritage*, 4(1), 140–159.
- Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). *SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and less than 0.5MB model size*. arXiv.
- Jalandoni, A., Zhang, Y., & Zaidi, N. A. (2022). On the use of machine learning methods in rock art research with application to automatic painted rock art identification. *Journal of Archaeological Science*, 144, 105629.
- Kowalski, B. R., Schatzki, T. F., & Stross, F. H. (1972). Classification of archaeological artifacts by applying pattern recognition to trace element data. *Analytical Chemistry*, 44(13), 2176–2180.
- Li, Y. C., Liew, A. W. C., & Su, W. P. (2012). The digital museum: Challenges and solution. In *2012 8th International Conference on Information Science and Digital Content Technology (ICIDT2012)* (Vol. 3, pp. 646–649). IEEE.
- Mikhailova, N., Mikhailova, E., & Grafeeva, N. (2019). The application of clustering techniques to group archaeological artifacts. In *World Conference on Information Systems and Technologies* (pp. 50–57). Springer International Publishing.
- Mnasri, Z., & D'Andrea, A. (2025). Automatic inventory of archaeological artifacts based on object detection and classification using deep and transfer learning. *Digital Applications in Archaeology and Cultural Heritage*, 39, e00458.
- Nani, P., Das, S., & Dey, S. (2025). Enhancing object recognition: A comprehensive analysis of CNN based deep learning models considering lighting conditions and perspectives. *Evolutionary Intelligence*, 18(4), Article 72.
- Radosavovic, I., Kosaraju, R. P., Girshick, R., He, K., & Dollár, P. (2020). Designing network design spaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10428–10436).
- Resler, A., Yeshurun, R., Natalio, F., & Giryas, R. (2021). A deep-learning model for predictive archaeology and archaeological community detection. *Humanities and Social Sciences Communications*, 8, Article 295.
- Rouse, I. (1960). The classification of artifacts in archaeology. *American Antiquity*, 25(3), 313–323.
- Tan, M., Chen, B., Pang, R., Vasudevan, V., Sandler, M., Howard, A., & Le, Q. V. (2019). MnasNet: Platform-aware neural architecture search for mobile. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2820–2828).
- Tsigkas, G., Sfikas, G., Pasialis, A., Vlachopoulos, A., & Nikou, C. (2020). Markerless detection of ancient rock carvings in the wild: Rock art in Vathy, Astypalaia. *Pattern Recognition Letters*, 135, 337–345.
- Yu, W., Zhou, P., Yan, S., & Wang, X. (2024). InceptionNeXt: When Inception meets ConvNeXt. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5672–5683).
- Zhang, C., Kaeser-Chen, C., Vesom, G., Choi, J., Kessler, M., & Belongie, S. (2019). *The iMet Collection 2019 Challenge Dataset* [Preprint]. arXiv.

Zhou, Y., Chen, S., Wang, Y., & Huan, W. (2020). Review of research on lightweight convolutional neural networks. In *2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)* (pp. 1713–1720). IEEE.

Author(s) Information

Fatma Ozge Ozkok

Erciyes University, Department of Computer Engineering,
Kayseri, Türkiye

ORCID iD: <https://orcid.org/0000-0002-1421-4670>

*Contact e-mail: fozgeozkok@erciyes.edu.tr

*Corresponding author(s) contact email

To cite this article:

Ozkok, F.O. (2026). Deep learning-based classification of archaeological artifacts using image processing. *The Eurasia Proceedings of Science, Technology, Engineering and Mathematics (EPSTEM)*, 42, 229-236.