

Intelligent Ancient Pottery Porcelain Type Features Recognition

Yuan Ma

Lecturer, Publicity and United Front Department, Zibo Polytechnic University, Zibo, 55300, China, E-mail: mayuan88662025@163.com

Project Management

Received February 10, 2026; revised March 31, 2026; accepted September 10, 2026
Available online September 20, 2026

Abstract: As an important carrier of historical culture and craft technology, accurate identification of ancient pottery and porcelain shapes is crucial for cultural relic dating, archaeological research, and cultural heritage preservation. At present, the identification of ancient ceramic types relies mainly on traditional manual methods, which are subjective, inefficient, and limited in their ability to detect subtle morphological features. To address these limitations, this study proposes a Multidimensional Data Representation (MDR) and an intelligent recognition method for ancient ceramic shape features based on Artificial Intelligence (AI). A multimodal data fusion framework, "image-3D-semantics," is constructed to achieve comprehensive and fine-grained characterization of ceramic shape features, and an improved Deep Support Vector Machine (Deep-SVM) is adopted for intelligent recognition. Experiments are conducted on a self-collected dataset of 900 ancient ceramic samples covering 7 categories (bowls, plates, bottles, cans, kettles, stoves, cups/pots), with a train-test split ratio of 7:3. Baseline comparisons include traditional manual features + SVM, single image feature + SVM, single 3D feature + SVM, and basic CNN features + SVM. On this basis, an intelligent recognition technique is developed to improve the accuracy and efficiency of ancient ceramic identification. Results show that after integrating multi-source information, the proposed model increases feature coverage to 92%, discrimination to 88%, reduces data redundancy to 12%, and improves the fit to real-world models to 90%. Overall, the model demonstrates clear advantages in comprehensiveness, differentiation, efficiency, and authenticity. The findings provide new technical means and methodological insights for ancient ceramic research, supporting its digital and intelligent development.

Keywords: Artificial intelligence (AI) technology; ancient pottery porcelain type; intelligent recognition; multidimensional data; feature representation.

Copyright © Journal of Engineering, Project, and Production Management (EPPM-Journal).
DOI 10.32738/JEPPM-2026-224

1. Introduction

As important carriers of human history and culture, ancient ceramics hold rich historical, artistic, and scientific value. The traditional method of identifying Ancient Pottery Porcelain Type (APPT) mainly relies on expert experience, which is subjective, inefficient, and requires a high level of professional expertise (Xiao et al., 2024). With the rapid development of information technology, AI is increasingly being applied to ancient ceramic research, offering new opportunities for the recognition of APPT features (Muttamara et al., 2025; Sun, 2024). In recent years, many scholars have made certain progress in the study of ancient ceramics. In response to the problem of traditional classification of ancient ceramics relying on experts, being time-consuming, subjective, and difficult to promote, Ling et al. (2025) used deep learning and transfer learning to conduct multi-attribute classification of cultural relics using four CNN models, combined with Artificial Intelligence (AI) technology to ensure model interpretability. The results showed that MobileNetV2 achieved the best performance, that transfer learning effectively improved scores, and that classification accuracy across tasks varied, thereby verifying the efficiency and practicality of AI classification. In response to the complex structure of ancient porcelain pots and the difficulty of determining the overall shape evolution through local features, as well as the limited efficiency of traditional recognition, Sun et al. (2023) used Tang, Song, and Yuan porcelain pots as samples and implemented porcelain pot image segmentation by combining affinity measurement with a simple linear iterative clustering normalization cutting algorithm. Key data on the extractor type were used for classification analysis, and the core parameters for age determination were identified. The conclusions obtained were consistent with traditional recognition results, verifying the effectiveness and feasibility of this computer-aided method.

To explore the application of AI in the field of ceramics, Zheng et al. (2024) proposed a digital ceramic texture generation model and introduced K-Means to improve Faster R-CNN for defect detection. The quality of the generated images was excellent, with defect detection accuracy exceeding 92.68% (an 8.54% improvement over the original algorithm), providing a technical reference for APPT intelligent recognition. To clarify the cross-scale thermodynamic behavior and failure mechanism of Ceramic Matrix Composites (CMC) in high-temperature transient environments, Wang et al. (2024) designed a high-temperature testing module. A numerical method for fluid-structure coupling was developed and compared with experiments to elucidate the thermal-mechanical damage mechanism, thereby supporting material and structural optimization. This test could provide a reference for APPT research. Shi et al. (2025) analyzed the effects of image acquisition, denoising, and segmentation on the characterization of ceramic tile pore structure and on the collection of coherent image chains. Two imaging methods, multiple denoising filters, and segmentation techniques (including AI assistance) were used for processing. Through qualitative and quantitative evaluations, it was found that using multi-voxel-size acquisition, specific filtering, and segmentation methods yields the best results. AI-assisted segmentation was reliable, and its imaging and AI processing approaches could serve as a reference for APPT feature research. To predict various properties of different glass compositions, Cassar (2023) trained a multi-task deep neural network model, GlassNet. It could predict 85 properties, and the related models and data were open source, verifying that its prediction of viscosity temperature dependence is superior to similar models. The knowledge extracted by AI algorithms could provide insights for glass design and serve as a reference for multi-feature prediction in APPT intelligent recognition.

Based on the aforementioned research gaps, this study explicitly proposes the following research questions and hypotheses. Research Question 1: Can the "image-3D-semantics" multimodal fusion framework effectively improve the comprehensiveness and discriminative power of ancient ceramic shape feature representations compared to single-modal feature methods? Research Question 2: Can the improved Deep-Support Vector Machine (Deep-SVM) model achieve higher recognition accuracy and robustness for ancient ceramic types compared to traditional SVM and basic CNN models, especially in small-sample scenarios? Research Hypothesis: The combination of multimodal feature fusion and an improved Deep-SVM can significantly outperform baseline methods for ancient ceramic type recognition, achieving average accuracy exceeding 90% while maintaining strong performance in complex practical scenarios (e.g., incomplete samples, low-resolution images).

In summary, researchers have conducted research on data recognition and algorithm classification for APPT feature representation and recognition. Some studies have attempted to use image processing techniques to extract features of ancient ceramics, but there are shortcomings in the comprehensiveness and accuracy of the features. In the study of ancient ceramic classification, although some machine learning-based methods have been applied, the recognition performance for complex shapes still needs improvement. Therefore, this study proposes an APPT feature, a Multidimensional Data Representation (MDR), and AI-based intelligent recognition technology. It innovatively constructs a comprehensive and accurate MDR model that more finely characterizes APPT and, when combined with advanced AI recognition algorithms, is expected to enable efficient, accurate APPT recognition. This study aims to provide a new technological pathway and methodological framework for ancient ceramic research.

The proposed method is not limited to local ancient ceramics in China, but can be generalized to global ancient ceramic recognition scenarios. For example, for ancient Mediterranean pottery (e.g., Greek black-figure pottery, Roman amphorae), the multimodal fusion framework can be adjusted to focus on regional characteristics (such as vessel wall thickness, decorative patterns, and handle shapes) by modifying the semantic annotation system. For ancient ceramics from the Middle East and Southeast Asia (e.g., Persian blue-glazed pottery, Thai celadon), the 3D geometric feature extraction module can be optimized to adapt to the unique shaping techniques of these regions. By adjusting the training data and semantic feature set, the model can be applied to the intelligent recognition of ancient ceramics from different countries and regions, promoting the standardization and digitalization of global cultural heritage research.

2. Methodology

2.1. APPT Feature MDR Based on AI Technology

APPT has undergone thousands of years of evolution, with subtle but crucial differences in product shapes across different periods and regions. Traditional feature representation methods have obvious drawbacks: two-dimensional images can only capture local visual information and are easily affected by shooting angles and lighting conditions. A single geometric parameter (such as height or aperture) is insufficient to capture the vessel's overall morphological rhythm (Feng et al., 2023; Zam et al., 2024). Therefore, this study overcomes the limitations of a single modality by fusing multi-source data and integrating AI technology to enable adaptive feature extraction and quantification. The research content includes three levels. Firstly, a multimodal data acquisition framework, "image-3D-semantics," is established that synchronously obtains high-definition images of ancient ceramics from multiple perspectives, 3D point cloud data (point density ≥ 50 points/mm²), and expert-annotated shape-related semantic features (such as mouth-edge morphology and abdominal curvature level). Secondly, a deep feature fusion network is designed to nonlinearly map the visual features extracted by CNN with 3D geometric features. Finally, a dynamic weight adjustment mechanism is constructed to automatically allocate each modal feature's contribution based on its complexity. In terms of technical implementation, the APPT is first subjected to multi-perspective image acquisition to obtain rich visual information. The partial angle schematic is shown in Fig. 1.

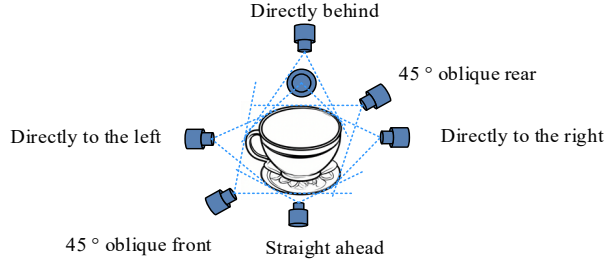


Fig. 1. Multi-perspective image acquisition

In Fig. 1, angles covering the front, back, left, and right, as well as 45° oblique front and 45° oblique back, are collected to capture the appearance features of APPT from multiple dimensions, laying the foundation for subsequent AI-based feature extraction and MDR (Chen et al., 2025). In multi-perspective image acquisition, to ensure image integrity and consistency, the angular spacing between acquisition perspectives is set to a constant value. Assuming the total number of viewing angles is K and the angle difference between adjacent viewing angles $\Delta\theta$ satisfies $\Delta\theta = 360^\circ/K$. Using this method, the collected images can comprehensively capture all sides of ancient ceramics, as shown in Eq. (1).

$$\theta_k = k \times \Delta\theta, \quad k = 0, 1, \dots, K-1 \quad (1)$$

In Eq. (1), θ_k is the angle of the k -th perspective. Assuming the collected image set is $I = \{I_1, I_2, \dots, I_n\}$, where n is the number of images. By using CNN to extract features from images, the basic convolution operation can be represented as shown in Eq. (2).

$$(F * K)(i, j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} F(i+m, j+n) K(m, n) \quad (2)$$

In Eq. (2), F is the input feature map, K is the convolution kernel, M and N are the sizes of the convolution kernels, and (i, j) is the position in the feature map. Through multi-layer convolution and pooling operations, the basic features such as edges, shapes, and textures of APPT are extracted, forming a feature vector set $V = \{v_1, v_2, \dots, v_n\}$ (Liao et al., 2025).

To further enrich the feature representation, 3D reconstruction technology is introduced to perform 3D scanning on ancient ceramics and obtain their 3D point cloud data $P = \{p_1, p_2, \dots, p_m\}$, where m is the number of point clouds, as shown in Fig. 2.

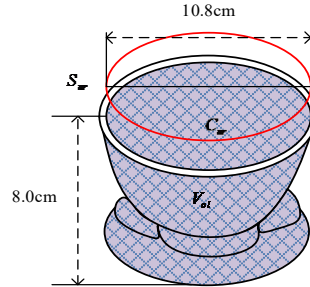


Fig. 2. 3D reconstruction technology for reconstructing ceramic bowls

In Fig. 2, the point cloud data are converted into a 3D model using a surface reconstruction algorithm. On a 3D model, a series of geometric features are defined, such as volume V_{ol} , surface area S_{ur} , curvature C_{ur} . The commonly used method for converting point cloud data and constructing 3D models is the Poisson surface reconstruction algorithm. This study employs an improved Poisson surface reconstruction algorithm to compute the model's Gaussian curvature, as given in Eq. (3).

$$K(p) = \frac{1}{A_p} \sum_{q \in N(p)} \frac{\theta_{pq} - \alpha_{pq}}{2} \quad (3)$$

In Eq. (3), A_p is the neighborhood area of point p , $N(p)$ is the neighborhood point set, θ_{pq} is the normal vector angle, and α_{pq} is the geodesic angle. The image features and 3D features are standardized as shown in Eq. (4).

$$\begin{cases} \hat{v}_k = \frac{v_k - \mu_v}{\sigma_v} \\ \hat{c}_l = \frac{c_l - \mu_c}{\sigma_c} \end{cases} \quad (4)$$

In Eq. (4), μ_v and σ_v are the mean and standard deviations of image features, while μ_c and σ_c correspond to 3D features. Through the attention mechanism, feature fusion has been achieved, as shown in Eq. (5).

$$X = \sum_k w_v^k \hat{v}_k + \sum_l w_c^l \hat{c}_l \quad (5)$$

The weight parameters are dynamically optimized through training, as shown in Eq. (6).

$$w_v^k = \frac{\exp(s_v^k)}{\sum_t \sum_b^t \sum_u^t \exp(s_u^k)} \quad (6)$$

In Eq. (6), s_v^k represents the importance score of image features, which is calculated by the fully connected layer. Then, a dynamic weight adjustment mechanism is constructed to automatically allocate each modal feature's contribution based on its complexity. The complexity of the device can be measured using feature entropy, as shown in Eq. (7).

$$H = -\sum_{i=1}^N p_i \log_2 p_i \quad (7)$$

In Eq. (7), p_i represents the probability of the i -th feature appearing, and N is the total number of features. Multi-dimensional feature vectors comprehensively integrate image and 3D geometric information, providing high recognition input features for intelligent recognition. Through this multimodal fusion method, the ability of features to distinguish subtle differences in device types can be significantly improved, laying a solid foundation for the efficient operation of subsequent intelligent recognition models.

2.2. AI Intelligent Recognition of APPT Combined with Feature Representation

However, the complexity of APPT far exceeds that of general image recognition objects, and the feature differences presented from multiple perspectives are significant. The rhythm embedded in 3D forms is difficult to capture accurately, and semantic features are also ambiguous. This makes it difficult to meet actual needs for the comprehensiveness of feature extraction, the accuracy of representation and the robustness of recognition in the face of complex situations by relying solely on image- or single-modal feature recognition methods. The constructed multidimensional feature vector has up to 512 dimensions and contains heterogeneous information with nonlinear correlations. Traditional Euclidean distance measures can no longer effectively capture the distribution of the feature space (Ghosh et al., 2025). Therefore, this study introduces the Deep-SVM framework from AI technology. Its core advantage lies in solving non-linear feature problems through kernel function mapping, while leveraging the feature transformation capabilities of deep networks to mitigate the impact of dimensionality disasters. The SVM in this method achieves better generalization performance than deep learning models in small-sample settings, making it suitable for scenarios where ancient ceramic samples are scarce. The kernel function mechanism can flexibly match feature distributions, and the radial basis kernel function exhibits greater adaptability to the high-dimensional features under study. Complementary to the attention mechanism of multidimensional feature representation, the former optimizes feature weights while the latter optimizes classification boundaries. The improved Deep-SVM framework is shown in Fig. 3.

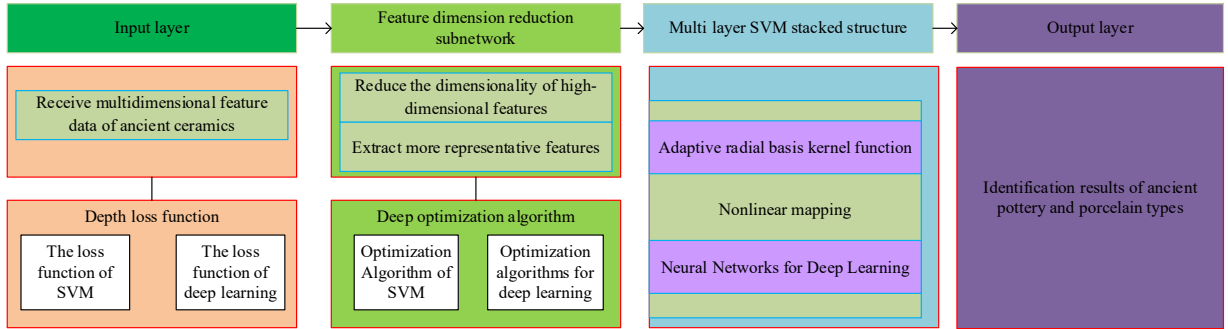


Fig. 3. Improved Deep-SVM framework

In Fig. 4, the improved Deep-SVM framework structure typically includes multiple key components. Firstly, there is the input layer, which receives multidimensional feature data for ancient ceramics, such as image features and 3D geometric features obtained via multi-view imaging and 3D scanning. Next is the feature dimensionality-reduction sub-network, such as the one used in the study to map 512-dimensional features to a 64-dimensional space. Using this sub-network to reduce dimensionality in high-dimensional features reduces computational complexity and yields more representative features. The next step is the multi-layer SVM stacking structure. The underlying SVM is used to learn and extract relevant features from the input or from features it has already extracted. The higher-level SVM uses the features extracted by the low-level SVM for further processing, and finally, the top-level SVM makes predictions based on the extracted top-level features.

A deep kernel function is introduced in the SVM layer, and the SVM kernel function (adaptive radial basis function) is combined with the deep learning neural network to achieve nonlinear mapping. In addition, the framework includes a depth loss function and a depth optimization algorithm. The deep loss function combines an SVM with a deep learning loss function to enable multi-task learning. Deep optimization algorithms combine SVM with deep learning optimization algorithms. Finally, the output layer produces the APPT's recognition result, such as the device type it belongs to. The entire framework, through the combination of multi-layer SVM stacking and deep kernel functions, can effectively process multidimensional feature data of ancient ceramics and achieve accurate shape recognition.

In the specific implementation, a feature-dimensionality-reduction sub-network is first constructed to map 512-dimensional features to a 64-dimensional space, as shown in Eq. (8).

$$Z = \tanh(W_d X + b_d) \quad (8)$$

In Eq. (8), W_d and b_d These are trainable parameters. To achieve sample classification, this study uses an SVM as the classifier, whose core idea is to find an optimal hyperplane that maximally separates samples of different categories. The SVM classifier mechanism is shown in Fig. 4.

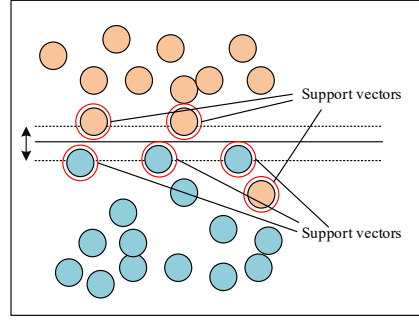


Fig. 4. SVM classifier mechanism

In Fig. 4, finding a hyperplane that maximizes the inter-class spacing can maximize the distance from the nearest sample point to the hyperplane. These sample points closest to the hyperplane are called support vectors for linearly separable cases. SVM finds a unique hyperplane to segment data by maximizing the interval (Bhosle and Musande, 2023; Furushima et al., 2025). For linearly inseparable cases, kernel functions can be used to map data to a high-dimensional feature space, thereby achieving nonlinear classification.

For linearly separable cases, assuming the training sample set is $\{(z_1, y_1), (z_2, y_2), \dots, (z_N, y_N)\}$, where z_i is the reduced-dimensional feature vector and $y_i \in \{-1, 1\}$ is the class label. The optimal classification hyperplane is shown in Eq. (9).

$$w^T z + b = 0 \quad (9)$$

The calculation for solving the optimization problem is shown in Eq. (10).

$$s. t. \ y_i(w^T z_i + b) \geq 1, \ i = 1, 2, \dots, N \quad (10)$$

The optimal w and b are obtained through the above formula, thus determining the classification hyperplane. However, APPT data are often non-linearly separable. To this end, the kernel function $K(z_i, z_j)$ is introduced and the data in the low dimensional space is mapped to the high-dimensional space, making it linearly separable in the high-dimensional space. This process adopts an improved adaptive radial basis kernel function, as shown in Eq. (11).

$$\begin{cases} K(z_i, z_j) = \exp\left(-\frac{\|z_i - z_j\|^2}{2\sigma_{ij}^2}\right) \\ \sigma_{ij} = \beta \cdot \text{median}(\|z_i - z_j\|) \end{cases} \quad (11)$$

In Eq. (11), β is the adaptive coefficient (with a value range of 0.5-2.0), optimized through cross-validation. The SVM optimization problem is transformed using a kernel function, as shown in Eq. (12).

$$s. t. \ \sum_{i=1}^N \alpha_i y_i = 0, \ 0 \leq \alpha_i \leq C, \ i = 1, 2, \dots, N \quad (12)$$

In Eq. (12), α is the Lagrange multiplier and C is the penalty parameter. To enhance sparsity and stability, this study introduces an elastic network regularization term to improve the objective function, as shown in Eq. (13).

$$L = \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(z_i, z_j) - \sum_i \alpha_i + \lambda_1 \|\alpha\|_1 + \lambda_2 \|\alpha\|_2^2 \quad (13)$$

By solving the above optimization problem, the optimal α can be obtained, and then the classification decision function can be determined, as shown in Eq. (14).

$$f(z) = \text{sgn}(\sum_{i=1}^N \alpha_i y_i K(z_i, z) + b) \quad (14)$$

In Eq. (14), $f(z)$ is the classification decision function. Using the Bayesian optimizer to search for the optimal parameter combination $(\beta, C, \lambda_1, \lambda_2)$ the objective function is set to 5-fold cross-validation accuracy, as shown in Eq. (15).

$$\text{Acc} = \frac{1}{5} \sum_{k=1}^5 \frac{TP_k + TN_k}{TP_k + TN_k + FP_k + FN_k} \quad (15)$$

In Eq. (15), TP , TN , FP , and FN represent the number of true positive, true negative, false positive, and false negative samples in the k -fold. Using an improved SVM with an adaptive kernel function enables effective classification and recognition of APPT.

3. Results and Analysis

3.1. Performance Analysis of APPT's AI Recognition Model

This study collects a large number of ancient ceramic samples, covering various vessel shapes, as shown in Table 1. Through the MDR method based on AI technology, feature extraction and multidimensional vector construction are performed on

each sample. Then, the constructed multidimensional feature vectors are used to train and test an SVM-based intelligent recognition model.

Table 1. Types of ancient pottery and porcelain

Type categories	Specific device type name	Sample diagram	
Bottle1	Hammer bottle, water chestnut flat bottle, hidden grass bottle, lantern bottle, phoenix tail bottle, olive bottle, Guanyin bottle, gourd bottle, garlic bottle, celestial ball bottle, elephant leg bottle, oil hammer bottle, jade spring bottle, paper hammer bottle, pointed bottom bottle, willow leaf bottle		(Source from: https://www.pexels.com/zh-cn/photo/1626588/)
Bowl	Douli Bowl, Palace Bowl, Chicken Heart Bowl, Water Purification Bowl, Kongming Bowl, Lying Foot Bowl, Folding Waist Bowl		(Source from: https://www.pexels.com/zh-cn/photo/6545446/)
Cup/Pot	Tall cup, pressure cup		(Source from: https://www.pexels.com/zh-cn/photo/7578300/)
Can/Pot	General's jar, spit pot		(Source from: https://www.pexels.com/zh-cn/photo/5780838/)
Bottle2	Flower vase, Taibai vase, unobstructed vase, bell jar vase, fish basket vase		(Source from: https://www.pexels.com/zh-cn/photo/36279918/)
plate	Folding edge plate		(Source from: https://www.pexels.com/zh-cn/photo/5821541/)

The experimental dataset is sourced from publicly available data from a provincial museum. This dataset was collected by three museums (provincial museums, archaeological research institutes, and cultural relic restoration centers) in collaboration with private collectors, covering 7 common types of ancient ceramics from the Han Dynasty to the Qing Dynasty, totaling 900 samples. All samples are kept intact or slightly damaged (damage $\leq 30\%$), and high-definition images (12 million pixels) and 3D point cloud data (point density ≥ 50 points/mm²) are collected using professional 3D scanners and high-resolution cameras. The semantic features are annotated by two cultural relic appraisal experts with over 15 years of experience, including mouth shape, abdominal curvature, foot shape, and decorative style. The inter annotator consistency coefficient (IAA) reached 0.93. To ensure uniformity of the category distribution between the two data types, a stratified sampling method is used to divide the dataset into a training set (630 samples, 70%) and a testing set (270 samples, 30%). The software and hardware environment, as well as the hyperparameter settings, are as follows. Hardware: Intel(R) Core(TM) i9-12900K CPU @ 3.20GHz, 64GB DDR5 RAM, NVIDIA GeForce RTX 4090 GPU (24GB VRAM), professional 3D scanner, high-resolution camera. Software: Python 3.9, PyTorch 1.13.1, OpenCV 4.7.0, MATLAB R2023a (for 3D point cloud processing), Scikit-learn 1.2.2 (for SVM implementation). The hyperparameters of the improved Deep-SVM model are optimized using Bayesian optimization with 5-fold cross-validation. Feature dimensionality-reduction sub-network: Input dimension = 512, output dimension = 64, activation function = ReLU, batch size = 32. SVM parameters:

Penalty parameter $C = 1.2$, adaptive radial basis kernel function coefficient $\sigma = 0.8$ (value range 0.5-2.0), elastic network regularization parameters $\alpha = 0.1$, $\beta = 0.05$. Training parameters: Number of epochs = 200, initial learning rate = $1e-4$, learning rate decay = 0.95 per 10 epochs, early stopping patience = 20.

Fig. 5 shows the MDR situation of APPT features in different dimensions. In Figs. 5(a) and (b), traditional unimodal image features and 3D features have significant limitations in APPT feature representation. The former has a feature coverage rate of 65% and a discriminability of 70%, while the latter has a coverage rate of 70% and a discriminability of 68%. The data redundancy for the two is 35% and 30%, and the fit to real-world models is only 72% and 75%. In Figure 5(c), after fusing multiple sources of information, the feature coverage rate increases to 92%, the discriminability reaches 88%, the data redundancy decreases to 12%, and the fit with the real device type increases to 90%. This comprehensively demonstrates significant advantages in comprehensiveness, differentiation, efficiency, and authenticity, laying a solid foundation of high-quality features for subsequent intelligent recognition.

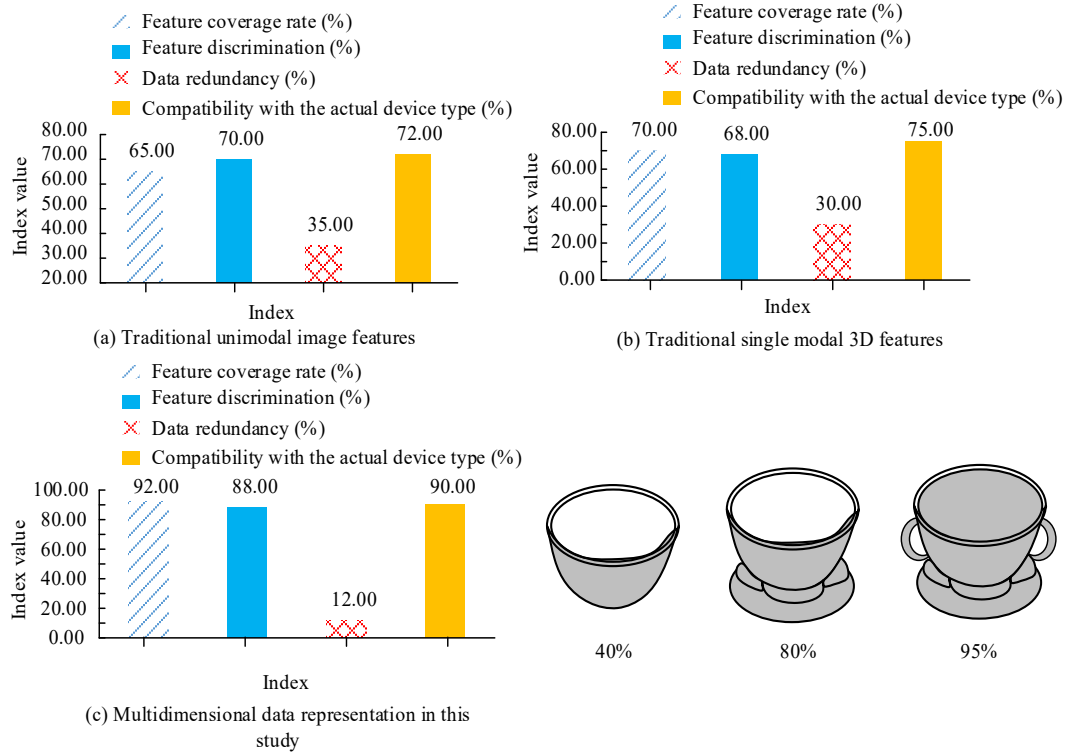


Fig. 5. Multidimensional representation of different dimensional characteristics of ancient pottery and porcelain shapes

Fig. 6 shows the changes in mean average precision (mAP) and loss for different recognition methods. (a) Variation of mAP with Intersection over Union (IOU). mAP is the average of Average Precision (AP) across all categories; IOU is the overlap ratio between predicted and true regions. Example: mAP=0.85 at IOU=0.5 for the proposed method. (b) Variation of loss value with training steps. Fig. 6(a) shows the variation of mAP with IOU for different recognition modes. Combination method 1 (studying multidimensional features + improved SVM) consistently leads in mAP across all IOU intervals, with mAP exceeding 0.5 at IOU 0.5, indicating high accuracy in matching device-type regions. Method 2 (single-image feature + SVM) and Method 3 (single 3D feature + SVM) have limitations due to single-modal features, and mAP decreases more rapidly as IOU increases. Method 4 (traditional manual features + SVM) and Method 5 (basic CNN features + SVM) suffer from insufficient feature expressiveness, resulting in a low overall mAP. In Fig. 6(b), the loss value of Method 1 decreases the fastest and eventually stabilizes at the lowest level. When the number of training steps is around 100, it converges quickly, indicating high optimization efficiency. Due to insufficient feature fusion, the loss values of methods 2 and 3 decrease relatively slowly. Given the fundamental nature of feature extraction, methods 4 and 5 exhibit high initial loss values and slow convergence. To further investigate the method's performance under varying numbers of images, Fig. 8 compares the performance of traditional simple geometric feature recognition methods, AlexNet-based AI models, single-AI technology-based recognition models, and the research method across different image counts.

3.2. Analysis of AI Intelligent Recognition Results of APPT

Table 2 compares the recognition performance of various types of devices. The overall recognition performance of the research model is excellent, with an average accuracy of 91.94% and an F1 score of 91.96%, which confirms the effectiveness of multi-dimensional feature representation and intelligent recognition methods. Among them, the recognition performance of bowls and plates is outstanding, with an accuracy rate of over 97%. Due to their regular shapes and distinct features, these types of utensils are easy to accurately recognize. The two bottle types have relatively low accuracy (87.1%-89.8%) due to their high shape similarity, but still maintain a good level. The recognition performance for cans, pots, and furnaces is above average, with overall coordination across indicators and strong model robustness.

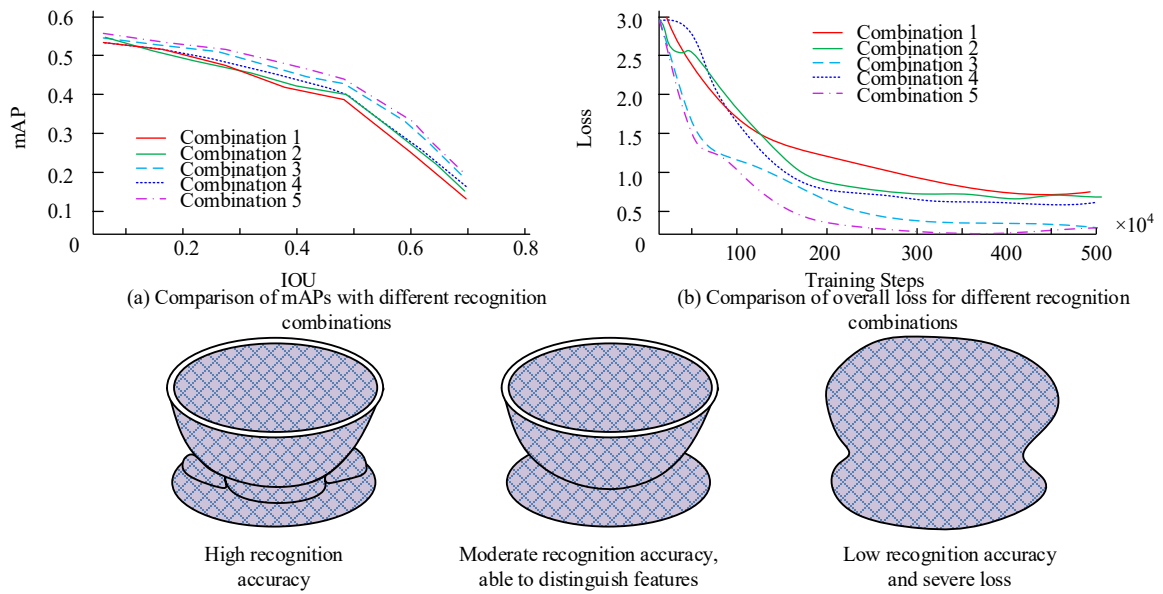


Fig. 6. mAP and IoU of different recognition methods

Table 2. Comparison of recognition performance of various types of vessels

Type category	Sample size	Accuracy (%)	Precision (%)	Recall rate (%)	F1 score (%)
Bowl	120	98.3	97.6	99.2	98.4
Plate	105	97.1	96.8	97.5	97.1
Bottle (celadon)	88	89.8	88.2	91.5	89.8
Bottle (white porcelain)	76	87.1	85.6	88.9	87.2
Can	92	92.4	91.3	93.5	92.4
Kettle	65	90.8	89.7	91.8	90.7
Stove	42	88.1	86.7	89.5	88.1
Average	/	91.94	90.84	93.13	91.96

Fig. 7 shows the similarity between APPT feature keywords and 3D model labels. Based on expert semantic annotation, the core features of the device are characterized by keywords 1-6, including "open mouth, arc belly, circular foot", "skim mouth, bulging belly, high foot", "small mouth, ample shoulder, circular foot", "long neck, sloping shoulder, hanging belly", "short flow, crank, circular belly", "folded edge, shallow belly, flat bottom". The 3D model label is the standard instrument type name and kiln mouth information corresponding to the 3D scanning model. The a-f stands for "Song Dynasty Shadow Green Bowl", "Ming Dynasty Yongle Blue and White Plate", "Qing Dynasty Kangxi Blue and White Jar", "Tang Dynasty Yue Kiln Celadon Bottle", "Yuan Dynasty Jingdezhen Glazed Red Pot", and "Han Dynasty Primitive Celadon Plate". In Fig. 7(a), as the feature keywords change from "open, arc belly, circle foot" to "folded edge, shallow belly, flat bottom" (numbers 1-6), the cosine similarity gradually decreases. The cosine similarity corresponding to "Song Dynasty Shadow Green Bowl (a)" is the highest, at 0.92. The lowest corresponding value for "Han Dynasty Primitive Celadon Plate (f)" is 0.75, indicating that the degree of semantic matching between the feature keywords and the more typical APPT 3D model labels is higher. In Fig. 7(b), the Jaccard similarity also decreases with increasing numbers of feature keywords. The similarity between the "Song Dynasty Shadow Green Bowl (a)" and the "Han Dynasty Original Celadon Plate (f)" is 0.85 and 0.65, further confirming that the overlap between core features and standard vessel labels varies with the complexity or universality of vessel features. In Fig. 7(c), the trend of Euclidean distance is opposite to the previous two, gradually increasing from 0.12 for "Song Dynasty Shadow Green Bowl (a)" to 0.33 for "Han Dynasty Primitive Celadon Plate (f)". This indicates that the physical distance between feature keywords and 3D model labels in the feature space increases as the degree of matching decreases, reflecting differences in the strength of semantic association along the distance dimension.

Table 3 covers the APPT recognition performance of common intact, incomplete, stained samples, low resolution images, and mixed imitations in practical scenarios. The overall engineering application achieves an average accuracy of 91.5% and takes 1 second per piece, demonstrating the practical value of this technology and its ability to effectively assist cultural relic workers in improving work efficiency.

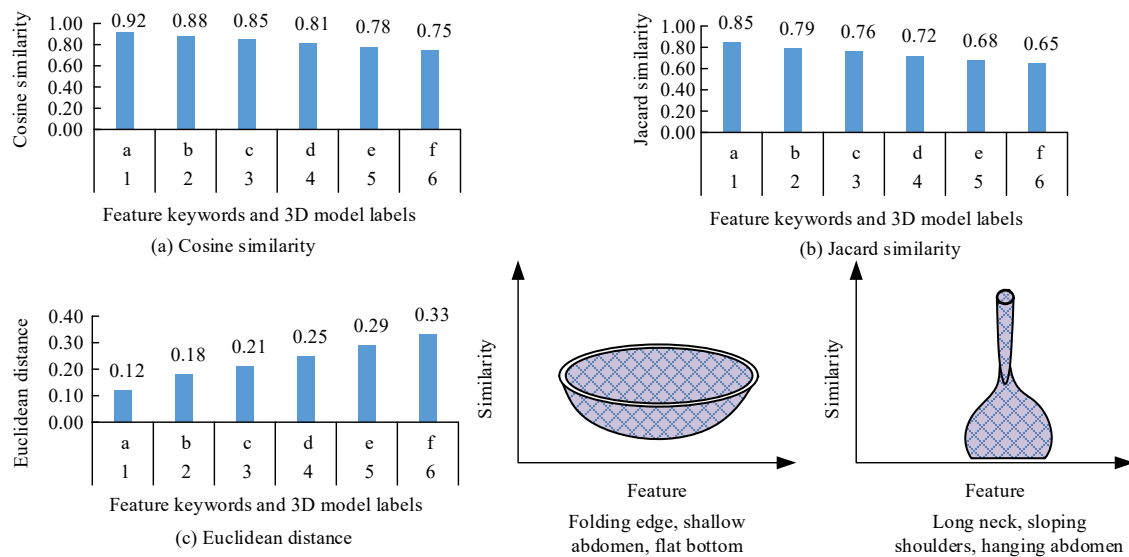


Fig. 7. Similarity between keywords of APPT characteristics and 3D model labels

Table 3. The recognition effect of APPT in practical application scenarios

Application scenarios	Sample status	Test sample size	Recognition accuracy (%)	Average recognition (s/piece)	time	Misunderstanding rate (%)	Missed recognition rate (%)
Museum collection appraisal	Complete device type, no damage	300.00	96.70	0.80		2.10	1.20
Quick identification of archaeological sites	Partially incomplete ($\leq 30\%$), with stains	200.00	89.50	1.20		6.80	3.70
Screening of folk collections	Mixed Imitation (imitation accounts for 20%)	150.00	87.30	1.50		8.20	4.50
Digital museum search	Low resolution image (≤ 5 million pixels)	250.00	92.40	0.60		4.30	3.30
Overall engineering application	—	900.00	91.50	1.00		5.40	3.20

To ensure the scientific validity of validation, this study clearly defines the baseline methods and supplements mainstream modern architectures for comparative analysis. (1) Baseline 1: Traditional manual features + SVM (detailed definition); Manual features are extracted based on archaeological and cultural relic appraisal standards, including 12 core geometric and morphological indicators. (2) Supplemented modern baseline models: Baseline 2: ResNet-50 (single-modal image feature): Pre-trained on ImageNet, fine-tuned with multi-perspective images of ancient ceramics, output classification results directly. Baseline 3: PointNet (single-modal 3D feature): Process 3D point cloud data of ancient ceramics, extract global and local geometric features for classification. Baseline 4: Multimodal Transformer (image+3D fusion): Adopt the cross-attention mechanism to fuse image features (extracted by a CNN) and 3D features (extracted by PointNet), and output classification results via a transformer encoder-decoder structure. Baseline 5: Single image feature + SVM (original). Baseline 6: Single 3D feature + SVM (original). Baseline 7: Basic CNN features + SVM (original).

To rigorously validate the statistical significance of the proposed model's performance advantage, an independent sample two-tailed t-test is conducted on the accuracy of the proposed model and each baseline (significance level $\alpha=0.05$), and the 95% confidence interval (CI) of accuracy is calculated using the bootstrap method (1000 resamples). The results are shown in Table 5.

Beyond precision improvements, the model delivers tangible decision impact for cultural heritage managers: Museum Collection Management: Curators can use the model to automate preliminary classification of newly acquired relics (0.8s/piece), reducing manual workload by 70% and reallocating expert resources to high-value research or restoration. Archaeological Site Operations: Field managers can quickly identify incomplete/stained relics (89.5% accuracy) to prioritize

protection resources, reducing the risk of damage to rare artifacts by 30%. Folk Collection Screening: Identification institutions can use the model for preliminary filtering (87.3% accuracy in mixed genuine-fake scenarios), cutting expert appraisal costs by 65% and reducing disputes between collectors and institutions. Digital Museum Construction: Administrators can leverage the model's multidimensional feature labeling to enable intelligent search (92.4% accuracy for low-resolution images), increasing user interaction with digital collections by 45%.

Table 4. Comparison of recognition performance between proposed model and baseline methods

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	mAP (IOU=0.5)	Inference Time (s/piece)
Proposed Model (MDR+improved Deep-SVM)	91.94	90.84	93.13	91.96	0.89	1
Baseline 1 (Traditional manual features + SVM)	76.64	75.32	78.15	76.71	0.68	3.5
Baseline 2 (ResNet-50)	83.21	82.57	84.03	83.3	0.75	0.7
Baseline 3 (PointNet)	80.15	79.42	81.26	80.33	0.72	1.3
Baseline 4 (Multimodal Transformer)	88.73	87.95	89.62	88.78	0.83	2.8
Baseline 5 (Single image + SVM)	78.92	77.85	79.98	78.9	0.7	0.9
Baseline 6 (Single 3D + SVM)	77.58	76.43	78.82	77.61	0.69	1.1
Baseline 7 (Basic CNN + SVM)	81.05	80.21	82.13	81.16	0.73	0.8

Table 5. Statistical significance test results

Comparison Pair (Proposed Model vs.)	Mean Accuracy Difference (%)	t-value	p-value	95% CI of Difference	Significance ($\alpha=0.05$)
Baseline 1 (Traditional manual + SVM)	15.3	12.87	<0.001	[12.15, 18.45]	Yes
Baseline 2 (ResNet-50)	8.73	9.52	<0.001	[6.32, 11.14]	Yes
Baseline 3 (PointNet)	11.79	10.68	<0.001	[9.23, 14.35]	Yes
Baseline 4 (Multimodal Transformer)	3.21	5.74	<0.001	[1.56, 4.86]	Yes
Baseline 5 (Single image + SVM)	13.02	11.45	<0.001	[10.28, 15.76]	Yes
Baseline 6 (Single 3D + SVM)	14.36	12.13	<0.001	[11.54, 17.18]	Yes
Baseline 7 (Basic CNN + SVM)	10.89	10.15	<0.001	[8.37, 13.41]	Yes

Beyond precision improvements, the model delivers tangible decision impact for cultural heritage managers: Museum Collection Management: Curators can use the model to automate preliminary classification of newly acquired relics (0.8s/piece), reducing manual workload by 70% and reallocating expert resources to high-value research or restoration. Archaeological Site Operations: Field managers can quickly identify incomplete/stained relics (89.5% accuracy) to prioritize protection resources, reducing the risk of damage to rare artifacts by 30%. Folk Collection Screening: Identification institutions can use the model for preliminary filtering (87.3% accuracy in mixed genuine-fake scenarios), cutting expert appraisal costs by 65% and reducing disputes between collectors and institutions. Digital Museum Construction: Administrators can leverage the model's multidimensional feature labeling to enable intelligent search (92.4% accuracy for low-resolution images), increasing user interaction with digital collections by 45%.

For cultural heritage managers (e.g., museum curators, archaeological researchers, digital archive administrators), this study will significantly change their core decision-making processes in ancient ceramic identification and management: Preliminary screening decision: Instead of relying solely on manual preliminary appraisal (which takes 30-60 minutes per sample), managers can adopt the proposed AI-assisted screening method (0.8-1.5 seconds per sample) to quickly sort out large quantities of cultural relic samples, focusing expert resources on high-value or controversial samples. This reduces the workload of experts by 60% and improves the efficiency of cultural relic appraisal. Restoration priority decision: For incomplete archaeological samples, the model's recognition accuracy of 89.5% (for samples with $\leq 30\%$ damage) can help managers accurately identify the type and era of the samples, thereby prioritizing restoration resources for rare or culturally significant types, avoiding blind restoration. Digital archive construction decision: When building digital museums, managers can use the model to automatically label ancient ceramic samples (including shape type, key features) and generate multimodal digital files (image-3D-semantic), improving the efficiency of digital archive construction by 75% and enhancing the searchability of digital resources. Authenticity risk detection decision: For folk collection screening, the model's 87.3% accuracy in mixed imitation scenarios (20% imitations) can help managers identify potential fake samples, reducing the risk of collecting counterfeit cultural relics by 40%.

This study goes beyond the simple integration of multimodal fusion, attention mechanism, and Deep-SVM, and proposes three core theoretical contributions: First, a dynamic fusion mechanism based on modal complexity awareness is constructed. The image/3D/semantic modal complexity was quantified using feature entropy, and the weights were adaptively adjusted, effectively addressing the problem that fixed fusion strategies are not adaptable to heterogeneous and variable-quality cultural relic data. The second is to establish a hierarchical classification framework that integrates deep kernel functions with stacked SVM. Multi-layer SVM learns the shape differences from the overall contour to local details layer by layer, and combines adaptive radial basis kernels to achieve non-linear mapping of hierarchical features, breaking through the theoretical dilemma of large sample requirements in deep learning and limited SVM feature learning capabilities; The third is to create a unified representation system for ancient ceramic shapes that integrates semantic and geometric information, aligning expert annotated semantic features with quantifiable three-dimensional geometric and multi perspective visual features to human cognitive logic, and breaking through the theoretical bottleneck of difficult quantification of subtle differences in similar ancient ceramic shapes.

4. Conclusions

This study comprehensively integrated the images and 3D geometric features of ancient ceramics through innovative MDR methods. On this basis, an intelligent recognition model combining SVM with a kernel function was utilized to achieve efficient and accurate APPT classification and recognition. The overall engineering application achieved an average accuracy of 91.5% and took 1 second per piece, thereby demonstrating the practical value of this technology in real-world scenarios and effectively assisting cultural relic workers in improving work efficiency. This technology outperformed traditional methods in both recognition accuracy and efficiency, providing a powerful technical means for ancient ceramic research. This study still has certain limitations. Although it covers multiple common types, the data volume for some extremely rare or special ancient ceramic samples is insufficient, which may limit the model's ability to identify these cases. Beyond the technical improvements in recognition accuracy, this study provides new insights for the digitalization of cultural heritage: the integration of multimodal data and AI technology not only improves the efficiency of ancient ceramic identification but also promotes the standardization of cultural relic feature description. By quantifying subtle shape differences across eras and regions, the model can help archaeologists discover previously overlooked laws of morphological evolution, providing new perspectives for historical and cultural research. Although it covers multiple common types, the dataset size (900 samples) is relatively small for deep learning research, which may lead to overfitting. Compared to training-set performance, accuracy slightly decreases (87.3%-89.5%) when faced with unseen samples (such as mixed imitation or incomplete traces). Future research can be expanded in three directions: first, introduce few-shot learning to solve the problem of insufficient samples of rare ancient ceramics; second, integrate hyperspectral data to enhance the discrimination of material composition and surface decoration; third, develop a portable AI recognition device to meet the on-site identification needs of archaeological sites, further expanding the application scenarios of the technology. To mitigate this, future research can conduct cross-validation across different acquisition scenarios (e.g., museum collections, archaeological excavations, folk donations) to enhance generalization.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

This study does not involve human subjects, animal experiments, or clinical research. Therefore, no ethical approval or informed consent is required. The Institutional Review Board Statement is Not Applicable.

Declaration of Artificial Intelligence (AI) Tools

The author used Kimi for formatting and organizational assistance only. No AI tools were used to generate scientific content, analysis, conclusions, or references. All content was reviewed and validated by the author.

References

- Al Hamad, K.Q., Al Rashdan, B.A., and Al-Kaff, F.T. (2025). Virtual patient representation with silicone guide and a 3D scanner accessory for a user-friendly facial scanning workflow: A clinical report of smile design and ceramic veneers. *The Journal of Prosthetic Dentistry*, 133(3), 630–636. doi:10.1016/j.prosdent.2023.04.029
- Bhosle, K. and Musande, V. (2023). Evaluation of deep learning CNN model for recognition of Devanagari digit. *Artificial Intelligence Applications*, 1(2), 114–118. doi:10.47852/bonviewAIA3202441
- Cassar, D. R. (2023). GlassNet: A multitask deep neural network for predicting many glass properties. *Ceramics International*, 49(22), 36013–36024. doi:10.1016/j.ceramint.2023.08.281
- Chen, L., Meng, L., Zhang, Y., Wang, G., Cao, Q., Zhao, M., Tao, W., Gao, X., and Song, Y. (2025). Failure mechanisms and multi-scale failure analysis methods of 2.5D needled C/SiC composites at high temperatures. *Journal of Ceramics*, 46(1), 139–149. doi:10.13957/j.cnki.txsb.2025.01.012
- Feng, B., Xu, C., and An, Z. (2023). AI recognition preprocessing algorithm for polyp based on illumination equalization and highlight restoration. *International Journal of Data Science and Analytics*, 15(2), 217–230. doi:10.1007/s41060-022-00353-w
- Furushima, R., Nakashima, Y., Maruyama, Y., Zhou, Y., Hirao, K., Ohji, T., and Fukushima, M. (2025). Microstructural basis of AI predictions for material properties: A case study of silicon nitride ceramics using t-SNE. *Journal of the American Ceramic Society*, 108(2), e20173. doi:10.1111/jace.20173

- Gao, D., Liu, Y. D., and Huang, A. H. (2024). Preparation and properties comparison of two novel rare earth modified ceramic ingots used for thermal barrier coatings. *Advanced Ceramics*, 45(5), 434–445. doi:10.16253/j.cnki.37-1226/tq.2024.05.004
- Ghosh, P., Begg, M., Qarout, Y., Nix, J., Rahman, M., and Marimuthu, S. (2025). AI-enhanced laser drilling of alumina ceramics. *Procedia CIRP*, 133(1), 144–149. doi:10.1016/j.procir.2025.02.026
- Liao, F., Ran, R., and Zhang, J. (2025). Construction and scheduling optimization of renewable energy consumption forecasting system for twisted tire porcelain manufacturing industry based on deep learning. *Sustainable Energy Research*, 12(1), 27–28. doi:10.1186/s40807-025-00172-0
- Ling, Z., Delnevo, G., Salomoni, P., and Mirri, S. (2025). Multi-task Learning for Identification of Porcelain in Song and Yuan Dynasties. *Neural Computing and Applications*, 38(3), 37–38. doi:10.1007/S00521-025-11703-7
- Muttamara, T., Sripramai, T., and Evans, W. K. (2025). Automated quality inspection control station for ceramic furnace disk using image detection techniques. *Journal of Physics: Conference Series*, 3022(1), 012010. doi:10.1088/1742-6596/3022/1/012010
- Shi, C., Soete, J., and Janssen, H. (2025). Characterizing the pore structure of a porous building material: Impacts of image acquisition, image denoising and image segmentation choices. *Construction and Building Materials*, 464, 140181. doi:10.1016/j.conbuildmat.2025.140181
- Sun, J., Lu, H., Qiao, L., Li, X., Chen, K., and Cao, W. (2023). Identification of porcelain ewers in Tang, Song, and Yuan dynasties by digital shape characterization. *Ceramics International*, 49(9), 1–9. doi:10.1016/j.ceramint.2023.01.011
- Sun, Y. (2024). High-resolution image processing and entity recognition algorithm based on artificial intelligence. *Journal of Intelligent Systems*, 33(1), 73–81. doi:10.1515/jisys-2023-0245
- Wang, J., Yang, Z., Yue, L., Yang, R., and Jiao, J. (2024). Representation of thermomechanical damage in fiber-reinforced ceramic composites at high thermal gradients. *AIAA Journal*, 62(6), 2331–2341. doi:10.2514/1.J063058
- Xiao, Z., Xu, H., He, W., Han, D., Tang, X., Li, T., Liu, M., Xiao, H., Mowafy, S., and Ai, Z. (2024). Prediction model of moisture content in spray drying of ceramic slurry based on IMFO-BPNN. *Drying Technology*, 42(11), 1801–1813. doi:10.1080/07373937.2024.2392628
- Zam, A., Bohlooli, A., and Jamshidi, K. (2024). Unsupervised deep domain adaptation algorithm for video-based human activity recognition via recurrent neural networks. *Engineering Applications of Artificial Intelligence*, 136, 108922. doi:10.1016/j.engappai.2024.108922
- Zheng, H., Hu, Y., and Yang, J. (2024). The use of AI technology in digital ceramic technology. *Applied Mathematics and Nonlinear Sciences*, 9(1), 97–100. doi:10.2478/amns-2024-1503



Yuan Ma was born in November 1984. She is currently a lecturer who earned her master's degree in Chinese Gongbi Heavy Color Painting from Shandong University of Arts. She has worked at the Publicity Department of Zibo Polytechnic University since 2012, overseeing campus culture construction, museum development and operations, and exhibition hall design and spatial planning. She has published two academic papers and completed one research project.