

<https://doi.org/10.1038/s40494-025-01886-6>

Integrating deep learning and machine learning for ceramic artifact classification and market value prediction

Yanfeng Hu^{1,3}, Siqi Wu^{2,3}✉, Zhuoran Ma² & Si Cheng¹

This study proposes an intelligent framework for the automated classification and valuation of ceramic artifacts, integrating deep learning and machine learning techniques. An improved YOLOv11 model was constructed to identify key ceramic attributes such as decorative patterns, shapes, and craftsmanship styles. The model achieved a mean Average Precision (mAP@50) of 70.0% and a recall of 91.0%, demonstrating strong capability in detecting complex visual features. Based on the extracted visual attributes, a Random Forest classifier was employed to predict price categories using multi-source auction data, achieving a test accuracy of 99.52%. Feature importance analysis further revealed manufacturing techniques and shape as key predictors of market value. The integrated framework effectively combines visual feature extraction and market-informed valuation, providing a scalable solution for intelligent ceramic appraisal and digital heritage curation. This approach supports both expert and non-expert applications, laying a foundation for future development of intelligent cultural heritage management systems.

Ceramics are an important symbol of Chinese culture, embodying thousands of years of artistic and technological heritage. They encompass a wide range of types, from pottery and painted ceramics to porcelain¹. Swanson and Timothy² (p. 45) highlight that ceramics serve as both artistic and utilitarian symbols. Beyond representing the esthetic aspirations of different historical periods, ceramics also play a crucial role in cultural preservation³.

In addition to their cultural significance, the ceramics industry is a vital part of China's manufacturing sector. Statistics indicate that the annual production of daily-use ceramics in China grew from 49.1 billion pieces in 2017 to 67.9 billion pieces in 2023, with an average annual growth rate of 5.55%⁴. According to Grand View Research⁵ (2021), the global ceramics market is projected to reach USD 347 billion by 2028, demonstrating immense economic potential. The compound annual growth rate (CAGR) between 2021 and 2028 is expected to be approximately 4.4%. These figures reveal that the ceramics market has vast potential for growth and development.

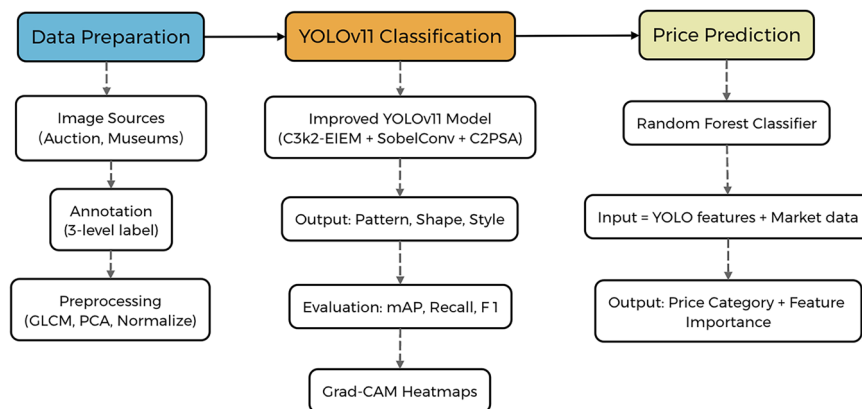
With the rapid advancement of deep learning and computer vision technologies, image-based ceramic classification has become increasingly prevalent. These techniques demonstrate efficient and objective classification capabilities through methods such as feature extraction, image segmentation, and image enhancement^{6,7}. Prior to deep learning dominance, researchers had already begun exploring how computer vision (CV) techniques could facilitate automated craftsmanship identification and address

the inefficiencies of traditional visual methods^{8–11}. Traditional approaches to ceramic classification included empirical identification—highly reliant on expert knowledge and subjective visual judgment^{12,13}—as well as scientific identification methods such as X-ray fluorescence, thermoluminescence dating, and spectral analysis, which, although precise, require complex instrumentation and domain expertise, limiting accessibility for non-professionals¹⁴. Recent studies have further enriched scientific identification methods. For example, stereoscopic and polarizing microscopes have been used to analyze celadon from different dynasties¹⁵, and compositional analysis has helped determine kiln origins¹⁶. Additionally, diffuse reflectance spectral data have been employed to capture color characteristics across ceramic types¹⁷.

Early computational methods applied hand-crafted feature descriptors such as Gradient Vector Flow (GVF) and Local Binary Patterns¹⁸, Gray-Level Co-occurrence Matrices¹⁹, and morphological profile curves^{20–22}. While these techniques showed promise in small-scale contexts, they lacked adaptability and interpretability. More recent studies have leveraged convolutional neural networks (CNNs)^{7,13,23,24}, transfer learning²⁵, and capsule networks²⁶, achieving superior performance on complex ceramic datasets. Some studies have automated the analysis of ceramics using visual attributes such as texture, color, and shape^{27–29}. Kernel mean shift clustering and Bag of Visual Words (BoVW) has performed well in ceramic feature extraction^{30–32}. Liu⁷ applied CNNs to classify Yaozhou kiln ceramics with

¹Faculty of Design and Architecture, Universiti Putra Malaysia, Serdang, Selangor, Malaysia. ²Arts, Design & Architecture, UNSW Sydney (The University of New South Wales), Oxford St & Greens Rd, Sydney, NSW, Australia. ³These authors contributed equally: Yanfeng Hu, Siqi Wu. ✉e-mail: z5601507@ad.unsw.edu.au

Fig. 1 | Overview of ceramic classification workflow. This figure illustrates the full pipeline adopted in this study, comprising three sequential phases: data preparation, YOLOv11 based classification, and regression-based price prediction.



high accuracy, while Cui et al.³³ utilized a deep learning model to enhance image clarity and detect surface defects. In terms of decorative patterns recognition, Chaowalit and Kuntitan³⁴ demonstrated improved accuracy using CNNs, while Santos et al.³⁵ extended CNN applications to classify both undecorated Roman amphorae and Portuguese faience with visually subtle patterns.

In the fields of engineering and manufacturing, research on ceramics has primarily focused on defect detection and the optimization of automated production processes, often prioritizing classification accuracy while overlooking interpretability and artistic value^{36,37}. Although models such as YOLO and ResNet can accurately distinguish surface patterns, they seldom explain why particular decorative motifs signify specific value tiers or dynastic origins.

Similarly, in archeology and cultural studies, Yi et al.²⁹ and Mei et al.³⁸ observe that most existing research concentrates on production periods and kiln origins, with limited exploration of complex craftsmanship styles. This aligns with broader archaeometric efforts that utilize XRF, TL dating, and petrographic analysis to study ceramic provenance and function^{39–41}. Zhan et al.⁴² highlighted how motif complexity and decorative technique reflect regional kiln styles and influence historical price trends, underscoring the need to integrate esthetic and economic factors into computational models. While disciplines such as ceramic archeology and art history have long addressed the symbolic and historical significance of ceramics^{43,44}, such cultural insights are rarely integrated into computational frameworks. As Finlay⁴⁵ emphasized, porcelain has historically played a socio-political role as a medium of global exchange, highlighting that ceramic valuation should incorporate both material analysis and cultural interpretation. Although the exceptional performance of convolutional neural networks (CNNs) in general image classification has been widely recognized^{37,46}, their application to ceramic datasets often remains limited to basic texture recognition, lacking both explainability and integration of artistic-level analysis^{1,7}. Furthermore, while market-based valuation approaches in machine learning offer practical utility, they are not universally applicable. For instance, in archeological research, the value of an artifact is typically assessed based on its historical and cultural significance rather than its market price⁴⁷.

Despite these technological advancements, several key challenges persist. First, there is a lack of large-scale ceramic datasets that encompass a wide range of craftsmanship styles and high-resolution images. Second, existing models often lack interpretability, with most focusing on classification accuracy while failing to provide intuitive classification criteria. Third, user experience remains insufficiently addressed. Current studies are primarily designed for academic or professional applications, with a notable lack of tools or solutions tailored to non-professional enthusiasts.

This study aims to address the identified research gaps through the following approaches: (1) Develop an integrated framework for ceramic

artifact valuation: the primary objective of this research is to establish a robust, data-driven framework that integrates both artistic features and market data for the accurate evaluation of ceramic artifacts. By combining traditional art historical attributes (such as decorative patterns, shapes, and craftsmanship style) with advanced machine learning models, this study aims to provide a systematic and objective approach to ceramic price categorization. (2) Enhance ceramic classification using YOLO model improvements: this study seeks to refine the process of ceramic classification by leveraging and modifying the YOLO (You Only Look Once) model. The research focuses on optimizing the model's performance in detecting and classifying ceramic types, which are crucial inputs for subsequent price prediction tasks. (3) Predict price categories using random forest classification: by incorporating the features extracted from the YOLO model and structured auction data, the objective is to employ Random Forest classification to predict the price categories of ceramic artifacts with high accuracy and interpretability.

A key innovation in this research is the hybrid integration of visual modeling and economic reasoning. YOLOv11 is optimized using attention enhancements, enabling it to focus on intricate design motifs and subtle manufacturing traits. These features are passed to a Random Forest classifier trained on multi-year auction data from institutions such as Christie's, Sotheby's, Poly Auction, and China Guardian.

Unlike conventional valuation tools, this model offers not only high predictive performance but also interpretability, identifying which craftsmanship attributes most influence valuation. Ultimately, this study bridges the gap between deep learning and art-historical appraisal, contributing a culturally aware, technically rigorous framework for the digital future of ceramic classification and valuation.

Methods

This study adopts a three-stage pipeline: dataset preparation, ceramic classification, and price prediction, as shown in Fig. 1. The stages are seamlessly integrated to support robust ceramic analysis and valuation.

The first stage of our pipeline involved the construction and annotation of a high-quality ceramic image dataset, combining automated acquisition with expert-informed labeling strategies. Chinese ceramics have a rich history, with significant differences in craftsmanship, forms, and decorative patterns across various historical periods¹⁸. Building on the ceramic classification frameworks proposed by Mu et al.²⁷ and Yi et al.²⁹, this study categorized ceramics based on shape, decorative patterns, and production techniques. A high-quality dataset of 8213 high-resolution images was constructed, representing 20 distinct craftsmanship styles and decorative patterns selected for their historical significance and visual distinctiveness. These 20 styles include both kiln-specific categories and decorative techniques, such as Blue and White Porcelain, Doucai, Wucai, Fencai, Ru Kiln, Guan Kiln, Ge Kiln, Jun Kiln, Ding Kiln, Longquan Kiln, Yingqing Porcelain, White Porcelain, Sacrificial Blue/Red Porcelain, Langyao Red, Tea-dust Glaze, Reticulated

Porcelain (Linglong), Cizhou Kiln Porcelain, Falangcai, and Fahua. The sample distribution among the 20 craft styles is moderately imbalanced. For instance, well-represented styles such as Blue and White Porcelain and Longquan Kiln Porcelain each contain over 600 samples, comprising approximately 8% of the dataset. In contrast, rarer types such as Ru Kiln Porcelain and Langyao Red Porcelain contain fewer than 100 images, each contributing less than 1.5% of the total. To mitigate potential issues related to class imbalance that may affect classification and regression stability, a set of data augmentation strategies, including image rotation, flipping, and brightness variation, was applied to underrepresented categories. This brought all classes closer to a more uniform distribution during model training. A detailed overview of both raw and augmented class distributions is provided, as shown in Supplementary Table 2.

Images were sourced from the following three channels. First, Auction houses (42.6%, 3500 images), including Christie's, Sotheby's, Bonhams, China Guardian, Poly Auction, and Beijing Rongbaozhai. Second, museums and cultural heritage databases (24.3%, 2000 images), such as the Palace Museum, the British Museum, the Metropolitan Museum of Art, the National Museum of China, and the ICOM database. Third, ceramic art stores and field photography (33.1%, 2713 images), obtained from platforms such as Taobao, Xianyu, Amazon, Pixabay, Wikimedia Commons, and private collectors. A hybrid data acquisition approach was adopted, for example, automated web scraping was performed using the Scrapy framework to extract structured ceramic image data from public databases. Meanwhile, manual photography was conducted in collaboration with ceramic experts and photographers to capture high-definition, high-value, and rare ceramic artifacts. All images were acquired in strict compliance with copyright regulations and are intended solely for academic research purposes.

Challenges encountered in data collection and corresponding solutions:

- (1) Issues with light reflection and shadow → Applied bilateral filtering for noise reduction, effectively preserving edge details while minimizing noise.
- (2) Interference from complex backgrounds → Utilized background segmentation algorithms, such as GrabCut, to remove distracting elements.
- (3) Inconsistent image resolution → Standardized all images to 1024×1024 pixels to ensure uniform model input and maintain data consistency.

For data annotation, this study adopted a hybrid approach combining AI pre-annotation and expert correction:

- (1) AI pre-annotation: a YOLO pre-trained model was utilized for initial object detection, automatically generating bounding boxes for ceramic contours.
- (2) Manual annotation and verification: a team of ceramic appraisal experts and data scientists refined the annotations using the LabelImg tool, following a structured, expert-informed guideline. Annotation was conducted across three hierarchical levels.

First-level classification: craftsmanship styles (e.g., blue-and-white porcelain, famille rose, doucai) were labeled based on characteristic features such as overglaze techniques and historical production periods. Experts referred to standard typologies drawn from authoritative museum collections (e.g., the Palace Museum) and academic literature to ensure consistent classification. As shown in Fig. 2, each craftsmanship style is contextualized within its corresponding dynastic period, from the Tang Dynasty to the modern era, highlighting the evolution of ceramic aesthetics and kiln-specific innovations over time.

Second-level classification: vessel shapes (e.g., bottles, jars, plates, bowls, cups, pots) were defined according to neck-body proportion, base structure, and handle or spout presence. In cases of borderline shape types, consensus was reached through group review. Reference images were

compiled into an internal labeling handbook to guide decisions, as shown in Fig. 3.

Third-level classification: decorative patterns (e.g., plants, animals, landscapes, portraits, geometric designs), as shown in Fig. 4. Rather than labeling individual motifs (e.g., lotus, peony, dragon), each image was annotated at the category level, based on the most visually dominant pattern types present in the overall design. For example, if a vessel featured both floral and tiger elements, the image was annotated as “plant” and “animal.” This multi-label, category-level annotation strategy balances annotation efficiency with classification relevance, enabling the model to learn from the dominant stylistic features without requiring exhaustive fine-grained motif annotation.

- (4) Supplementary attributes: additional features such as color complexity (monochrome vs. polychrome), structural intricacy (simple, moderate, intricate), and estimated price range (low-end collectibles vs. high-value antiques) were also annotated to enhance dataset richness. Structural intricacy was defined based on a combination of part count, curvature complexity, and decorative layering.
- (5) Annotation consistency and quality control: to ensure annotation consistency, all team members underwent a calibration phase using 300 sample images. Inter-Annotator Agreement (IAA) was assessed throughout the process, achieving a final Cohen's Kappa coefficient of 0.91, indicating strong agreement and reliable label quality.

To ensure data quality and enhance model robustness, this study implemented a series of preprocessing techniques, including image preprocessing, bounding box optimization, price normalization, outlier detection, and feature extraction. The complete preprocessing formulas, parameter configurations, and empirical evaluation details are provided in the Supplementary Information.

First, the training dataset comprised 8213 images, which were partitioned into training (70%, 5749 images), validation (20%, 1642 images), and test (10%, 822 images) sets, ensuring balanced representation across the 20 defined craftsmanship styles. To further improve model robustness and evaluate generalization capability, 5-fold cross-validation was employed. Additionally, K-means clustering was used to optimize anchor box dimensions, thereby enhancing detection accuracy for ceramic objects of varying shapes. A cyclic learning rate scheduler was applied to stabilize gradient updates, while early stopping with model checkpointing was implemented to prevent overfitting, for example, training was terminated if validation loss failed to improve over 10 consecutive epochs. In each iteration, four folds (80%, 6571 images) were used for training and one fold (20%, 1642 images) for validation. Following data partitioning, images were normalized to the [0,1] range, filtered to reduce surface noise, and augmented using geometric (rotation, flipping, cropping) and color transformations (brightness adjustment, HSV conversion). To enhance the spatial generalization ability, Mosaic, GridMask, and MixUp image enhancement techniques were also applied for verification, as shown in Supplementary Table 1. Anchor box dimensions were optimized using K-means clustering to adapt to varying ceramic shapes, and non-maximum suppression (NMS) was employed to reduce redundant detection.

Second, for price data, outliers were detected and removed using the interquartile range (IQR) method, followed by log transformation to reduce skewness and Z-score normalization to account for cross-auction house variability. Key features, including color (GLCM), shape (Hu Moments), and decorative pattern encodings (one-hot), were extracted and reduced via Principal Component Analysis (PCA) to improve computational efficiency and maintain prediction accuracy.

Although these normalization techniques improved the numerical stability and convergence of the training process, they may introduce trade-offs. For instance, the log transformation compresses the scale of high-value artifacts, potentially reducing the model's sensitivity in distinguishing between upper-tier price categories. Similarly, Z-score normalization across auction houses may obscure house-specific pricing nuances, such as branding premiums or regional valuation patterns.

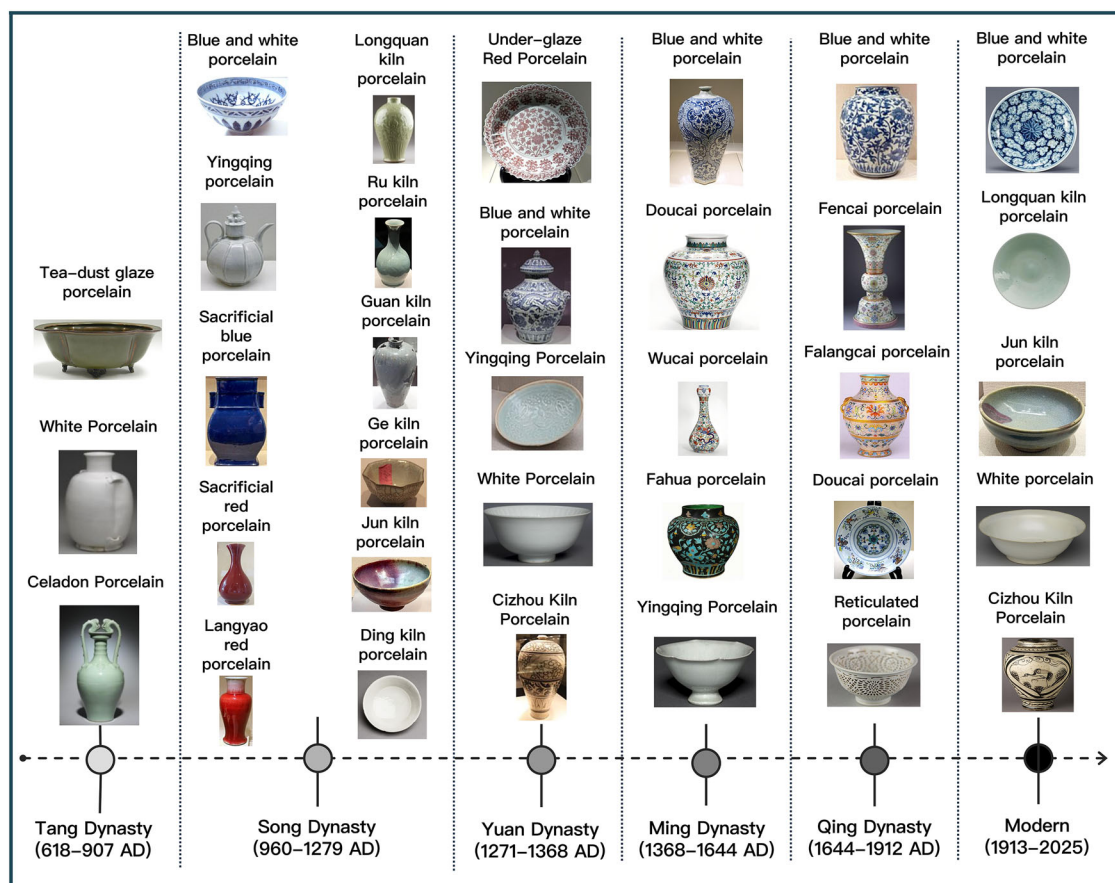


Fig. 2 | Craftsmanship styles classification. This figure visualizes 20 representative Chinese ceramic craftsmanship styles across seven historical periods, ranging from the Tang Dynasty (618–907 AD) to the Modern era (1913–2025). This figure

includes elements that were redrawn or adapted from copyright-free sources such as Wikimedia Commons and Pixabay, ensuring no copyrighted content is used.

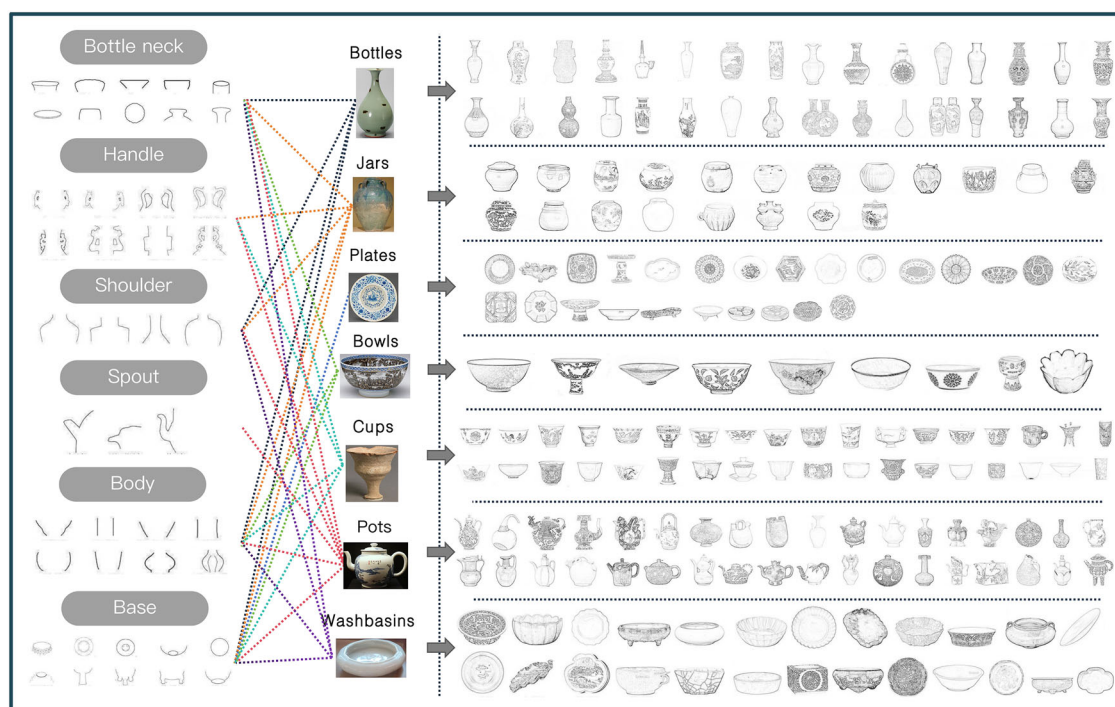


Fig. 3 | Shape classification. This figure illustrates the typological classification system used for ceramic vessel shapes, based on modular morphological parts. This figure includes elements that were redrawn or adapted from copyright-free sources such as Wikimedia Commons and Pixabay, ensuring no copyrighted content is used.

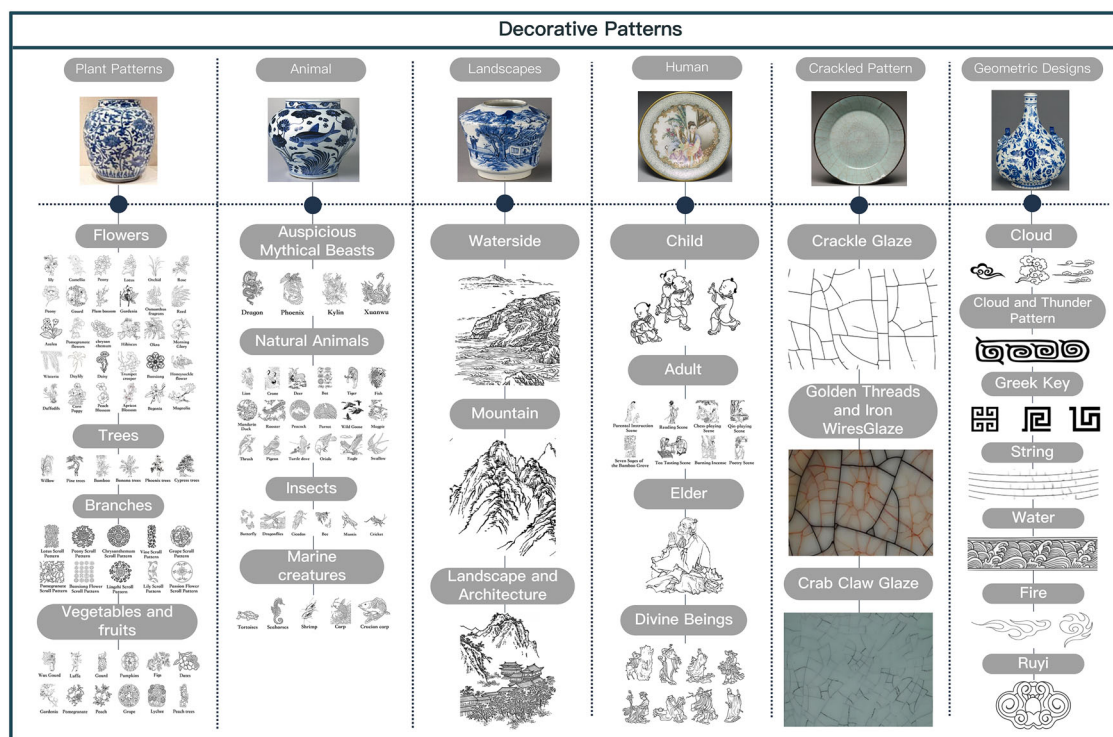


Fig. 4 | Decorative patterns classification. This figure presents a taxonomy of decorative patterns commonly found on Chinese ceramic artifacts, organized into six major categories: plant patterns, animal motifs, landscapes, human, cracked glaze patterns, and geometric designs. Each column shows representative visual

motifs and subtypes based on iconographic content, glaze texture, or symbolic form. This figure includes elements that were redrawn or adapted from copyright-free sources such as Wikimedia Commons and Pixabay, ensuring no copyrighted content is used.

These decisions were made to ensure model robustness and minimize the influence of extreme values and scale inconsistencies.

Building upon the curated and preprocessed dataset, the second stage of our pipeline deployed an improved YOLOv11 model to classify ceramic images based on their decorative patterns, structural forms, and craftsmanship features.

The classification of ceramic artifacts involves a comprehensive analysis of decorative patterns, shapes, and craftsmanship techniques. However, the complexity of surface textures, diversity of artistic styles, and intricacy of manufacturing details pose significant challenges to traditional classification methods. Conventional techniques based on handcrafted feature extraction and rule-driven algorithms often rely on low-level features (e.g., edges, color histograms, and shape descriptors) and fail to effectively capture the subtle decorative differences inherent in ceramics, thereby limiting classification accuracy and effectiveness. Additionally, studies utilizing machine learning methods such as Support Vector Machines (SVM), Random Forests (RF), or K-Nearest Neighbors (KNN) have achieved partial improvements in classification performance. However, these methods still face limitations when addressing the fine structures of high-resolution ceramic images.

Recent breakthroughs in deep learning, particularly Convolutional Neural Networks (CNN) and their extensions in object detection, have made automated ceramic classification feasible. YOLOv11 (You Only Look Once, Version 11), one of the most advanced real-time object detection models, capable of simultaneously detecting multiple ceramic attributes such as decorative patterns, object shapes, and production techniques within a single image. Compared to traditional CNN-based classifiers, YOLO significantly optimizes the object detection process by integrating object localization and classification into a single forward pass, thereby reducing inference time while maintaining high accuracy. The core advantages of YOLO in ceramic classification tasks include: (1) Real-time detection. YOLO possesses end-to-end object detection capabilities, making it well-suited for large-scale ceramic classification applications, such as museum digitization, online antique authentication, and automated

valuation systems. (2) Multi-object recognition. Ceramic artifacts often feature multiple decorative elements; for example, a single artifact may include floral motifs and geometric engravings. YOLO can simultaneously detect multiple categories, enhancing classification interpretability and robustness. (3) Efficient inference. YOLOv11 performs detection and classification within a single image simultaneously, offering greater computational efficiency compared to two-stage detection models such as Faster R-CNN and Mask R-CNN. This makes YOLO suitable for deployment on edge devices and mobile platforms.

For the improved YOLOv11 model architecture, it consists of three primary components: Backbone, Neck, and Head, each playing a critical role in the multi-object recognition process, as shown in Fig. 5.

To further optimize feature extraction efficiency and detection accuracy, this study integrates C3k2-EIEM (CSP with k2 convolution and Edge-Information Enhanced Module) into the YOLOv11 backbone, the C3k2-EIEM is presented, as shown in Supplementary Note 1. This enhancement improves edge detection, spatial feature retention, and overall classification performance.

The improved YOLOv11 employs ResNet50 as the backbone network to extract both low-level and high-level features from ceramic images. To improve feature representation and computational efficiency, this study introduces the following architectural enhancements:

- (1) C3k2-EIEM module: this module replaces conventional CSP (Cross-Stage Partial) blocks at the P3, P4, and P5 feature levels, enhancing the detection of fine-grained decorative details, such as engravings and inscriptions, by explicitly capturing edge information and preserving spatial details. This module consists of three key components: Edge Information Learning (SobelConv Branch), which integrates Sobel filters to refine edge detection, improving the recognition of object contours and decorative engravings; Spatial Feature Preservation (Conv Branch), which maintains high-resolution spatial details to ensure robust classification of intricate ceramic patterns; and a Feature

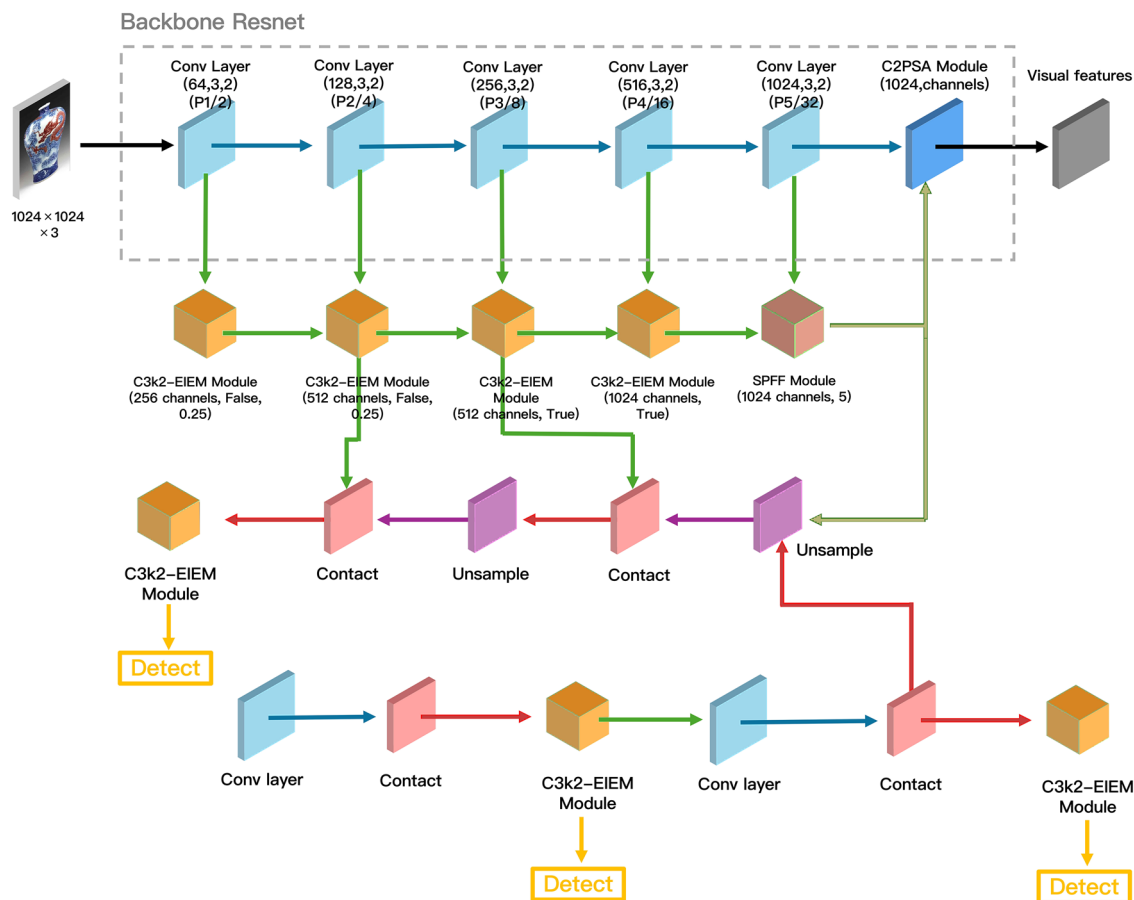


Fig. 5 | Improved YOLOv11 architecture. This figure illustrates the structural design of the improved YOLOv11 framework used for ceramic classification. The architecture integrates a ResNet backbone, multiple feature enhancement modules,

and optimized detection heads tailored for fine-grained ceramic attributes such as patterns, shapes, and craftsmanship styles. The entire figure was originally created by the authors.

Fusion Strategy, which combines edge-based and spatial-based features, resulting in a more comprehensive and discriminative object representation.

- (2) SPPF (Spatial Pyramid Pooling Fast): by performing multi-scale pooling, this module extracts ceramic object features at different scales, improving the detection of ceramics with varying sizes, such as plates, jars, and bowls.
- (3) C2PSA (Cross-Stage Partial Attention Mechanism): this module integrates channel attention and spatial attention to enhance the model's sensitivity to intricate decorative patterns, such as underglaze painting and hollow carvings. By adaptively adjusting weights, the model focuses more effectively on critical decorative regions, reducing background interference.

The Neck component is responsible for aggregating feature information from different levels and further optimizing the fusion of deep and shallow features. The improved YOLOv11 integrates a Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) in the Neck module. By combining the top-down feature propagation of FPN with the bottom-up feature enhancement of PAN, the model improves its ability to detect multi-scale ceramic objects. This ensures that the model maintains high precision when simultaneously recognizing large-scale objects (e.g., overall ceramic shapes) and small-scale decorative details (e.g., patterns and inscriptions).

The detection head (Head) of the improved YOLOv11 consists of three parallel output branches, each of which has been optimized to simultaneously detect large ceramic objects (e.g., vases, bowls, plates) as well as fine-grained features (e.g., floral patterns, geometric engravings,

calligraphic inscriptions) while suppressing background noise (reducing false positives and improving bounding box localization). Optimization strategies include improved convolutional layers (Conv, $k=3$, $s=2$) to enhance bounding box prediction accuracy and reduce detection errors, as well as the C3k2 detail enhancement module to strengthen the model's classification capability for decorative elements and improve its performance in complex backgrounds.

For the training of the improved YOLOv11 model, this study adopted its framework with architectural modifications tailored to the structural characteristics of ceramic artifacts. Specifically, the conventional Cross Stage Partial (CSP) blocks at feature pyramid levels P3, P4, and P5 were replaced with the C3k2-EIEM module. This module integrates edge enhancement, inter-scale feature fusion, and efficient spatial encoding, thereby improving the model's ability to capture fine-grained decorative patterns and subtle craftsmanship traits. The loss function followed the standard YOLO composition, incorporating cross-entropy loss for multi-class classification, generalized IoU loss for bounding box regression, and quality focal loss (QFL) to address class imbalance and emphasize difficult samples. These components were combined using fixed weighting, ensuring a balanced optimization of both classification and localization objectives.

To enhance generalization and model robustness, advanced data augmentation strategies were employed during training. These included Mosaic augmentation (merging four images to increase contextual diversity), MixUp augmentation (blending two images to produce soft-labeled samples), and GridMask augmentation (applying structured occlusions to encourage feature robustness under partial visibility). The optimal training

configuration consisted of a batch size of 16 and an initial learning rate of $1e-4$ with a scheduled decay. A total of 200 training epochs were conducted using the Adam optimizer, with momentum set at 0.9 to stabilize gradient updates and reduce training oscillations.

The evaluation metrics of research focused on measuring classification accuracy, detection precision, and model robustness across different ceramic craftsmanship style, shapes, and craftsmanship attributes. Additionally, ablation studies were conducted to analyze the interpretability of the model and its alignment with price prediction, particularly by examining how attention-enhanced modules contribute to the identification of high-value features in ceramic artifacts.

(1) Classification metrics (pattern, shape, craftsmanship style): to evaluate the categorization of ceramic attributes, the following metrics were used:

Accuracy measures the overall classification performance, reflecting the proportion of correctly classified ceramic attributes.

$$Accuracy = \frac{TruePositives(TP) + TrueNegatives(TN)}{Total\ Number\ of\ Samples} \quad (1)$$

Precision measures the proportion of correctly classified ceramic features among all predicted instances.

$$Precision = \frac{TruePositives(TP)}{TruePositives(TP) + FalseNegatives(FP)} \quad (2)$$

Recall evaluates the model's ability to correctly retrieve all relevant ceramic attributes.

$$Recall = \frac{TruePositives(TP)}{TruePositives(TP) + FalseNegatives(FN)} \quad (3)$$

The F1-Score is used as a balanced metric combining precision and recall, particularly suited for imbalanced ceramic categories. In this study, the Best-F1 score refers to the highest F1 value achieved across varying confidence thresholds (e.g., 0.0 to 1.0), with the optimal performance observed at a threshold of approximately 0.34.

$$F1\ Score = 2 \cdot \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

(2) Object detection metrics (Bounding Box Evaluation): since YOLOv11 performs both classification and object localization, it is essential to evaluate bounding box precision using the following metrics: Mean Average Precision at IoU 0.5 (mAP@50). This metric evaluates the model's object detection performance by measuring the average precision when predicted bounding boxes have at least 50% Intersection over Union (IoU) with ground truth, which reflects the model's ability to correctly detect and classify ceramic elements under a moderate localization threshold. Higher mAP@50 scores indicate better accuracy in identifying and localizing decorative patterns such as floral, geometric, or calligraphic motifs.

(3) Ablation studies and robustness analysis of model: we conducted a series of ablation studies to assess the individual and combined effects of attention mechanisms on the model's interpretability and price prediction accuracy. The classification performance of model was measured using accuracy, precision, recall, F1-score, and AUC, while a feature importance analysis was conducted to understand the key factors that influence the price estimation.

To support the interpretability goal of this research, particularly in bridging visual craftsmanship features and market valuation logic, this study integrates Gradient-weighted Class Activation Mapping (Grad-CAM) into the YOLOv11-based ceramic classification process. As a post-hoc visualization technique, Grad-CAM enables us to generate heatmaps that highlight the regions within ceramic images that activate the model's attention most strongly during classification. In this study, Grad-CAM serves two specific

purposes, first, to verify the effectiveness of the enhanced attention modules (C2PSA, C3k2-EIEM, and SobelConv) introduced into the YOLOv11 framework. By visualizing which image regions contribute most to the detection of decorative patterns, structural elements, or glaze details, the method confirms whether the model learns semantically meaningful patterns. Second, to establish an interpretability bridge between deep learning outputs and traditional expert valuation logic. The attention regions identified by Grad-CAM are cross-referenced with key features (e.g., manufacturing complexity, shape structure) used in Random Forest price prediction, confirming that the visual focus of the model aligns with empirically important price determinants.

In the third phase, we implemented a RF-based regression framework to predict market price categories based on extracted visual features. This component aimed to bridge the visual characteristics of ceramic artifacts with their appraised monetary value.

Valuing ceramic artifacts is a complex task influenced by multiple factors, including artistic craftsmanship, historical significance, and market trends. Traditional valuation methods rely heavily on expert assessments and historical auction data, which can introduce subjectivity and inconsistencies. To address these challenges, this study employs a regression-based machine learning approach to systematically predict the collectible value of ceramics.

Although ceramic prices are inherently continuous, this study adopts a classification-based prediction approach for practical and methodological reasons. First, auction price distributions are highly skewed and heavy-tailed, with a small number of exceptionally high-value items distorting regression outputs, as shown in Supplementary Table 3. Treating price as a continuous variable under these conditions often leads to poor generalization and unstable predictions, particularly for rare samples. Second, in real-world appraisal and auction settings, ceramic values are typically communicated in discrete price brackets (e.g., "less than \$10,000", "\$10,001–100,000", "\$100,001–500,000", "\$500,001–1,000,000", and "more than \$1,000,000"), rather than as precise numerical values. The selection of these five price brackets was informed by both empirical auction practice and exploratory data analysis. We surveyed historical ceramic auction catalogs from major auction houses (e.g., Sotheby's, Christie's, Poly Auction), where such price groupings are routinely used to segment market levels. To validate this segmentation, we conducted a quantile analysis on the training data distribution, which revealed natural inflection points aligning with these ranges. Alternative schemes, such as equal-width bins or quartile-based grouping, were also tested during pilot runs, but resulted in lower classification accuracy and higher misclassification between adjacent categories. The final five-tier segmentation thus balances real-world interpretability with statistical alignment to price distribution characteristics, supporting both communication clarity and model performance.

While this classification-based approach aligns with industry practices, it is important to acknowledge that discretizing continuous price data involves trade-offs. Specifically, the conversion of continuous prices into discrete brackets can reduce the granularity of prediction and obscure subtle value differences between adjacent price levels. This assumption was made to improve model stability and interpretability in real-world applications such as auction valuation. However, alternative methods, such as hybrid classification, regression models, or ordinal regression, could be explored in future work to preserve more fine-grained price information while maintaining classification robustness. Lastly, classification models allow for clear evaluation using confusion matrices and AUC scores, which offer intuitive insights into misclassification patterns across value levels. Future work may explore hybrid models that combine categorical classification with probabilistic regression for finer-grained valuation.

To build a reliable and interpretable ceramic price prediction model, it is essential to identify and structure features that capture both the artistic attributes of artifacts and the dynamics of the auction market. This section summarizes the feature engineering process into the following four points.

- (1) **Categorical features:** to construct an effective price prediction model for ceramic artifacts, key features influencing valuation were identified and structured into categorical and numerical factors, ensuring a comprehensive and data-driven approach to ceramic price estimation. The classification attributes of decorative patterns, craftsmanship style, and kiln origins are crucial for capturing the artistic and historical significance of ceramics. Specifically, decorative patterns include plant motifs, geometric patterns, mythical creature designs, human figures, landscape motifs, animal patterns, linear stripes, and glaze surface decorations. Craftsmanship style refers to the processes that impart color and decorative effects to ceramics through different glaze formulations, glazing methods, and firing techniques, including monochrome-glazed ceramics and multicolored-glazed ceramics. Additionally, kiln origins such as Ru, Guan, Ge, Ding, Jun kilns, as well as Jingdezhen, Longquan, Yaozhou, and Cizhou kilns are incorporated. These classification features are processed using One-Hot encoding to convert them into machine-readable numerical representations.
- (2) **Auction price normalization and outlier handling:** the auction data used in this study were collected from six major auction houses (Christie's, Sotheby's, Bonhams, China Guardian, Poly Auction, and Rongbaozhai) and span a temporal range from 2000 to 2024, covering over two decades of ceramic artifact transactions. All auction prices were converted and normalized to 2024 US dollars (USD) using historical exchange rates and Consumer Price Index (CPI) data published by international financial databases such as the World Bank and OECD. To ensure stable model training and reduce the effect of extreme values, auction price data were cleaned using the Interquartile Range (IQR) method, with any values lying below $Q1 - 1.5 \times IQR$ or above $Q3 + 1.5 \times IQR$ flagged as outliers and excluded. Additional duplicate removal and price format normalization were performed. After cleaning, the remaining dataset included 7812 valid price-labeled samples. A ceramics auction price range summary table is shown in Supplementary Table 3, showing a right-skewed distribution with most items valued between USD 10,001 and USD 100,000, and fewer high-end pieces exceeding USD 1 million.
- (3) **Quantitative features:** in addition to categorical attributes, quantitative factors were incorporated to capture the physical and market-driven influences on valuation. These include three independent variables—shape, decorative motifs, and manufacturing complexity—and one dependent variable, price range. Furthermore, historical auction data from Sotheby's, Christie's, Poly Auction, and China Guardian were analyzed to integrate market trends, ensuring that price estimations reflect real-world demand fluctuations. To maintain numerical consistency and enhance model performance, all quantitative features were normalized using Min-Max Scaling. By systematically combining both artistic characteristics and empirical market data, the proposed model establishes a robust and interpretable framework for the valuation of ceramic artifacts.

To evaluate the contribution of different features to price prediction, we conducted an initial correlation analysis and feature importance assessment using RF's built-in Gini importance ranking. Features with near-zero variance or strong collinearity (Pearson $r > 0.9$) were removed to reduce redundancy and mitigate overfitting risks. Furthermore, Principal Component Analysis (PCA) was applied to the normalized numerical feature space to improve computational efficiency. The first 10 principal components were retained, accounting for 92.7% of the total variance. This dimensionality reduction step ensured that the most informative aspects of shape complexity, glaze richness, and structural integrity were preserved, while reducing noise and irrelevant variations. Feature selection and encoding strategies were guided by both domain knowledge (e.g., auction expert feedback) and

empirical analysis of model performance under different combinations of features.

- (4) **Data encoding:** to ensure effective integration of categorical and numerical attributes into the regression-based price prediction model, a structured data encoding strategy was implemented. Categorical features, including decorative patterns, craftsmanship style, and shapes, were processed using OHE to transform discrete, non-numeric values into binary feature representations. This approach prevents the model from imposing ordinal relationships on inherently non-ordered attributes, ensuring that categories such as floral patterns, geometric patterns, dragon motifs, and glaze-based decorations are treated as independent variables. Similarly, craftsmanship style and shapes were encoded using OHE, allowing the regression model to capture stylistic and historical variations without introducing artificial numerical relationships. Meanwhile, numerical attributes such as physical dimensions, integrity scores, and market-based valuation factors were normalized using Min-Max Scaling, ensuring that all numerical values were rescaled to a standardized range of [0,1]. This preprocessing step prevents scale imbalances, stabilizes model convergence, and preserves the relative influence of different valuation factors, resulting in a robust and interpretable regression model for ceramic price prediction.

For regression model training, an RF classification model was employed, as shown in Fig. 6, which illustrates the architecture and workflow used to ceramic price prediction. The pipeline is structured into four key stages: data preprocessing, feature extraction, ensemble training, and prediction. The pipeline is structured into four key stages: data preprocessing, feature extraction, ensemble training, and prediction. The RF algorithm constructs multiple decision trees during training and outputs the mode of the classes for classification tasks, which effectively reduces overfitting and improves generalization. RF achieves this by combining bagging and random feature selection: each decision tree is trained on a bootstrap sample of the data, and at each split, a random subset of features is considered. This ensemble strategy increases model diversity, reduces variance, and avoids overfitting, which is useful for high-dimensional, mixed-type datasets such as ours, where features span both categorical and continuous domains. Moreover, RF inherently supports multiclass classification tasks and does not require feature scaling, which simplifies integration with one-hot encoded decorative attributes and numerical complexity indicators.

It should be noted, however, that while RF offers a favorable balance between predictive performance and interpretability, it is not the only viable option for price prediction tasks. Ensemble learning models such as XGBoost and LightGBM can provide enhanced accuracy, better handling of class imbalance, and finer control over overfitting through gradient boosting mechanisms. In this study, RF was deliberately chosen to prioritize transparency and explainability, key considerations in the context of cultural heritage valuation, where trust and interpretability are critical. Nevertheless, future research could conduct a systematic comparison of ensemble methods to determine whether performance gains from more complex models justify trade-offs in interpretability and computational cost.

The dataset was first divided into training and testing sets using an 80/20 hold-out method, ensuring that 80% of the data was used for training while the remaining 20% was reserved for testing. The training set consisted of both categorical features (e.g., decorative patterns, kiln origins) and quantitative variables (e.g., shape, decorative motifs, manufacturing process). Categorical features were converted into numerical representations using OHE, and quantitative features were normalized via Min-Max Scaling to maintain consistency and improve model performance.

The RF model was trained using MATLAB's TreeBagger function, which allows for flexible parameter tuning and efficient handling of large datasets. The training process involved iteratively adjusting key hyperparameters to optimize model performance: Number of Trees (numTrees):

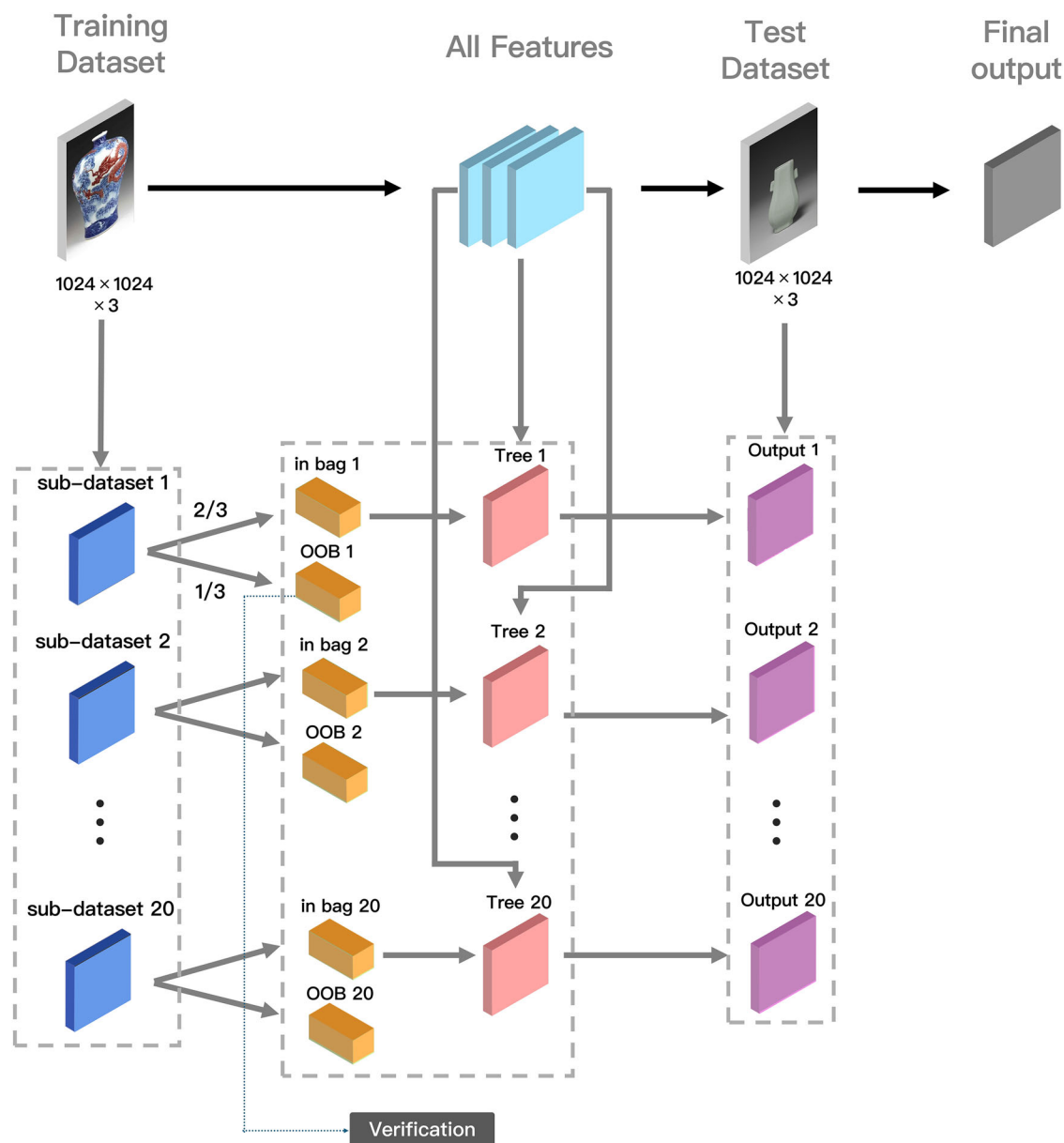


Fig. 6 | Random Forest architecture. This figure visualizes the full workflow of the Random Forest classification model used to predict ceramic price categories. The model integrates visual-semantic features extracted from improved YOLO 11 with

historical market data, structured into a supervised ensemble learning framework. The entire figure was originally created by the authors.

The number of trees in the forest was varied among 10, 50, 100, and 150. Increasing the number of trees generally improves performance but also increases computational cost. Maximum Depth (max depth): The maximum depth of each decision tree was set to 2, 5, 10, and 20, controlling the complexity of the model and preventing overfitting. Maximum Features (max features): The number of features considered for splitting at each node was tested using two strategies: one-third of the total features and the square root of the total features, aligning with standard practices for regression and classification tasks, respectively. Minimum Samples per Leaf (min samples leaf): This parameter was varied among 1, 5, 10, 20, 50, and 100 to control the minimum number of samples required at a leaf node. Smaller values tend to capture more intricate patterns, while larger values promote generalization.

After hyperparameter tuning via grid search, the final RF model, configured with 50 trees, a maximum depth of 50, two features per split, and a minimum of one sample per leaf, achieved a classification accuracy of 75.47% on the held-out testing set. This result represents the standalone

predictive performance of the optimized model when applied to unseen ceramic artifact data.

Following model training, feature importance was assessed using the out-of-bag (OOB) permuted predictor importance provided by the TreeBagger function. This analysis quantified the contribution of each feature to the model's predictive performance. The ranked feature importances were visualized using a bar plot, facilitating a clear interpretation of the factors influencing price predictions. By systematically tuning hyperparameters and analyzing feature contributions, the Random Forest model provided a robust and interpretable framework for ceramic price category prediction, integrating both artistic characteristics and market-driven data.

Although this study focuses on RF due to its balance of interpretability and performance, we conducted preliminary comparisons with Support Vector Machines (SVM) and logistic regression. These models yielded lower classification accuracies (SVM: 63.2%, Logistic Regression: 58.7%) and showed higher variance across folds.

Table 1 | The performance comparison

Model version	mAP@50 (%)	Recall (%)	Precision (%)	Best-F1 (conf \approx 0.34) (%)
YOLOv11	69	89	99	62
YOLOv11(improved)	70	91	98	64

Best-F1 is the maximal value along the confidence sweep.

Table 2 | The result of K-fold cross-validation

K-fold cross-validation result	mAP@50 (%)	Recall (%)	Precision (%)	F1 score (%)
1st discount	69	90	99	66
2nd discount	69	92	99	62
3rd discount	69	91	98	65
4th discount	70	92	99	63
5th discount	68	90	99	65

Best-F1 is the maximal value along the confidence sweep.

Results

Performance evaluation of YOLO-based ceramic classification

This subsection presents the quantitative performance evaluation of the optimized YOLOv11 model, comparing it against the baseline YOLOv11 to assess improvements in ceramic classification. The evaluation focuses on key object detection metrics, including mean Average Precision (mAP), Precision, Recall, and F1-score, measured across different ceramic attributes, such as decorative patterns, shapes, and craftsmanship techniques (Table 1).

To ensure a comprehensive and consistent evaluation, the dataset of 8213 annotated ceramic images was randomly divided into three subsets, 70% (5749 images) for training, 20% (1643 images) for validation, and 10% (821 images) for testing or performance comparison. The results demonstrate that the C3k2-EIEM-enhanced YOLOv11 model achieves consistent but modest box-level gains over the baseline detector. Specifically, mAP@50 increased by 1% (from 69% to 70%), indicating an enhanced overall detection precision for ceramic attributes. Although the 1% improvement in mAP@50 may seem marginal, it is consistent across validation folds and indicates greater reliability in fine-grained feature detection. Recall improved by 2% (from 89% to 91%), reducing false-negative detections of subtle shape or glaze details. Although precision slightly decreased by 1% (from 99% to 98%), this trade-off results in better model generalization by balancing false positives and false negatives. Additionally, the Best-F1 score (the highest F1 value obtained along the confidence sweep) increased from 62% to 64%, confirming a more balanced harmony between precision and recall at the detection layer. These results suggest that the improved YOLOv11 model offers enhanced feature representation in a controlled experimental setting. However, further validation in diverse real-world contexts is needed to confirm its broader applicability.

To ensure the robustness and generalization capability of the C3k2-EIEM-enhanced YOLOv11 model, a 5-Fold Cross-Validation was conducted, with the results summarized in Table 2. The mAP@50 values remained within a stable range of 68% to 70%, averaging 69%, indicating consistent detection performance across different validation splits. The Recall values ranged from 90% to 92%, demonstrating a high retrieval rate of relevant ceramic objects with minimal variance, while Precision consistently remained between 98% and 99%, confirming the model's high confidence in its predictions with few false positives. The Best-F1-score fluctuated between 62% and 66%, reflecting a moderate balance between precision and recall. Notably, the highest Best-F1 value (66%) was observed when recall was relatively lower (90%), suggesting that even with consistently high precision,

minor fluctuations in recall may significantly affect the harmonic mean due to class imbalance. These results indicate that the proposed enhancements likely contribute to more consistent classification performance under current dataset conditions. It is important to clarify the nature of the Best-F1 values, as shown in Table 2. These scores represent the model's highest harmonic mean of precision and recall observed near the confidence threshold of 0.34, rather than macro-averaged F1 across all classes. Although both precision and recall individually exceeded 90%, the corresponding Best-F1 values ranged from 62% to 66%. This outcome reflects the fact that F1 is highly sensitive to threshold settings: when the confidence threshold is set too low, recall may be inflated while precision suffers. When too high, the reverse occurs. The reported values indicate that at conf \approx 0.34, the model achieves its most balanced trade-off between retrieving relevant ceramic elements and minimizing false positives. This behavior is common in object detection settings and suggests that future optimization could benefit from adaptive thresholding or confidence calibration to further refine this balance.

Performance evaluation of random forest for price classification

The RF model was trained with the following optimal hyperparameters: 50 trees (numTrees), a maximum depth of 50, a maximum of 2 features per split (max features), and a minimum of 1 sample per leaf (min samples leaf). Under these settings, the highest overall accuracy achieved was 75.47%, indicating a reasonable level of performance for categorical price prediction under the given conditions. For the ceramic price classification task, the dataset was constructed based on cleaned and standardized auction data, resulting in 7812 valid samples with labeled price categories. The dataset was randomly split into two parts: 80% (6250 samples) training set was used to fit the RF classifier and perform a feature importance analysis. Twenty percent (1562 samples) test set served as an independent benchmark to evaluate the final classification performance, including accuracy, precision, recall, F1-score, and AUC.

On the training set, the model's accuracy reached 99.65%, with accuracy, recall and F1 scores all of which were 99.65%. On the independent test set, the model's accuracy was 98.91%, and the accuracy, recall and F1 score were also 98.91%. This consistency stems from the micro-averaging calculation method for all indicators, that is, true positives, false positives and false negatives are summarized in all categories, resulting in the same scores for each indicator when the prediction is highly accurate and the data in each category is balanced (after data enhancement).

To further evaluate the model's ability to distinguish between ceramic price categories, we generated Receiver Operating Characteristic (ROC) curves for the training set, test set, and the full dataset, as shown in Fig. 7. The Area Under the Curve (AUC) serves as a key metric for assessing the model's classification separability. The training set (red curve, AUC = 0.9965) and the all-sample curve (yellow, AUC = 0.9952) demonstrate near-perfect classification capacity, indicating strong internal pattern learning. The test set (blue curve, AUC = 0.9891) also shows high true positive rates across thresholds, confirming the model's generalizability to unseen ceramic artifacts. The close alignment between all three curves suggests that the model effectively captures price-related feature distributions without severe overfitting and performs robustly in predicting market value categories of ceramic vessels. According to the AUC result, a broad five price range is essentially easier than predicting precise continuous prices.

Fig. 7 | ROC curve analysis for price classification. This figure presents the ROC curve used to evaluate the classification performance of the Random Forest model in predicting ceramic artifact price categories. The ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR) across varying classification thresholds.

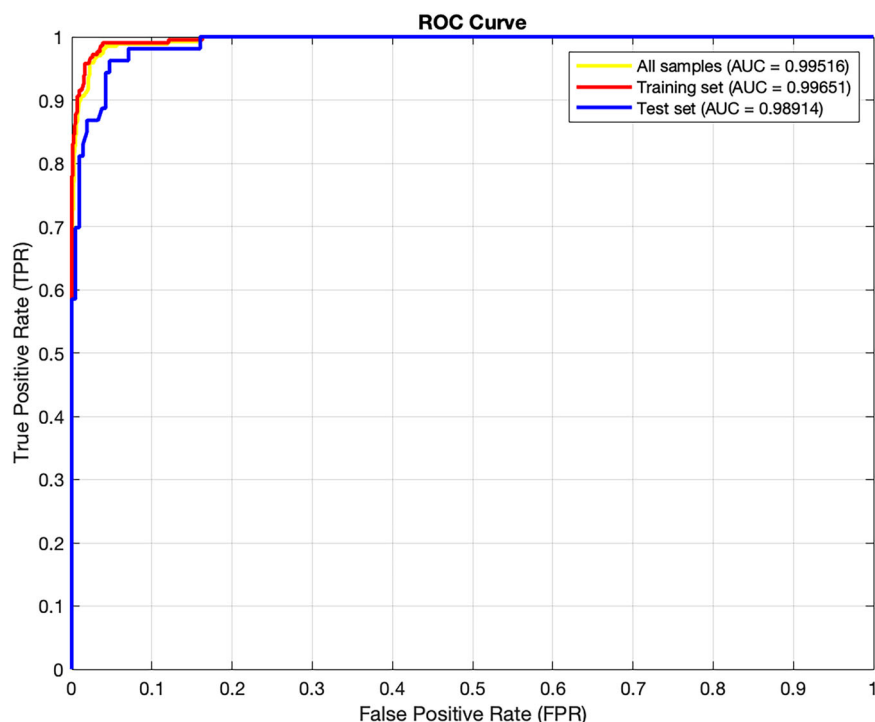
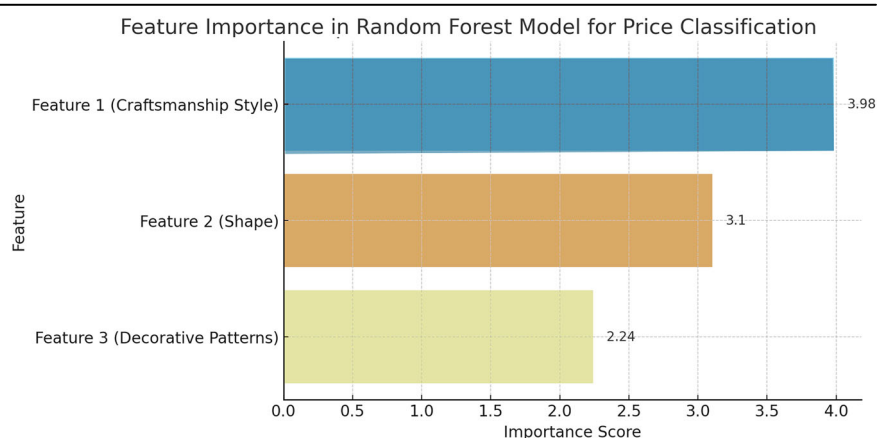


Fig. 8 | Feature importance in Random Forest model for price classification. This figure displays the relative contribution of three key features, such as Craftsmanship Style, Shape, and Decorative Patterns, to the price prediction task as determined by the Random Forest model.



Additionally, the feature importance ranking analysis revealed that Feature 1 (3.9826) was the most influential factor, followed by Feature 2 (3.1026) and Feature 3 (2.2396), as shown in Fig. 8. This suggests that manufacturing techniques (Feature 1) play the most significant role in determining ceramic prices, while shape (Feature 2) and decorative patterns (Feature 3) also contribute meaningfully. The ranking underscores that the complexity of craftsmanship significantly impacts market valuation, a finding that aligns with historical auction trends.

In general, the RF-based price prediction model shows promising predictive performance and useful feature interpretability, making it well-suited for real-world applications in ceramic valuation, auction market analysis, and automated appraisal systems. While RF offers high classification accuracy under current settings, its limited ability to separate adjacent price brackets and sensitivity to class imbalance may constrain its applicability in high-stakes valuation contexts. Future work should explore ordinal classification or probabilistic modeling to better handle fine-grained pricing tiers.

Case study: high-value artifact prediction

High-Value Artifact Prediction presents a comparison between the baseline RF (combine YOLO v11) model and the improved RF model (combine improved YOLO v11), focusing on a Song Dynasty Celadon vase.

In the first chart, as shown in Fig. 9, the baseline model incorrectly predicted the artifact as a Medium-Value Artifact, while the optimized model classified it as a High-Value Artifact, aligning with historical auction results. The baseline model (represented in orange) misclassified the vase due to underestimating the significance of inscription details. In contrast, the optimized model (blue) correctly identified the vase's high value possibly due to its limited sensitivity to fine-grained features such as inscription detail.

The second chart, as shown in Fig. 10, presents the confusion matrix of the optimized RF model, revealing detailed classification patterns across the three price categories. Most high-value artifacts were correctly predicted (60/65), with 5 misclassified as medium-value. Among the medium-value artifacts, 10 were misclassified as

Fig. 9 | Comparison of model predictions for Song Dynasty celadon vase. This figure presents a case study comparing prediction outcomes between the baseline model and the optimized deep learning framework for a Song Dynasty celadon vase.

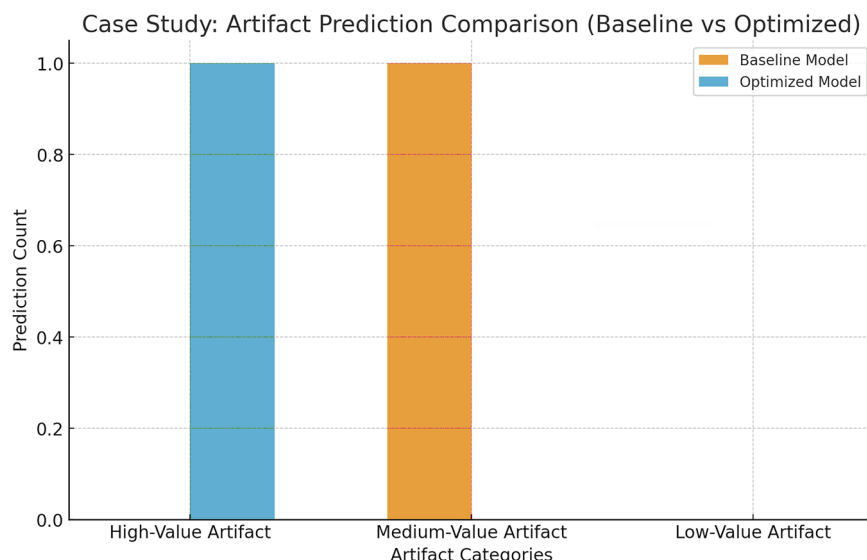
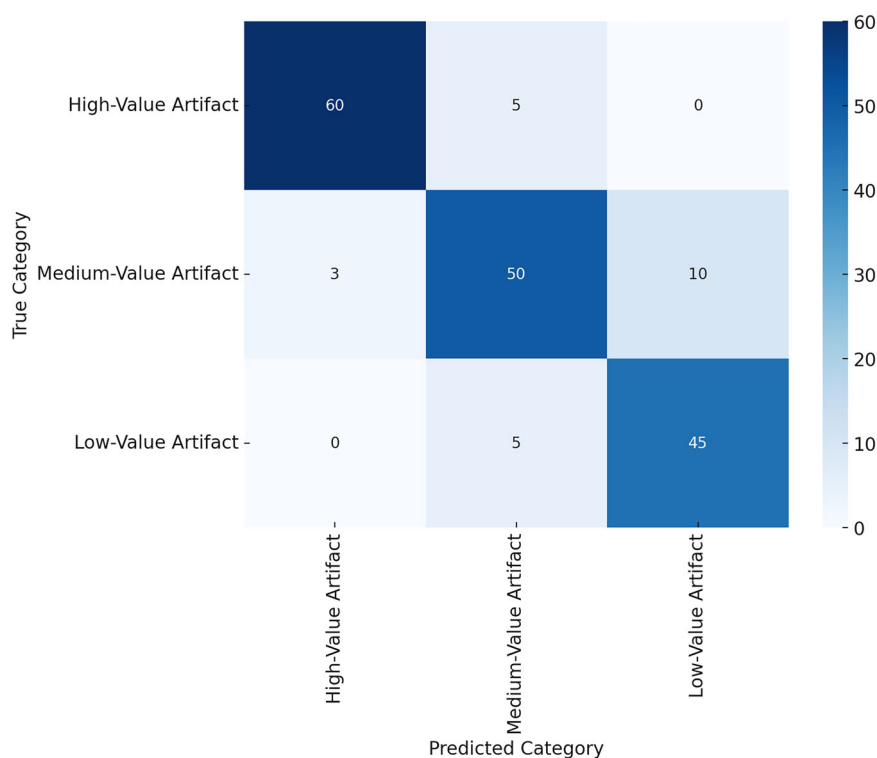


Fig. 10 | Confusion matrices for optimized models. This figure presents the confusion matrix summarizing the classification performance of optimized RF models across three artifact value categories, such as High-Value, Medium-Value, and Low-Value.



low-value, suggesting some overlap in feature distribution between these adjacent tiers. Notably, all low-value artifacts were either correctly classified or confused with the neighboring medium category, demonstrating clear model boundaries. The optimized model, benefiting from YOLO-enhanced feature embeddings, predicted the High-Value Artifact as shown by the accurate diagonal values in the confusion matrix.

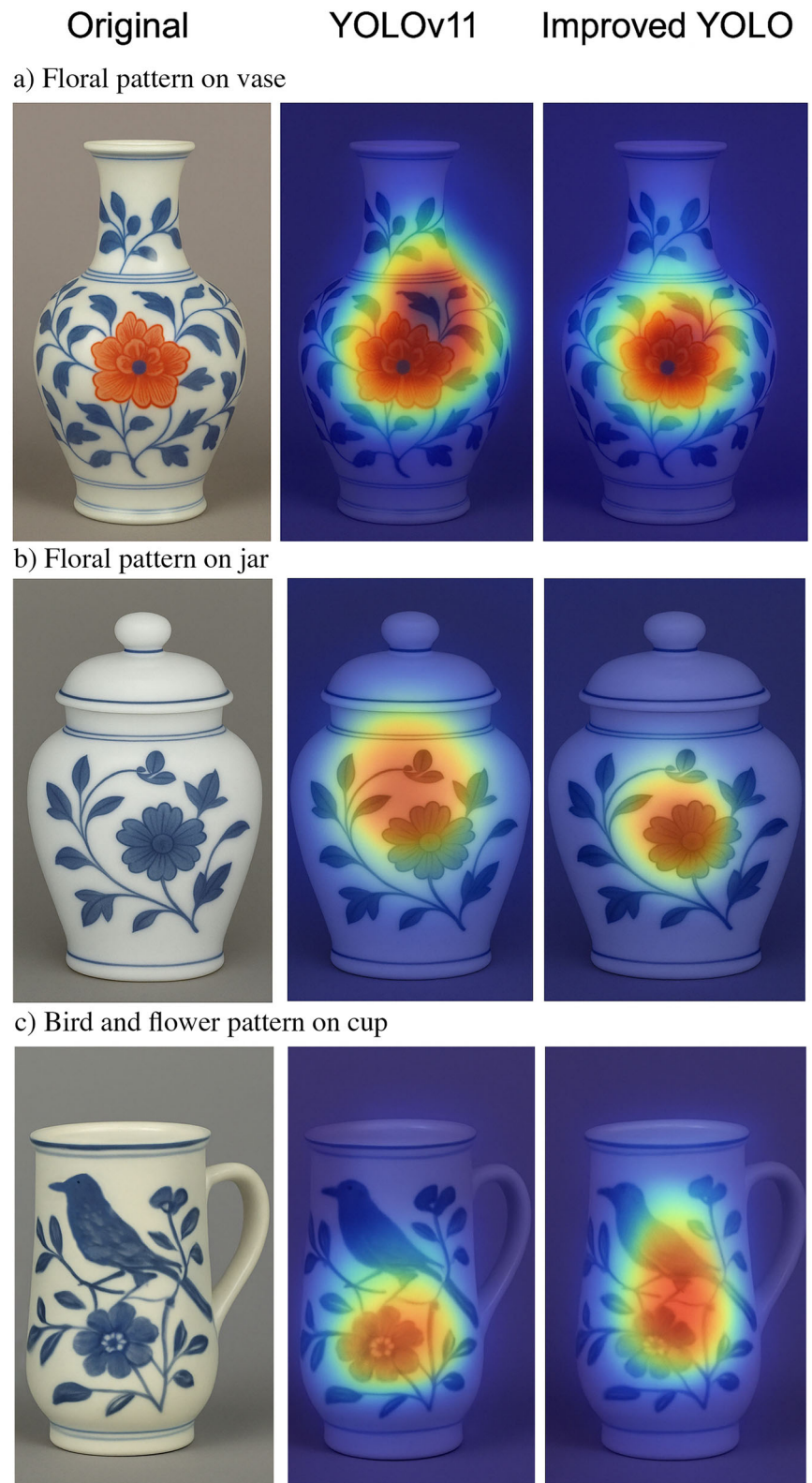
These findings suggest that integrating YOLO-extracted features may enhance the model's ability to recognize high-value artifacts, particularly those with complex decorative or inscription elements. However, as this case study is based on a single artifact, broader validation across diverse ceramic types and periods is necessary to confirm the generalizability of these observations.

Case study and error analysis: strengths and limitations of the model

To enhance the interpretability of the YOLOv11-based ceramic classification system, this study employed attention-based visualization techniques to identify which visual regions likely influenced the classification outcomes. Specifically, we applied Grad-CAM (Gradient-weighted Class Activation Mapping) to generate heatmaps over ceramic images, highlighting the regions most strongly activated by the attention layers.

As shown in Fig. 11, attention in both YOLOv11 and its improved variant tends to concentrate on intricate decorative zones such as floral engravings, avian motifs, and ornamental bands, depending on the object category. The heatmaps suggest that the

Fig. 11 | Grad-CAM attention visualization for ceramic artifact classification. This figure presents Grad-CAM heatmaps generated by YOLOv11 and the improved YOLO model to visualize attention regions across three ceramic object categories. Red to yellow hues indicate high attention focus areas, while blue regions reflect minimal attention. **a** For a vase with floral feature, both models focus on central motifs, but the improved YOLO yields more consistent and centered activation. **b** For a jar with floral feature, attention from the baseline YOLOv11 is slightly off-centered, whereas the improved model clearly aligns with the dominant flower feature. **c** For the bird and flower feature on a cup, the improved model demonstrates better attention spread across both symbolic elements. This figure is based on original fieldwork photographs taken by the author, it present Grad-CAM attention overlays generated by the author based on real-world imagery.



improved YOLO model, with C2PSA, C3k2-EIEM, and SobelConv modules, focuses more consistently on semantically meaningful regions compared to the baseline.

Furthermore, the Grad-CAM maps were cross-referenced with the top-ranking features in the RF model. For instance, samples identified with high manufacturing complexity in the RF model also displayed visual

attention concentrated on structural protrusions or rare color-glaze combinations in the YOLO heatmaps, as shown in Fig. 12. This preliminary alignment suggests coherence between the visual attention of the deep learning model and the feature importance in the RF classifier.

Beyond this alignment, we conducted a qualitative error analysis to investigate the model's limitations in recognizing visually degraded or

Fig. 12 | Visual attention on protrusions and glaze characteristics. This figure shows Grad-CAM heatmaps highlighting structural protrusions and glaze details in three ceramic artifacts. Red and yellow indicate regions of high attention, while blue represents low attention intensity. **a** In the reclining baby pillow, the model focuses on the curvature and facial relief, consistent with high structural complexity. **b** In the square legged ritual vessel, attention is concentrated on the central embossed feature and lower legs, highlighting both symmetrical relief and form. **c** In the lion footed censer, the model emphasizes the upper glaze layering and leg junctions, aligning with the complex construction and unique glazing style. This figure is based on original fieldwork photographs taken by the author, it present Grad-CAM attention overlays generated by the author based on real-world imagery.

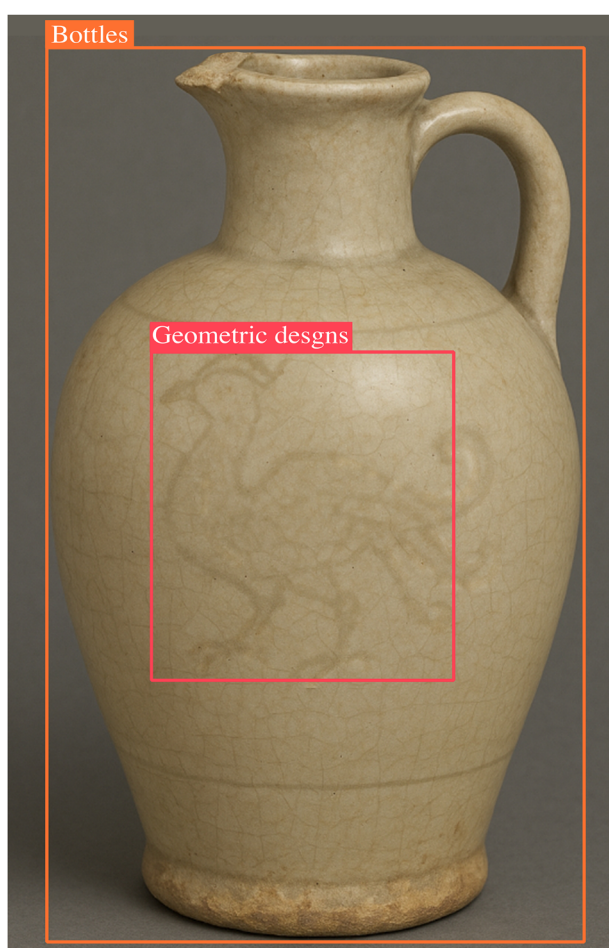
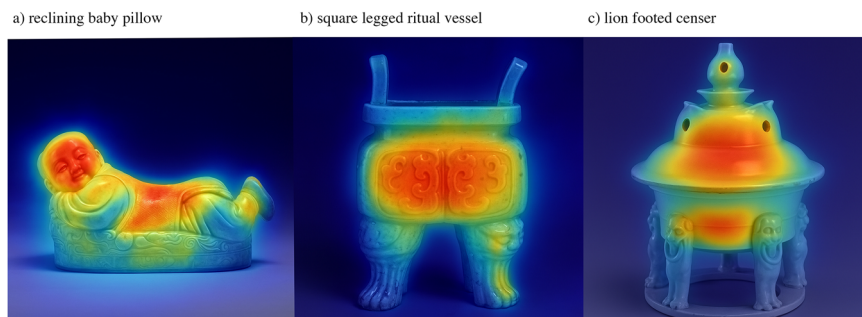


Fig. 13 | Blurred motif misclassified as geometric. This figure illustrates a case of misclassification by the optimized YOLOv11 model, where a blurred phoenix motif on a bottle was incorrectly identified as a geometric design. This figure is based on original fieldwork photographs taken by the author, its feature bounding boxes and classification annotations produced using the author's improved YOLOv11 model.

incomplete artifacts. This analysis revealed three representative types of failure cases: As shown in Fig. 13, when avian decorative motifs, such as a phoenix, appear faint and indistinct due to glaze degradation, aging, or low pattern contrast, the model may fail to classify the motif correctly. In this case, although the vessel clearly features a stylized bird in the center, the model misidentified it as having “geometric designs.” This misclassification



Fig. 14 | Broken contour misclassified shape. This figure presents a failure case in the object shape classification of improved YOLOv11, where the ceramic vessel with a damaged spout was incorrectly classified as a “Pot” instead of its correct category (e.g., “Bottle”) due to disrupted neck-body continuity. This figure is based on original fieldwork photographs taken by the author, its feature bounding boxes and classification annotations produced using the author's improved YOLOv11 model.

likely stems from the blurred edges and weak visual contrast between the motif and the background glaze, which confuse the detection module. This indicates a limitation in recognizing low-contrast features, especially in artifacts with worn glazes.

Artifacts with damaged rims or missing parts (e.g., chipped plates or broken necks) posed challenges in shape classification. As shown in Fig. 14, a bottle without a handle was mistaken for a pot. This suggests that the model's reliance on full-contour information limits its robustness against shape deformation caused by physical deterioration.



Fig. 15 | Decorative pattern not detected due to glaze peeling. This figure presents a failure case in the object shape classification of improved YOLOv11, where the model overlooked the decorative layers due to glaze peeling and uneven surface texture. This figure is based on original fieldwork photographs taken by the author, its feature bounding boxes and classification annotations produced using the author's improved YOLOv11 model.

As shown in Fig. 15, it shows an example where the model overlooked the decorative layers due to glaze peeling and uneven surface texture. A celadon vessel with glaze peeling failed to detect the original underglaze pattern.

Discussion

This study aims to bridge the gap between automated ceramic classification and price prediction, addressing key challenges within the field. The performance of the improved YOLOv11 model in the task of ceramic classification demonstrated significant improvements, particularly with the integration of C3k2-EIEM, SobelConv, and C2PSA enhancements. These modifications contributed to a slight but meaningful improvement in mAP@50 and recall, showing that the model's capacity to accurately detect ceramic features was enhanced. Although precision slightly dropped, this was an acceptable trade-off, as it led to a better balance between false positives and false negatives, improving the Best-F1 score from 62% to 64%. This indicates that the improved model not only provides higher detection accuracy but also demonstrates robustness in a variety of ceramic feature detection tasks. Further analysis using K-fold cross-validation consistently affirmed the model's stability and generalizability. The mAP@50 across all splits remained between 68% and 70%, supporting the model's consistency and generalization under controlled dataset conditions.

In the second part of the study, Random Forest was employed for price prediction. The RF classifier also demonstrated high performance, with an accuracy of 99.65% on the training set and 98.91% on the testing set. The AUC values of the ROC curves indicated that the model effectively captures

price-related feature distributions without severe overfitting, and performs robustly in predicting market value categories of ceramic vessels. The result of the confusion matrix of the improved RF model suggests that most high-value artifacts were correctly predicted. However, there is an overlap in feature distributions and confusion between categories in both medium-value and low-value artifacts. Despite these challenges, feature importance analysis revealed that manufacturing style (Feature 1) played the most significant role in determining ceramic prices, followed by shape (Feature 2) and decorative patterns (Feature 3). This aligns with market trends observed in historical auction data, further validating the model's relevance in real-world applications.

This study contributes a framework combining deep learning-based feature extraction with machine learning-based price prediction for ceramic artifacts. The improved YOLOv11 model effectively captures stylistic attributes, while the RF model highlights interpretable pricing factors. These findings underscore the potential of automated classification and valuation systems in the ceramic field.

Despite the encouraging results, the model also exhibited limitations in certain real-world scenarios. First, the current classification framework treats price categories as discrete labels, which may overlook ordinal relationships between tiers. This simplification can lead to misinterpretation of class boundaries and limits the model's sensitivity to price gradation. Second, the model's performance may be affected by class imbalance, especially when high-value samples are relatively scarce. The misclassifications observed in the confusion matrix support this, showing moderate confusion between mid- and low-value artifacts. Furthermore, its applicability to earlier artifacts or heavily degraded specimens (e.g., with severe surface erosion or incomplete form) remains limited, as these items often lack the visual clarity and structural completeness required by the YOLO framework.

Future work should improve model generalizability and interpretability by exploring ordinal regression or Bayesian probabilistic frameworks to capture soft boundaries between price brackets, while integrating multi-source market data, such as historical sales, artist provenance, and expert confidence levels, to support fine-grained, explainable valuation systems. These efforts could facilitate broader adoption by collectors, museum curators, and non-expert users through intuitive interfaces and simplified tools, particularly within digital museums, auction platforms, and cultural heritage investment contexts.

Data availability

The image data that support the findings of this study are available in figshare with the identifier <https://doi.org/10.6084/m9.figshare.29122634>.

Received: 1 March 2025; Accepted: 15 June 2025;

Published online: 28 June 2025

References

- Li, Z., Liu, F., Yang, W., Peng, S. & Zhou, J. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **33**, 6999–7019 (2022).
- Swanson, K. K. & Timothy, D. J. Souvenirs: Icons of meaning, commercialization and commoditization. *Tour. Manag.* **33**, 489–499 (2012).
- Ganiyeva, F. Porcelaine of eastern and western lifestyles: similarities and differences. *Vakan'uviz-Uluslararası Tarih Araştırmaları Dergisi* **8**, 1513–1531 (2023).
- Intelligence, H. Statistical analysis of China's daily ceramic production and growth trends from 2017 to 2023. <https://www.huaon.com/channel/trend/1004154.html> (2024).
- Research, G. V. Ceramics market size, share & trends analysis report. <https://www.grandviewresearch.com/industry-analysis/ceramics-market> (2021).
- Ling, Z., Delnevo, G., Salomoni, P. & Mirri, S. Findings on Machine Learning for Identification of Archaeological Ceramics: A Systematic Literature Review. *IEEE Access* **12**, 100167–100185 (2024).

7. Liu, Q. Technological innovation in the recognition process of Yaozhou Kiln ware patterns based on image classification. *Soft Comput.* <https://doi.org/10.1007/s00500-023-08528-8> (2023).
8. Anichini, F. et al. The automatic recognition of ceramics from only one photo: the ArchAIDE App. *J. Archaeol. Sci. Rep.* **36**, 102788 (2021).
9. Barreau, J.-B. et al. Photogrammetry based study of ceramics fragments. *Int. J. Herit. Digit. Era* **3**, 643–656 (2014).
10. Karasik, A. A complete, automatic procedure for pottery documentation and analysis. In *CVPRW 2010: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop* (IEEE, 2010). <https://doi.org/10.1109/CVPRW.2010.5543563>.
11. Zhao, Z., Ma, X., Shi, Y. & Yang, X. Multi-scale defect detection for plaid fabrics using scale sequence feature fusion and triple encoding. *Vis. Comput.* **41**, 1–17 (2024).
12. Gilboa, A., Karasik, A., Sharon, I. & Smilansky, U. Towards computerized typology and classification of ceramics. *J. Archaeol. Sci.* **31**, 681–694 (2004).
13. Lyons, M. Ceramic fabric classification of petrographic thin sections with deep learning. *J. Comput. Appl. Archaeol.* **4**, 188 (2021).
14. Wu, J. et al. Source and chronology analysis of ancient Chinese ceramic. *J. Chin. Ceram. Soc.* **S1**, 39–43 (2007).
15. Ren, Z. et al. Study of sauce glazed wares from Yaozhou Kilns (Northern Song Dynasty, 960–1127 ce): probing the morphology and structure of crystals in the glazes. *J. Eur. Ceram. Soc.* **42**, 7352–7359 (2022).
16. Chang, D., Ma, R., Zhang, L., Cui, J. & Liu, F. Characterizing the chemical composition of Tang Sancai wares from five Tang Dynasty Kiln sites. *Ceram. Int.* **46**, 4778–4785 (2020).
17. Chakraborty, S. et al. Rapid assessment of regional soil arsenic pollution risk via diffuse reflectance spectroscopy. *Geoderma* **289**, 72–81 (2017).
18. Sun, J. et al. Identification of porcelain ewers in tang, song, and yuan dynasties by digital shape characterization. *Ceram. Int.* **49**, 14246–14254 (2023).
19. Niu, C. & Zhang, M. Using image feature extraction to identification of ancient ceramics based on partial differential equation. *Adv. Math. Phys.* **2022**, 3276776 (2022).
20. Lucena, M., Fuertes, J. M., Martínez-Carrillo, A. L., Ruiz, A. & Carrascosa, F. Classification of archaeological pottery profiles using modal analysis. *Multimed. Tools Appl.* **76**, 21565–21577 (2017).
21. Lucena, M., Fuertes, J. M., Martínez-Carrillo, A. L., Ruiz, A. & Carrascosa, F. Efficient classification of Iberian ceramics using simplified curves. *J. Cult. Herit.* **19**, 538–543 (2016).
22. Lucena, M., Fuertes, J. M., Martínez-Carrillo, A. L., Carrascosa, F. & Ruiz, A. Decision support system for classifying archaeological pottery profiles based on mathematical morphology. *Multimed. Tools Appl.* **75**, 3677–3691 (2016).
23. Cardarelli, L. A deep variational convolutional autoencoder for unsupervised features extraction of ceramic profiles. A case study from central Italy. *J. Archaeol. Sci.* **144**, 105640 (2022).
24. Cintas, C. et al. Automatic feature extraction and classification of Iberian ceramics based on deep convolutional networks. *J. Cult. Herit.* **41**, 106–112 (2020).
25. Navarro, P. et al. Learning feature representation of Iberian ceramics with automatic classification models. *J. Cult. Herit.* **48**, 65–73 (2021).
26. Li, R. et al. Lbcapsnet: a lightweight balanced capsule framework for image classification of porcelain fragments. *Herit. Sci.* **12**, 133 (2024).
27. Mu, T., Wang, F., Wang, X. & Luo, H. Research on ancient ceramic identification by artificial intelligence. *Ceram. Int.* **45**, 18140–18146 (2019).
28. Wan, G., Fang, H., Wang, D., Yan, J. & Xie, B. Ceramic tile surface defect detection based on deep learning. *Ceram. Int.* **48**, 11085–11093 (2022).
29. Yi, J. H., Lee, H. & Kim, S. An analysis of the appearance characteristics of Korean ceramics per era through statistical analysis of metadata annotated with a visual element classification system of ceramics. *Herit. Sci.* **10**, 52 (2022).
30. Teddy, S., Nikolas, S. & James, P. Computer vision classification of pottery fragments using bag of visual words. *J. Archaeol. Sci. Rep.* **3**, 58–66 (2015).
31. Piccoli, C. et al. Archeomatica project: a new approach for the study of pottery in archaeology. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **II-5/W3**, 213–218 (2015).
32. Zhou, P. & Wang, K. Porcelain image classification based on semi-supervised mean shift clustering. In *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, 791–797 (IEEE, 2017).
33. Cui, T., Niu, R., Fang, Z. & Fang, Y. Ceramic art based on digital technology image processing. *J. Image Process. Theory Appl.* **7**, 32–42 (2024).
34. Chaowalit, O. & Kuntitan, P. Using deep learning for the image recognition of motifs on the center of Sukhothai ceramics. *Curr. Appl. Sci. Technol.* **22**, 1–15 (2021).
35. Santos, J. et al. Automatic ceramic identification using machine learning. Lusitanian amphorae and faience. Two Portuguese case studies. *Sci. Technol. Archaeol. Res.* **10**, e2343214 (2024).
36. Abbas, S., Moumen, H. & Abbas, F. Efficient method using attention based convolutional neural networks for ceramic tiles defect classification. *Rev. d'Intelligence Artificielle* **37**, 53–62 (2023).
37. Huynh, N. T. A multi-subpopulation genetic algorithm-based CNN approach for ceramic tile defects classification. *J. Intell. Manuf.* **35**, 1781–1792 (2024).
38. Mei, D., Lu, L., Chen, W. & Cheng, Y. Study on the classification of Chinese glazed pagodas. *Buildings* **14**, 4084 (2024).
39. Chubarov, V. M. et al. Possibilities and limitations of various x-ray fluorescence techniques in studying the chemical composition of ancient ceramics. *J. Anal. Chem.* **79**, 262–272 (2024).
40. Galli, A., Sibilia, E. & Martini, M. Ceramic chronology by luminescence dating: how and when it is possible to date ceramic artefacts. *Archaeol. Anthropol. Sci.* **12**, 1–15 (2020).
41. Montana, G. Ceramic raw materials: how to recognize them and locate the supply basins—mineralogy, petrography. *Archaeol. Anthropol. Sci.* **12**, 175 (2020).
42. Zhan, T., Zeng, L., Xu, W., Xu, W. & Wang, Y. Statistical model on impacting the value of ancient ceramics artworks. In *2017 International Conference on Robots & Intelligent System (ICRIS)*, 298–301 (IEEE, 2017).
43. Clunas, C. *Superfluous Things: Material Culture and Social Status in Early Modern China* (University of Hawaii Press, 1993).
44. Vickers, M. & Gill, D. W. J. *Artful Crafts: Ancient Greek Silverware and Pottery* (Oxford University Press, 1994).
45. Finlay, R. *The Pilgrim Art: Cultures of Porcelain in World History*, Vol. 11 (University of California Press, 2010).
46. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* **25**, 1097–1105 (2012).
47. Crook, P. Approaching the archaeology of value: a view from the modern world. *Post Mediev. Archaeol.* **53**, 1–20 (2019).

Author contributions

Y.H. designed the study, conducted the deep learning experiments, and drafted the manuscript. S.W. provided methodological guidance and supervised the machine learning framework. Z.M. and S.C. collected the auction dataset and assisted in feature engineering. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at

<https://doi.org/10.1038/s40494-025-01886-6>.

Correspondence and requests for materials should be addressed to Siqi Wu.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025