

Conformal Prediction Improves Model Reliability in the Prediction of Art Auction Price

Tekkyu Lee¹, David Lee²

Received February 9, 2026

Accepted June 27, 2026

Electronic access August 31, 2026

Predicting the price of art auction is an important question in finance and socio-cultural analytics because of surging interest in art valuation, decision-making for investment, and risk assessment for art market. Recently, deep learning technology has enabled prediction of valuation tasks for art auction prices based on image-related features. However, existing methodologies are limited by low explainability and thus low reliability leading to uncertain estimation. In this study, we propose an explainable deep learning framework for art auction price prediction by quantifying uncertainty in model output. We trained a convolutional neural network (CNN) for the task of classifying artwork prices into ten classes, which are based on the quantiles of auction prices. Moreover, we used SHapley Additive exPlanations (SHAP) to quantify the contributions of the features toward valuation outcomes. To quantify uncertainty, we used conformal prediction based on Adaptive Prediction Sets (APS) and generated statistically valid set of prediction outcomes with coverage guarantees. Our framework achieved high predictive performance, with mean classification accuracy of 0.798 ± 0.012 , AUROC of 0.876 ± 0.009 , and AUPRC of 0.912 ± 0.007 . Moreover, the coverage guarantee on the test set was established at $91.3\% \pm 1.2\%$ at a target confidence level of 90%, and with a mean prediction set size of 3.03 ± 0.21 . Coverage guarantee remained stable across price deciles, although uncertainty tended to increase in higher price categories due possibly to data sparsity. SHAP analysis showed that our model primarily used high-level visual representations such as compositional structure, texture, and aesthetic characteristics. Our findings suggest that combining explainability measure and uncertainty quantification can improve the reliability and interpretability of AI-based art auction price prediction. Our proposed framework demonstrates the potential of uncertainty-aware AI for art markets, particularly for art price estimation and valuation.

Introduction

Predicting art auction prices has been an interest in diverse domains (e.g., finance, art, or socio-cultural analysis) since it has potential to inform asset valuation, investment decision-making, and cultural heritage management¹. Unlike other types of assets, art works demonstrate high degree of heterogeneity, illiquidity, and subjective criteria for valuation, leading to complex price settling process, which remains often opaque². Traditionally, art valuation used statistical modeling methods such as hedonic regression³, which decomposed prices into hand-selected rules such as artist reputation, dimensions, medium, historical significance, and characteristics of auction house. Although such approach offered interpretability and insight⁴, such method assumed linearity between the rules and the price, which made it unable to capture the complex and nonlinear relationships, often found in modern art works⁵.

As the size and the quality of databases for art auction price improved, recent studies have applied machine learning, par-

ticularly deep learning, methods for the task of predicting art value⁶. Various machine learning methods (e.g., ensemble, gradient boosting, or decision tree algorithms) have shown high predictive performance as they are capable of modeling complex relationships between features and art prices⁷. Likewise, advancement in computer vision has enabled the representation of visual information from artwork images⁸. However, despite these, most existing models operate as black box models, leading to deterministic predictions without explanations or uncertainty quantification.

In fact, despite high predictive performance, the explainability and reliability (as an inverse of uncertainty) remain underexplored in the art valuation⁹. Previous studies often report performance metrics (e.g., mean absolute error or classification accuracy), but they tend to ignore interpretability measures and confidence in output¹⁰. This lack of attention is particularly concerning in art valuation, since the level of uncertainty and transaction values are often high.

To address the limitations, this study proposes a machine learning-based framework that combines convolutional neural networks (CNN), SHAP-based explainability assessment, and conformal prediction-based uncertainty quantification. First,

¹ Choate Rosemary Hall, USA

² The Prism School

we developed a CNN-based classifier which learns the representations pertaining to auction prices. Second, SHAP was used to estimate the contribution scores of the features toward auction prices, thereby ensuring model explainability. Third, we used conformal prediction based on Adaptive Prediction Sets (APS) to derive statistically valid prediction sets with coverage guarantees. Using this framework, we aimed to propose a systematic methodology to obtain calibrated prediction sets pertaining to each artwork valuation. This is especially important in the art domain characterized by high subjectivity, sparse data, and highly skewed price distribution, since our framework minimizes the risk of overconfident and unreliable prediction outcomes domains characterized by subjectivity, sparse high-value samples, and heavy-tailed price distributions. Our findings suggest that reliable uncertainty quantification and explainability can be achieved through this framework.

Methods

Dataset and Data Source

We used a publicly available benchmark dataset called `art_auction_valuation` (https://github.com/jasonshil0/art_auction_valuation). The dataset is curated for image-based art auction price prediction. The dataset contains over 5,000 artwork images obtained from public auction websites, each of which is annotated with a corresponding auction price. Moreover, each art sample in the dataset comes with a corresponding artwork image and an identifier. However, only limited metadata, such as artist and auction house information, are available for certain samples. Therefore, this study focused on features associated with images and mapped them to the target output, which is the art auction value and its uncertainty. We preprocessed the dataset by removing duplicate samples and corrupted images. Additionally, we removed samples with missing or invalid auction prices. Finally, only the samples incompatible with standardized RGB color channels were used. After preprocessing and filtering, approximately 5,000 valid artwork images remained for downstream analysis. Our final dataset has a size that is adequate for a computer vision dataset. Furthermore, the size of our dataset was comparable to that of prior studies in ML-based art price prediction. Lastly, since we defined the primary objective of this study not as developing a foundation model for art price prediction using self-supervised learning but as investigating a methodological feasibility of explainable prediction framework equipped with uncertainty quantification, the size of the dataset was considered adequate.

Data Preprocessing

We converted all image samples in our final dataset into standardized RGB format. We resized the formatted samples into a spatial resolution of 32×32 dimension matrix. We intentionally selected such low resolution matrix so that we could reduce the computational burden, estimation complexity, and maintain compatibility with a small-scale CNN architecture (which is commonly used in image-classification tasks). Such resizing removed artistic details with higher granularity. However, our exploratory data analysis with 64×64 dimension matrix produced only modest performance improvement while substantially increasing computational cost and runtime. Each artwork image is represented as a tensor, such that:

$$X'(i, j, c) = \frac{X(i, j, c) - \mu_c}{\sigma_c} \quad (1)$$

, where the tensor X for the three channels (i, j, c) were normalized by the mean and the standard deviation.

Next, we normalized pixel intensities using the range $[0, 1]$. We then standardized the intensities using CIFAR-10 channel statistics¹¹. In order to increase the size of training data, we augmented the samples in our final dataset by using random crops, horizontal contrast, and perturbations. The data augmentation was also used to reduce the risk of overfitting. On the other hand, the validation and test sets were left static without augmentation.

We modeled auction prices using classification task. In the classification task, the continuous auction price values were categorized into ten quantile bins based on decile thresholds. We formulated our task as classification also because such multi-label classification task is compatible with conformal prediction in providing prediction sets with coverage guarantees. Moreover, the task formulation as classification task was able to minimize the impact of outliers in obtaining balanced feature representations across all price brackets.

We split our final dataset into training and testing sets using an 80:20 split ratio. For conformal prediction, the training set was further split into 80% for training and 20% for validation or calibration. We used stratified splitting so that we could preserve balanced class distributions across all data splits.

Model Architecture

We used a convolutional neural network (CNN) as the base algorithm for auction price modeling. We used CNN because the algorithm offers a way to obtain representations using convolutional operations and to efficient dimensionality reduction through repeated pooling. We used the CNN architecture with an initial convolutional layer with 32 filters and a 3×3 kernel, followed by max pooling. This layer is then connected with two deeper convolutional layers each of which has 64

filters and undergoes pooling operations to reduce spatial dimensionality. We flattened the matrix output of the final convolutional layer into a dense representation vector, which was passed through fully connected layers (i.e., 64 hidden units using Rectified Linear Unit activation). We set the final linear output layer (i.e., 10-dimensional layer), such that a softmax function value is used for auction price classification. We used this CNN architecture to enable hierarchical (and sequential) representation learning, with later layers focusing on obtaining high-level representations, such that:

$$H^{(l)} = \text{ReLU} \left(W^{(l)} * H^{(l-1)} + b^{(l)} \right) \quad (2)$$

, where the hidden representation is max pooled, such that:

$$P^{(l)} = \text{MaxPool} \left(H^{(l)} \right) \quad (3)$$

Moreover, the prediction was made such that:

$$\hat{y}_i = w^T u_i + b \quad (4)$$

, where the best output is selected based on the softmax value, such that:

$$p(y_i = k | X_i) = \frac{\exp(o_{i,k})}{\sum_{j=1}^{10} \exp(o_{i,j})} \quad (5)$$

The model was trained by minimizing the crossentropy loss:

$$L_{CE} = -\frac{1}{N} \sum_i \sum_k I(z_i = k) \log(p_{i,k}) \quad (6)$$

Model Training and Optimization

Our CNN model was trained using the Adam optimizer with learning rate initially set at 0.001 and batch size at 128. The primary objective of the classification task was optimization (or minimization) of multi-label cross-entropy loss function. We performed model validation during training after each epoch of training so that we could training progress and learning convergence. We trained the model for 200 epochs with early stopping condition based on loss convergence. We found that training and validation losses consistently reduced during early phase of training, while stabilizing without loss peaking or divergence. In fact, training convergence was found to occur within 20 epochs. Moreover, our training results showed that there was no evidence of model overfitting or underfitting.

All experiments were conducted using an NVIDIA RTX 3090 GPU with 24 GB memory and an AMD Ryzen 9 processor. Average time for training was approximately 18 minutes per epoch, while average time for inference on the test set was approximately 4.1 ms per sample.

We evaluated the model performance using classification accuracy, area under the receiver operating characteristic curve (AUROC), area under the precision-recall curve

(AUPRC), macro-averaged precision, recall, F1-score, and Brier score. Specifically, precision, recall, and F1 score were calculated, such that:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (8)$$

$$F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

Explainability Analysis

We evaluated model explainability using SHapley Additive exPlanations (SHAP). The SHAP analysis was used to assess the contribution of each input feature toward an accurate prediction output. We computed SHAP values based on game theory, which provided theoretical background for accuracy and statistically grounded coverage guarantees. For each prediction output, we used SHAP to decompose the model output into a baseline prediction and contributions of individual features. This produced contribution scores that explain which features of an artwork image had the highest influence on prediction. Such scores enable the assessment of contributions of features such as brushstroke density, color contrast, and structural composition, which are learned by the model. Moreover, this enables inspection into model behavior and helps determine the extent of alignment with domain intuition, such that:

$$f(x) = \phi_0 + \sum_j \phi_j \quad (10)$$

where the contribution is defined as:

$$\phi_j = \sum_S \frac{|S|!(M - |S| - 1)!}{M!} [f(S \cup \{j\}) - f(S)] \quad (11)$$

Conformal Prediction

We quantified uncertainty in model output to ensure model reliability. To this end, we used split conformal prediction for our model. We first divided the original training dataset into two independent data splits: one for training set and the other for calibration (validation) set. Our model was trained on the training set, and the conformity scores were calculated on the calibration set, in which we estimated the distribution of prediction errors. For each sample in the calibration set, we obtained conformity scores from softmax outputs as one minus the estimated softmax value, such that lower conformity scores are assigned to confident predictions and higher scores to uncertain predictions.

Our conformal prediction framework estimates coverage guarantees, such that there is a certain probability that the

true label exists within the prediction set with probability at least $1 - \alpha$, independent of the underlying distribution or being model agnostic. By introducing such statistically rigorous framework (i.e., conformal prediction) to estimate coverage guarantee, we intended to provide an efficient uncertainty quantification for auction price prediction. We calculated the nonconformity score as thus the following:

$$S_i = 1 - p_{y_i}(x_i) \quad (12)$$

The outputs meeting the quantile condition were selected based on:

$$\hat{q}_{1-\alpha} = \inf \left\{ q : \frac{1}{|I|} \sum_i I(S_i \leq q) \geq 1 - \alpha \right\} \quad (13)$$

where α was set at 0.1, such that the prediction set has 90% chance of containing the true output.

Results

Dataset Characteristics

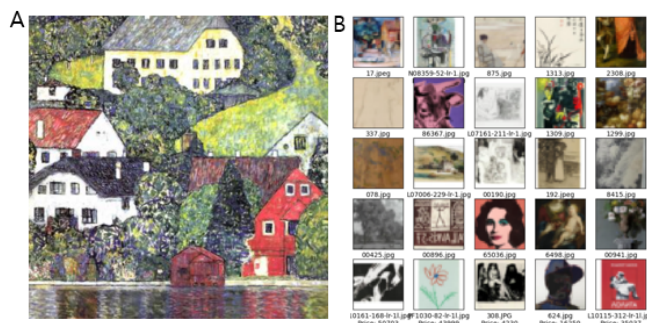


Figure. 1 Selected examples of art works in the dataset.

Selected examples of the samples in the dataset are summarized in Figure 1. The distribution of auction prices in the dataset is shown in Figure 2A. The distribution of auction prices is highly right-skewed (positively skewed), indicating that auction prices are clustered in relatively low-price ranges with sparse in higher values.

To mitigate distributional skewedness, we used a quantile-based categorization for the classification task. This strategy mitigated imbalance and ensured representations across price categories. This strategy also helped make sample sizes relatively homogeneous across bins, enabling conformal prediction for uncertainty quantification. We found that probabilistic modeling (e.g., using conformal prediction) is necessary, as deterministic point predictions (i.e., traditional neural network-based prediction) do not adequately reflect the uncertainty in model outputs, particularly those underrepresented in

the dataset. Also, as expected, we found that art work features were associated with auction price (Figure 2B).

Model Development and Performance

We found that model convergence was obtained at less than 20 epochs (Figure 3A). Of note, we did not observe any evidence of model overfitting (Figure 3A). The model demonstrated strong predictive accuracy (Figure 3B, 3C). We found that classification accuracy was 0.798, suggesting strong predictive performance of our model.

Furthermore, we found that AUROC was 0.876 (Figure 3B), indicating strong class discriminatory performance, while AUPRC was 0.912 (Figure 3C), suggesting favorable precision-recall trade-offs even under class imbalance. We also found that the Brier score was 0.142, suggesting that the model prediction was acceptably well calibrated, and that predicted class probabilities are well aligned with observed outcomes rather than being overconfident or underconfident.

In our class-wise analysis, we found that balanced predictive performance was achieved between the two categories. For class 0 (negative class), precision was 0.793, recall 0.769, and F1-score 0.781, while class 1 (positive class) exhibited precision of 0.802, recall of 0.824, and F1-score of 0.813. The slightly higher recall for class 1 suggests that the model is marginally more sensitive in detecting positive-class outcomes, whereas precision remains comparable in both categories. We found that macro F1-score was 0.797, and micro F1-score was 0.798, indicating stable performance regardless of the size of each class.

Lastly, of the 1,531 samples belonging to class 0 (negative class), our model correctly classified 1,177 instances, misclassifying 354 as class 1 (positive class). For class 1 (positive class), 1,438 of the 1,745 instances were correctly identified, with 307 false negatives. These results suggest that there was a slight asymmetry in classification behavior, with marginally better recall for the positive class. However, misclassification rates remained low in both groups, suggesting that our model does not systematically favor one class over the other.

Model Explainability

In our SHAP (SHapley Additive exPlanations) analysis results, we summarized the global feature importance and each feature's directional effect on estimated auction prices. The SHAP summary plot shows that as features are ranked by their mean absolute SHAP values, their overall influence on model predictions are well established (Figure 4). Of note, each point represents an individual artwork, and the color scale represents feature magnitude, with red color indicating higher values and blue color indicating lower values.

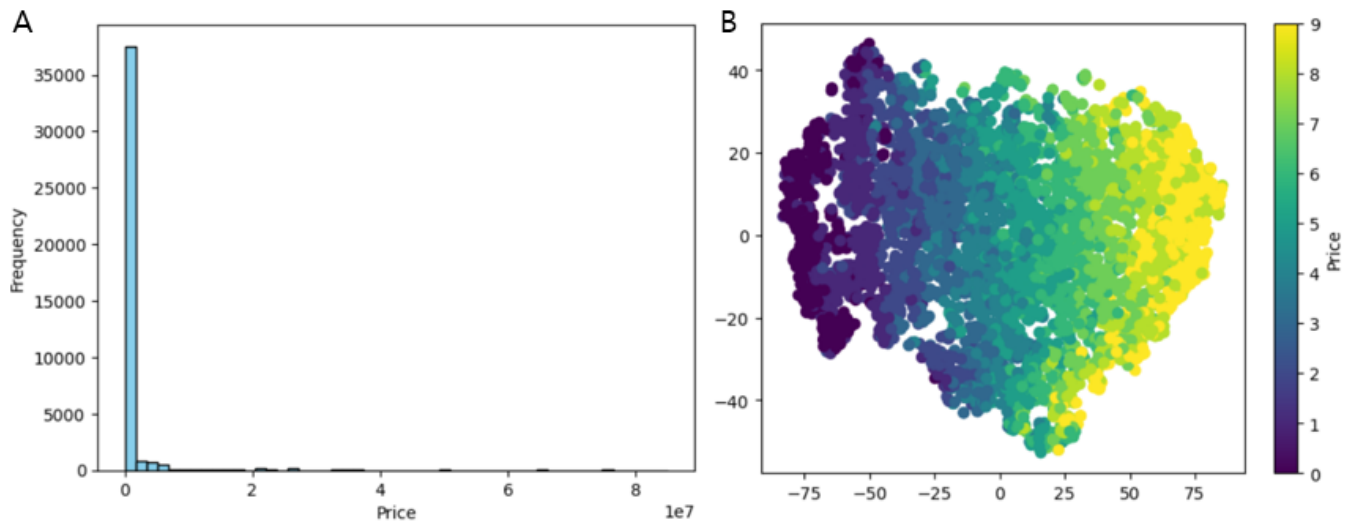


Figure. 2 A) the distribution of the price of the art works in the dataset. B) the 2 dimensional plot for art work clustering per their features for the prices.

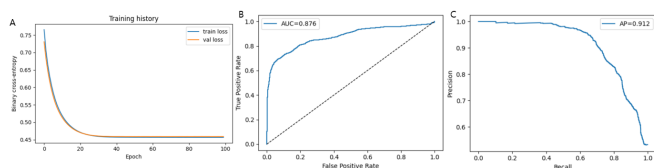


Figure. 3 A) Model training history. B) Receiver Operator Characteristics (ROC) curve. C) Precision-Recall Curve (PRC).

We found that the most influential feature was the first t-SNE embedding dimension (*tsne_1*), which substantially overcame all other inputs in terms of magnitude of impact and dispersion. High *tsne_1* values showed strong positive SHAP values, while low values were associated with negative contributions to predicted price. This indicates that our CNN model tends to model auction prices based on abstract visual patterns (representations) in the learned latent embedding space. These embeddings, despite being unable to explain, encode complex aesthetic, structural, or formatting characteristics of each artwork, such as composition, texture structure, color harmony. Moreover, it is highly probable that the embeddings encode semantic properties of artworks that are not directly observable in raw features. The strong separative performance by *tsne_1* suggests that high-level and complex learned representations outperform manually engineered features in determining auction prices. On the other hand, the second t-SNE dimension (*tsne_2*) had only marginal impact relative to *tsne_1*, indicating that valuation is primarily aligned along a dominant latent representation rather than a multi-dimensional representation with equal weighting.

Among the features associated with the images (i.e., those not learned embeddings by the model), the proportion of black pixels (*thresholdBlackPerc*) and overall brightness were among the most important features. Higher values of these features generally had higher SHAP values, suggesting that darker or high-contrast images were generally estimated at higher predicted auction prices. Moreover, edge density (*EdgePer*) and the ratio of unique colors further contributed to auction price prediction, suggesting that compositional complexity and visual richness were important in predicting auction prices.

Furthermore, temporal information, as encoded as the year, also demonstrated importance in prediction. Higher values (i.e., more recent creation dates) were associated with lower predicted auction prices, while older works were generally predicted to have higher auction prices. These findings are consistent with the current market behavior, in which historical or time-tested artworks tended to be better received. Also, brightness-related features were found to be important. For instance, a higher proportion of bright regions (*highbrightnessPerc*) was associated with higher auction prices, while a high proportion of low-brightness areas (*lowbrightnessPerc*) demonstrated no observable impact on determining auction prices.

Lastly, object- and structure-related features, such as width, height, corner density (*CornerPer*), and the number of human faces in the artwork were also important in determining (or predicting) auction prices. We found that the presence of identifiable subjects (e.g., human faces) alone was not a strong predictor of auction prices, whereas structural features such as image dimensions moderately contributed to predicting auc-

tion prices.

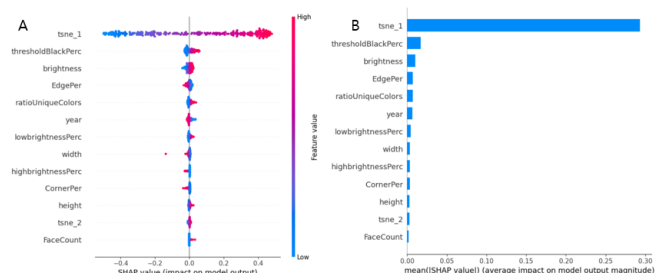


Figure. 4 Model explainability analysis results using Shapley Additive Explanation (SHAP).

Conformal Prediction Results and Model Reliability

In our analysis for uncertainty quantification using conformal prediction, we analyzed the distribution of normalized predictive entropy based on softmax class probabilities. We found that our model was relatively confident in correct prediction outputs, while the uncertainty increased (or confidence level) decreased for incorrect prediction outputs (Figure 5). Overall, the normalized entropy value was concentrated at lower values with relatively low degree of spread (Figure 5). Here, we used normalized entropy value as a surrogate metrics for model uncertainty (or lack of confidence), where higher normalized entropy value indicates higher uncertainty, or less confidence.

The distribution of normalized entropy showed that low entropy values were calculated for correct predictions, with the majority of correct predictions' normalized entropy values established below 0.05. This indicates that the model demonstrates high confidence when predicting auction prices correctly, assigning high posterior probabilities for the correct output. On the other hand, incorrect predictions are associated with relatively higher normalized entropy values with a much broader spread. The normalized entropy values of incorrect predictions generally exceeded 0.2 and extended to values above 0.4. This suggests that misclassifications (or incorrect predictions) are generally associated with uncertainty.

We found that the relationship between normalized entropy values and prediction set size from conformal prediction was monotonic and constant, with prediction sets with low normalized entropy values generally having one or two entries, while those with higher normalized entropy values leading to larger set sizes. This resizing tendency suggests that the conformal prediction is adequately working when the classifier model is predicting for uncertain outcomes, providing coverage guarantees only when applicable. Importantly, this behavior was achieved without any modification to the base classifier model we trained, underscoring conformal prediction being model agnostic.

In terms of model reliability, our results showed that our model is well calibrated. We found that there was a strong alignment between the normalized entropy values and the correctness of output. This indicates that our model has learned meaningful representations from the image inputs to predict auction prices with strong predictive performance rather than incorrect probability calculations. We also found that the conformal prediction enables the calculation of these confidence scores into prediction sets with coverage guarantee and the prevention of overconfident and unreliable predictions.

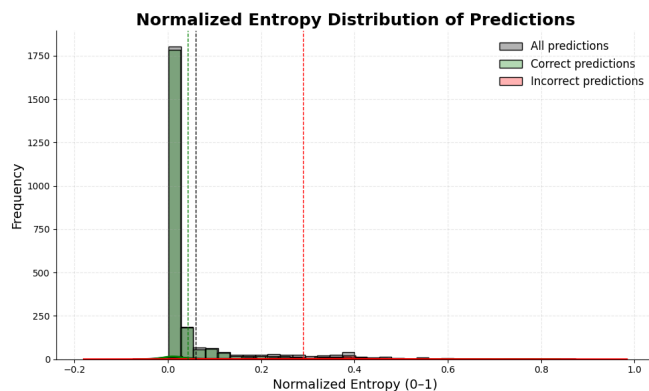


Figure. 5 Normalized entropy values for the model reliability for the correct and the wrong prediction results.

Discussion

In this study, we showed that incorporating uncertainty quantification and explainability assessment frameworks into art auction price prediction offered advantages over traditional prediction performance-driven modeling. We integrated conformal prediction and SHAP into a CNN-based auction price prediction model. Our results showed that we were not able to achieve strong predictive performance but also model the image features to predict auction prices in ways that are capable of providing calibrated estimates and interpretable outputs. These findings collectively indicate that reliability and transparency, which have previously been largely ignored in the art auction price prediction using ML, are both feasible and critical for improving model utility in high-stakes domains like art auctioning.

We showed that the observed auction price distribution was fundamentally heavy-tailed skewed, reflecting the reality of art markets¹², with a long trailing of right tail and a small number of highly priced artworks in the dataset. This data imbalance poses intrinsic challenges for modeling, as models trained on such imbalanced data tend to overweight, or sometimes ignore, underrepresented and rarely occurring samples. Our imbalanced data as evidenced in the histogram indicates

the need for methodology that does not rely only on point estimates based on the assumptions of Gaussian residuals. Instead, the evidence necessitates probabilistic and distribution-free approaches such as conformal prediction, which provided prediction sets with coverage guarantees without point estimates.

Despite the challenges of the problem (i.e., predicting for art auction prices due to high likelihood of involvement of confounders), our CNN model achieved strong classification and discriminatory performances. We showed that our modeling approach can reach an accuracy of approximately 80% and an AUROC of 0.876. Moreover, our model effectively differentiated between price categories using provided features alone without having to model confounders (whether observed or unobserved). Notably, the high AUPRC value of 0.912 further indicates our model's robustness in the presence of class imbalance. Similarly, the Brier score showed our model's predictions were reasonably calibrated, showing that the model softmax scores were not too arbitrary. Particularly, note that the precision and the recall were relatively symmetric, suggesting that the model did not systematically favor one class over another. This balanced model behavior is essential for downstream uncertainty modeling, since it is critical for conformal prediction to operate on well-calibrated probabilistic outputs.

Our SHAP analysis results showed that our model used learned embeddings for art auction price predictions more than other superficial image features. The feature importance analysis results showed that latent embeddings (tsne_1), which are high-dimensional representations capturing aesthetic, compositional, and semantic characteristics associated with each sample, were most important in predicting art auction prices. This finding aligns with our hypothesis that art auction price is more closely associated with abstract and complex stylistic and semantic cues combined rather than simple apparent attributes such as artworks' brightness or contrast alone. While superficial features related to the images contributed toward predicting art auction prices correctly, they were only marginal and secondary to the learned embeddings, indicating the effectiveness of deep representation learning for this task. Importantly, SHAP analysis results that the model's prediction was explainable and plausible, thereby minimizing the risk of shortcut learning based on spurious correlation relying on distantly related features.

Our uncertainty quantification analysis using normalized entropy values showed that there is a well observable separation between correct and incorrect predictions. Correct predictions were generally associated with lower normalized entropy values, while incorrect predictions were generally associated with higher normalized entropy values. This separation of normalized entropy values suggests that our model learned to predict art auction prices using the learned representations

instead of numerical artifacts. Such model behavior is critical, since deep learning models are often criticized for their tendency for overconfidence even for incorrect prediction outputs. Our normalized entropy analysis results minimize the risk of such concern.

Our conformal prediction analysis results showed that the statistically rigorous framework for uncertainty quantification can be effectively incorporated into standard machine learning (including deep learning) modeling work¹³. Our findings showed that the average prediction set size was 3.03, suggesting that our model showed relatively stable and usable prediction sets with coverage guarantees even at a high confidence interval (i.e., 90%). Notably, we found that the prediction set size increased as model uncertainty increased, suggesting an expected model behavior in response to uncertain samples. Compared to traditional point estimates (i.e., using softmax values) or Bayesian approximations that assume parametric noise distributions, our conformal prediction framework provided prediction sets with coverage guarantee without requiring assumptions about prior distributions or model retraining. This property of the framework is particularly important in the art markets, in which data-generating processes are often unknown, heterogeneous, and irregular^{14–17}.

In terms of applications of our model, our findings offer practical implications for auction houses, art collectors, and investment decision-makers. Instead of offering a single point estimate for each sample's prediction for art auction price, our proposed framework provides a prediction set with coverage guarantee (i.e., known error bounds) with explainable outcomes. This allows decision-makers to assess values and risks efficiently and informs bidding strategies¹⁸. Moreover, improved explainability of our model can enable users of our model to understand the aesthetic, compositional, structural, or semantic factors influencing predicted art auction prices, minimizing the risk of black-box phenomenon of ML models^{19,20}.

Our study has several limitations, which may suggest directions for future research. Our study used image data for modeling with only limited meta-data associated with each image sample. However, it must be acknowledged that the real-world pricing decisions often incorporate factors extraneous to the images' intrinsic qualities, such as provenance, history of the artists, market sentiment, and economic context. Integrating such heterogeneous information must be considered in future works, since they could serve as important confounders. Moreover, our study relied on unimodal data (images). It may be an interesting future research direction extending conformal prediction into a multimodal modeling framework using joint text–image multimodal representations. In addition, future work may explore quantile regression conformal intervals for continuous price prediction rather than classification tasks as we have defined in this study. Finally, larger and more di-

verse datasets may improve the generalizability of the prediction model and enable more granular uncertainty quantification.

In conclusion, our study showed that predictive performance alone is not sufficient to address the needs in the task for art auction price prediction. Explainability along with uncertainty quantification are critical to improving model utility by actual downstream uses in the art auction domain. By combining deep learning with SHAP and conformal prediction, we showed that our proposed framework contributes to the future of data-driven auction price prediction systems.

Conclusion

In this study, we showed that integrating explainability and uncertainty quantification improves the reliability and interpretability of ML-based art auction price prediction systems. By combining CNN with SHAP analysis and conformal prediction, our proposed framework was capable of obtaining useful representations and statistically rigorous uncertainty estimates in prediction outputs. Our findings suggest that predictive performance alone is insufficient for practical deployment of an ML model for art auction price prediction. Instead, a reliable uncertainty estimation and explainable model outputs are critical for informed decision-making, particularly in the domains characterized by subjectivity and sparsity of observations. Additional validations or future studies on larger multimodal datasets remain necessary. However, our study provides an evidence that uncertainty-aware and explainable modeling approach may offer a promising direction for next-generation art auction price prediction systems.

References

- 1 O. Ramašauskaitė, *Modeling investment strategies in contemporary art in the context of the Yugoslav wars*, 2024.
- 2 E. Coslor, *Transparency in an opaque market: Evaluative frictions between “thick” valuation and “thin” price data in the art market*, 2016.
- 3 W. Diewert, *Hedonic regressions: A review of some unresolved issues*, 2003.
- 4 W. Goetzmann, *Accounting for taste: Art and the financial markets over three centuries*, 1993.
- 5 A. Zorloni, *The Economics of Contemporary Art*, 2013.
- 6 C. Liu, *Prediction and analysis of artwork price based on deep neural network*, 2022.
- 7 D. Jang and M. Park, *Price determinant factors of artworks and prediction model based on machine learning*, 2019.
- 8 M. Nixon and A. Aguado, *Feature Extraction and Image Processing for Computer Vision*, 2019.
- 9 J. Allgaier et al., *How does the model make predictions? A systematic literature review on explainability in machine learning*, 2023.
- 10 A. Saranya and R. Subhashini, *A systematic review of explainable artificial intelligence models and applications*, 2023.
- 11 Y. Abouelnaga et al., *CIFAR-10: KNN-based ensemble of classifiers*, 2016.
- 12 Y. Romano, E. Patterson and E. Candès, *Classification with valid and adaptive coverage*, 2020.
- 13 M. Sadinle, J. Lei and L. Wasserman, *Least ambiguous set-valued classifiers with bounded error levels*, 2019.
- 14 H. Papadopoulos, *Inductive conformal prediction: Theory and application to neural networks*, 2008.
- 15 C. Guo, G. Pleiss, Y. Sun and K. Weinberger, *On calibration of modern neural networks*, 2017.
- 16 B. Lakshminarayanan, A. Pritzel and C. Blundell, *Simple and scalable predictive uncertainty estimation using deep ensembles*, 2017.
- 17 A. Angelopoulos, S. Bates, J. Malik and M. Jordan, *Uncertainty sets for image classifiers using conformal prediction*, 2021.
- 18 M. Ribeiro, S. Singh and C. Guestrin, *Why should I trust you? Explaining the predictions of any classifier*, 2016.
- 19 F. Doshi-Velez and B. Kim, *Towards a rigorous science of interpretable machine learning*, arXiv, 2017.
- 20 J. de Miranda Cardoso, J. Ying and D. Palomar, *Graphical models in heavy-tailed markets*, 2021.