International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 6099-6112

https://doi.org/10.70517/ijhsa463524

Color Analysis of Oil Paintings Combined with Decision Tree Algorithm and Its Practical Application in Art Teaching

Yuehai Wang[1] and Xiaoting Ren[2,*]
[1] Fine Arts Academy, Weinan Normal University, Weinan, Shaanxi, 714099, China
[2] Communist Youth League Committee, Weinan Normal University, Weinan, Shaanxi, 714099, China
Corresponding authors: (wnsy9963@163.com).

Abstract Scientific analysis of oil painting color by combining machine learning technology has become an important direction for the integration of art and technology, which provides new technical support for improving the quality of art teaching. This study combines the Gradient Boosted Decision Tree (GBDT) algorithm to deeply analyze the color features of oil paintings and explore its practical application in art teaching. Methodologically, the oil painting color prediction model based on GBDT is constructed, and the key factors affecting oil painting color are ranked in importance and visualized by SHAP interpretation method. At the same time, UV light-curing ink printing experiments were carried out on corrugated paper and gray-backed white board paper to establish the CIELAB color spatial relationship model for different paper substrates. The results show that the GBDT model outperforms the other six mainstream algorithms in terms of prediction accuracy, with an accuracy of 0.851 and an AUC value of 0.933. The SHAP analysis indicates that the color particle size, color type, canvas texture, pigment layer thickness, and the surrounding environment color are the top five key factors affecting the color of oil paintings. Printing experiments confirmed that there is a significant difference in the color rendering effect of ink on corrugated paper and gray background whiteboard paper, corrugated paper brightness is only 13% while the color deviation value is as high as 20%. The research results provide scientific support for the teaching of oil painting color in university art majors, and through the introduction of multimedia teaching methods and curriculum system reform, students' color perception and creative ability are effectively improved.

Index Terms Gradient boosting decision tree, Oil painting color analysis, SHAP interpretation method, Color teaching, Printing color presentation, Machine learning

I. Introduction

Color language has an important position in the art of oil painting, and the charm of color performance lies in the fact that emotions can be transmitted through color, so that the viewer can resonate with the painter, and the viewer can better appreciate the painter's spiritual world and emotional changes [1]. The degree of cognition generated by the subjective color cannot be separated from the continuous exploration and accumulation of artists, in different oil paintings, the subjective expression of color has different ways of embodiment, the painter through the objective object of the real color, according to their own thoughts and ideas will be transformed into color, so as to better express their own emotions [2], [3].

Traditional oil painting pays more attention to the realism of the painting object, ignoring the expression of the emotion of the painting itself. Many painters in the creation of oil paintings, only the pursuit of simple color tone warm and cold relationship, so that the painting itself can not fully reflect their own emotions and ideas [4], [5]. After the gradual development of oil painting color from Renaissance to Impressionism, the realistic color has been gradually improved in the continuous development. Many western painters detached oil painting color from the traditional way of creation, showing more natural effects, so that modern oil painting presents a strong emotional rhythm. Breaking down monochrome oil paintings into layers with distinct hues can produce subtle changes in the color gamut, thus showing bright colors [6]. After the creation of photographic technology, many painters brought out a more realistic color effect of the depicted objects by using color blocks and color juxtaposition [7]. Oil painting color from the Impressionists began to subjective expression gradually enhanced, the expressive nature of color has a very strong spiritual infectiousness, the continuous change of color hue levels, showing a clear contrast between warm and cold [8], [9]. Painters get various inspirations for color expression through the real feeling of actual objects.

Decision tree algorithms, a common class of supervised machine learning predictive models, are mapping relationships between object attributes and object values [10]. Applying the decision tree algorithm to the color analysis of oil paintings can simplify the color information into black, white, blue, yellow, red and other colors, which maximally retains the effective color language of oil paintings and greatly simplifies the original data [11], [12]. Related studies have shown that this algorithm greatly improves the contrast between the base color and the color of the oil painting, and improves the efficiency and accuracy of the oil painting color analysis [13].

As an important form of traditional western painting art, oil painting's unique color expression is the core feature that distinguishes it from other painting media. The formation of color in oil painting involves the complex interaction of multiple factors such as the physical properties of pigments, canvas materials, painting techniques, and environmental light, etc. This complexity makes it difficult for traditional empirical color analysis methods to meet the needs of modern art education and creation. In recent years, machine learning technology has made breakthrough progress in the fields of image processing and pattern recognition, providing a new technical path for oil painting color analysis. In particular, the advantages of decision tree-like algorithms in dealing with nonlinear relationships and high-dimensional features make them an ideal tool for analyzing the complex features of oil painting color. Currently, the teaching of oil painting color in the field of art education mainly relies on experience and subjective feelings, and lacks objective and quantitative means of analysis, which, to a certain extent, limits the students' in-depth understanding of the color laws and the enhancement of the ability to use them. At the same time, the use of color in artistic creation often relies on personal intuition, the lack of scientific theoretical guidance, resulting in greater randomness in the color control of the creative process. Introducing artificial intelligence technology into oil painting color analysis can not only reveal the color laws hidden behind art works, but also provide quantifiable and reproducible teaching tools for art teaching. In this study, the Gradient Boosted Decision Tree (GBDT) algorithm is used to construct a color prediction model for oil paintings, which establishes the mapping relationship between the color features and the final rendering effect through the comprehensive analysis of multi-dimensional features of oil paintings, such as physical attributes, painting techniques, and environmental factors. The study introduces the SHAP interpretable analysis method to deeply analyze the model prediction results and identify the key factors affecting the color of oil paintings and their contribution. At the same time, through the printing color experiment, to explore the influence of different substrates on the color performance mechanism. Finally, the research results are applied to the teaching practice of oil painting color in university art majors to improve students' color perception and creative ability through the reform of curriculum system and innovation of teaching methods.

II. Decision Tree Algorithm Based Color Prediction for Oil Paintings

II. A. ***Predictive modeling***

II. A. 1) **Decision trees**

Decision tree is a classical classification and regression algorithm in machine learning. In classification problems decision tree model is in the form of a tree structure. A classification problem requires a decision will start from the root node, after different internal nodes of the decision-making, will eventually get two different colors of the leaf nodes in one of the leaf nodes, the leaf nodes represent the classification problem of the two results. The composition of the decision tree is formed by the training data, which gives the decision tree new decision rules and finally realizes the classification effect of the classification problem [14].

For decision trees, as the tree splits, as the further down the division point the data set contains less information about unknown events, that is, the higher the purity, so that the final leaf node can more accurately respond to a category. A measure of the purity of a dataset in machine learning is the information entropy, the formula for which is shown in equation (1) below:

$$\mathrm{Ent}(D)=-\sum_{k=1}^{|y|}p_k\log_2 p_k \tag{1}$$

Equation (1) indicates that there are y categories in dataset D, and p_k refers to the proportion of the k th category samples in the dataset, and the higher the information entropy $Ent(D)$ calculated after the banding, the higher the purity of the dataset being divided.

After understanding the information entropy, the decision tree can select the optimal attribute at each division point through the information gain, and the formula of the information gain is shown in the following equation (2):

$$Gain(D,a)=Ent(D)-\sum_{m=1}^{m}\frac{|D^m|}{|D|}Ent(D^m) \tag{2}$$

In equation (2), a is the attribute of the node to be divided, and there are m different values of the attribute, calculate the information entropy of each value $Ent(D^m)$, weight and sum the information entropy of each value of the a attribute, and then subtract the weighted information entropy of the a attribute from the information entropy of the previous node, then we get the information gain of the a attribute. Comparing the information gain of each attribute on the node, the higher the information gain is the optimal attribute of the node, the typical ID3 decision tree algorithm is based on the information gain to divide the attributes.

II. A. 2) GBDT algorithm

Gradient Boosting Decision Tree (GBDT), an algorithm that combines gradient boosting algorithm with decision tree algorithm, is a typical integrated learning algorithm, which is composed of multiple weak learners linked together, which are decision tree learners, thus GBDT algorithm is a strong learner combining decision tree algorithms and integrated learning ideas [15].

However, the Boosting method of GBDT is different, it is an additive model composed of multiple CART decision regression trees, which can be simply understood as the final prediction of the whole model is the sum of the predictions of all the regression trees. Although GBDT is commonly used in regression problems, it can also be used in classification problems. The training process of GBDT classification problem is as follows: For a dataset, it is intended to use m decision tree to make decisions.

(1) First create the first tree $f_1(x)$, which is a priori information in a binary classification problem as in equation (3):

$$f_1(x) = \log\frac{p_1}{1-p_1} \tag{3}$$

(2) For the 2nd through m st regression trees, the training objective, which is the residual of the previous results, is calculated for each tree as in equation (4):

$$r_{mi} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f(x)=f_{m-1}(x)} = y_i - \hat{y}_i \tag{4}$$

For the current m st subtree, it is necessary to traverse its feasible cut points as well as thresholds to find the parameter corresponding to the optimal prediction c so that it approximates the residuals as much as possible, and the formula for the prediction c is shown in Equation (5):

$$c_{mj} = \arg\min_{c} \sum_{x_i \in R_{mj}} L(y_i, f_{m-1}(x_i) + c) \tag{5}$$

The R_{mj} in Eq. (5) refers to the set of predicted values of the leaf nodes in all the division methods of the m nd subtree, that is, the possible predicted values of the m rd tree, where j is in the range of $1, 2, 3, \ldots, J$. The optimal predicted value of the m th tree can be obtained and the m th tree can be updated on the $m-1$ th tree, as shown in Eq. (6):

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J} c_{mj} I(x \in R_{mj}) \tag{6}$$

(3) The final learner is obtained by adding up the predictions of m tree through an additive model, as shown in equation (7):

$$F_M(x) = f_M(x) = \sum_{m=1}^{M} f_m(x) = \sum_{m=1}^{M}\sum_{j=1}^{J} c_{mj} I(x \in R_{mj}) \tag{7}$$

This is the learner computed by regression, and for the classification problem an additional step is needed, which is to compute the probability that a particular sample category is 1 as shown in equation (8):

$$P(x_i) = \frac{1}{1+e^{-F_M(x_i)}} \tag{8}$$

II. B. ***SHAP interpretation method***

GBDT is an integrated learning model based on decision trees, which itself has a complex model structure, but its advantages and sophistication have been widely proved and accepted at the theoretical level, and it has yet to be tested in practice because of its weak interpretability. When we face a complex machine learning model like GBDT, mining the internal mechanism of model prediction can help us better understand the problem under study, the data we face, and the reasons why the model may fail. Only when the model prediction results are interpretable can we have more trust in the model, and the SHAP interpretation method can accomplish the interpretation of the GBDT model prediction results.

The SHAP interpretation method is both locally and globally interpretable. SHAP constructs a linear model based on the game-theoretically optimal Shapley value, which can be used to interpret the output of any machine learning model. For each sample, the machine learning model gives a predicted value, and SHAP treats all features as “contributors”, and the Shapley value is the value assigned to each feature in the sample [16].

We can understand this by assuming that the j nd feature of the i st sample takes the value of $x_{i,j}$, the predicted value of the machine learning model for the i th sample is $\hat{y}_i$, the base value of the model is ϕ_0, and the Shapley value of $x_{i,j}$ is ϕ_i, then the following equation holds:

$$\hat{y}_i = \phi_0 + \sum \phi_{i,j} = \phi_0 + \phi_{i,1} + \phi_{i,2} + \cdots + \phi_{i,m} \tag{9}$$

The basic idea of the Shapley value is to consider the mean value of the marginal contribution of a feature to the predicted value of the model when it is added to the model, the reason for the mean value is that for m feature we have 2^m ways of combining it and for all the ways of combining it, we need to calculate the marginal contribution of the feature to the predicted value of the model when it is added to the model.

SHAP specifies the interpretation in the following form:

$$f(x_i) = F(z') = \phi_0 + \sum^{m} \phi_{i,i} z'_{i,i} \tag{10}$$

where $f(x_i)$ is the predicted value of the machine learning model for sample x_i, F is the explanatory function, $\phi_{i,j} \in R$ is the Shapley value for sample i, the j th feature taking the value of $x_{i,j}$, and $z'_{i,j} \in \{0,1\}^m$ is the feature combination vector, where input 1 indicates that the corresponding feature value exists, and input 0 indicates that the corresponding feature value does not exist. In order to compute the Shapley value, it is necessary to simulate the situation where only some feature values exist and others do not, and this representation of the feature combination vector is the trick to compute $\phi_{i,j}$. If the j th feature is not in the current combination, $\phi_{i,j} = 0$, it means that the feature will not contribute to the sample predicted value at this time. For sample x_i, the Shapley value $\phi_{i,j}$ of the j th feature value $x_{i,j}$ is computed by satisfying the following equation:

$$\phi_{i,j} = \sum_{S \subseteq M \setminus \{x_{i,j}\}} \frac{|S|!(m-|S|-1)!}{m!} \left[f_{x_i}(S \cup \{x_{i,j}\}) - f_{x_i}(S) \right], j \geq 1 \tag{11}$$

where M is the set of all features in the dataset with dimension m. S is a subset drawn from M with size $|S|$. $f_{x_i}(S)$ is the predicted value of the model for the sample x_i when only the feature set S is used, and when S is the empty set, the value of $f_{x_i}(S)$ is called the base value ϕ_0, which corresponds to the average of the model's predicted value over all samples. $f_{x_i}(S \cup \{x_{i,j}\})$ is the average over all samples of the model's predicted value for the sample x_i when the feature value $x_{i,j}$ is added on top of the feature set S, and $\frac{|}{S}|!(m-|S|-1)!m!$ represents the weight of the difference between the values taken by $f_{xi}(S \cup \{x_{i,j}\})$ and $f_{xi}(S)$. Since there can be many scenarios for the value of S, it can be seen from the formula that the Shapley value $\phi_{i,j}$ of the eigenvalue $x_{i,j}$ is a combination of all possible values of s and the effect of features other than the eigenvalue $x_{i,j}$ on the eigenvalue $x_{i,j}$.

The following is a simple tree model to demonstrate the process of SHAP calculation.

Assume a decision tree as shown in Figure 1:

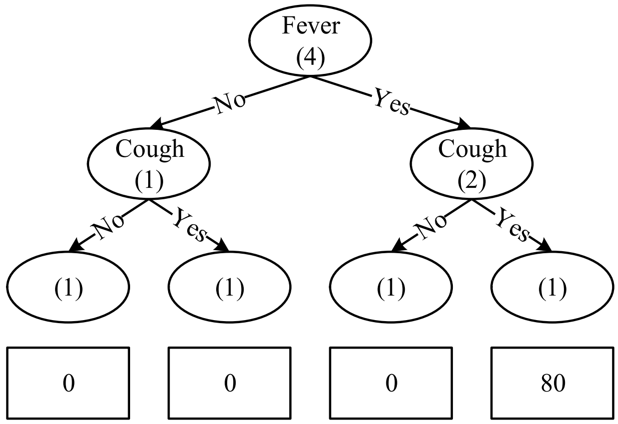


Figure 1: Example of Decision Tree

The root node has 4 samples, the left and right nodes have 2 samples each after the first fork, and all 4 leaf nodes have only 1 sample after the second fork. The leaf nodes take known values. The decision tree model contains two features Fever and Cough, and all possible feature combinations S are as follows:

(1) $\{\}$

(2) $\{Fever\}$

(3) $\{Cough\}$

(4) $\{Fever, Cough\}$

For the given sample $x=\{Fever: Yes, Cough: Yes\}$, calculate the SHAP value ϕ_{Fever} for Fever and ϕ_{Cough} for Cough.

In this example, $M=\{Fever, Cough\}$, m are the number of features and $m=2$, j are the j th feature.

The calculation process is as follows.

(1) Calculate $f_x(S)$

1) $S=\{\}$. shown in Fig. 2:

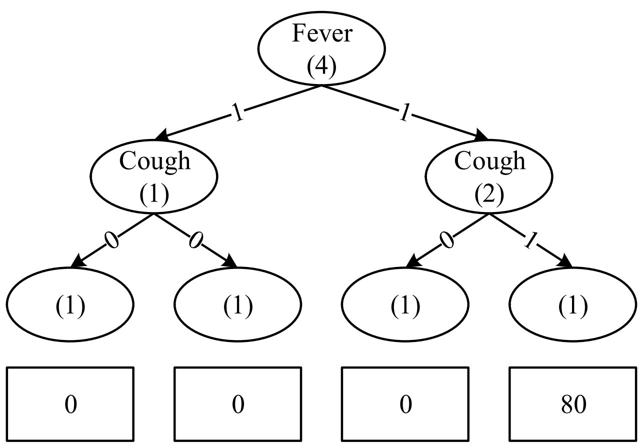


Figure 2: The weight of each edge when $S=\{\}$

In this example, $f_x(S=\{\})=1\times(0\times0+0\times0)+1\times(0\times0+1\times80)=80$.

2) $S=\{Fever\}$. shown in Figure 3:

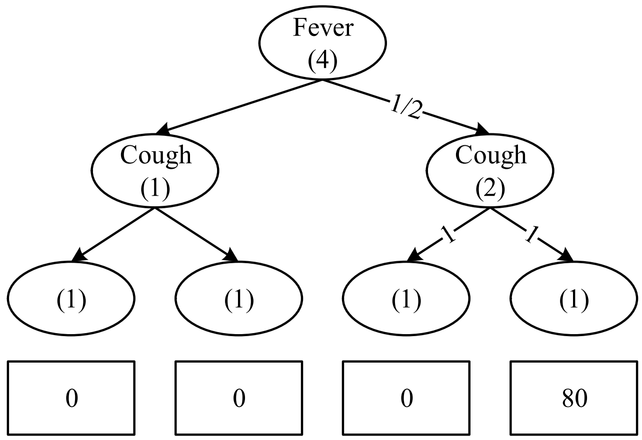


Figure 3: The weights of each side when $S=\{Fever\}$

In this example. $f_x(S=\{Fever\})=\frac{1}{2}\times(1\times0+1\times80)=40$.

3) $S=\{Cough\}$. shown in Figure 4:

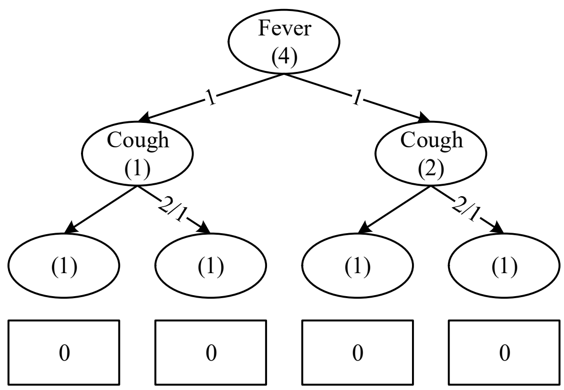


Figure 4: The weights of each side when $S=\{Cough\}$

In this example. $f_x(S=\{Cough\})=1\times(\frac{1}{2}\times0)+1\times(\frac{1}{2}\times80)=40$.

4) $S=\{Fever,Cough\}$. shown in Figure 5:

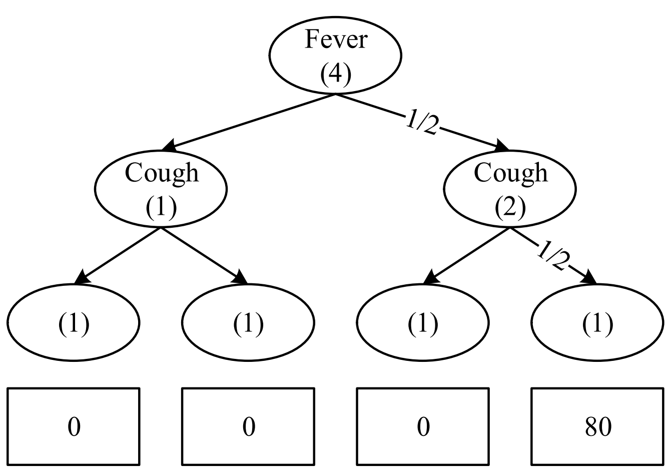


Figure 5: The weights of each side when $S=\{Fever,Cough\}$

In this example, $f_x(S=\{Fever, Cough\}) = \frac{1}{2}\times(\frac{1}{2}\times 80) = 20$.

(2) Calculate SHAP values ϕ_{Fever} and ϕ_{Cough}:

$$\begin{aligned}\phi_{Fever} &= \frac{0!\times(2-0-1)!}{2!} \\ &\times[f_x(S=\{Fever\}) - f_x(S=\{\})] \\ &+\frac{1!\times(2-0-1)!}{2!} \\ &\times[f_x(S=\{Fever, Cough\}) - f_x(S=\{Cough\})] \\ &= \frac{1}{2}\times(40-80)+\frac{1}{2}\times(20-40) \\ &= -30\end{aligned} \tag{12}$$

$$\begin{aligned}\phi_{Cough} &= \frac{0!\times(2-0-1)!}{2!} \\ &\times[f_x(S=\{Cough\}) - f_x(S=\{\})] \\ &+\frac{1!\times(2-0-1)!}{2!} \\ &\times[f_x(S=\{Fever, Cough\}) - f_x(S=\{Fever\})] \\ &= \frac{1}{2}\times(40-80)+\frac{1}{2}\times(20-40) \\ &= -30\end{aligned} \tag{13}$$

$$\phi_0 = f_x(S=\{\}) = 80 \tag{14}$$

(3) Prediction of samples using SHAP values.

$$\begin{aligned}\hat{y} &= \phi_0 + \phi_{Fever} + \phi_{Cough} \\ &= 80-30-30 \\ &= 20\end{aligned} \tag{15}$$

III. Experiments and analysis of results

III. A. Predictive model performance assessment

III. A. 1) Experimental environment and evaluation indicators

The experimental environment is Windows 11 operating system, 13th Gen Intel(R) Core(TM) i7-13700H, 2.40 GHz, 16 GB machine-band RAM, NVIDIA GeForce RTX 4060 Laptop GPU, and the programming language Python 3.8 is used for the simulation experiments through Jupyter and the Pycharm Pro editor for simulation experiments.

In practical application scenarios, other evaluation metrics are usually used auxiliary to comprehensively assess the performance of the model. In this experiment, the following five main classification assessment metrics are used: accuracy, precision, recall, F1 value and AUC value. These metrics reflect the quality and reliability of model predictions from different perspectives. The expressions of each metric are shown below:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{16}$$

$$\Pr ecision = \frac{TP}{TP+FP} \tag{17}$$

$$\text{Re}\,call = \frac{TP}{TP+FN} \tag{18}$$

$$F1 = \frac{2\times \Pr ecision \times \text{Re}\,call}{\Pr ecision + \text{Re}\,call} = \frac{2TP}{2TP+FP+FN} \tag{19}$$

III. A. 2) Comparative experimental analysis

The performance of the prediction results of this paper's algorithm (GBDT) and six current mainstream machine learning models are compared and analyzed, i.e., the model performance metrics are compared with Logistic Regression (LR), Random Forest (RF), Decision Tree (DT), KNN, LightGBM, and XGBoost, and the results of the comparison are shown in Table 1. Comparing and analyzing the performance metrics of each model, the results reveal that the GBDT model outperforms the other models in key performance metrics, such as accuracy, precision, recall, F1 value, and AUC value, which significantly demonstrates its high performance in prediction tasks. In addition, LightGBM and XGBoost models also show better performance, right after the GBDT model. In contrast, the logistic regression and decision tree models show weaker performance in the above performance metrics, especially in the key metrics of accuracy and recall. Based on the above analysis, it can be concluded that the GBDT model exhibits high performance in several of the evaluation metrics assessed in this paper.

Table 1: Model performance index comparison

Model	Accuracy	Precision	Recall	F1	AUC
GBDT	0.851	0.849	0.862	0.86	0.933
LightGBM	0.833	0.828	0.861	0.859	0.907
XGBoost	0.834	0.836	0.862	0.875	0.916
RF	0.83	0.845	0.85	0.853	0.895
KNN	0.82	0.837	0.808	0.825	0.774
LR	0.755	0.8	0.716	0.754	0.82
DT	0.766	0.76	0.768	0.765	0.77

In order to further validate this conclusion, a 10-fold cross-validation method based on accuracy is used in this paper, and the validation results are demonstrated by box-and-line diagrams, and the 10-fold cross-validation results are shown in Fig. 6. In this study, a preliminary comparative analysis of the model performance was conducted using the 10-fold cross-validation box-and-line diagram based on accuracy. The results show that the GBDT model outperforms the other comparative models in terms of accuracy.

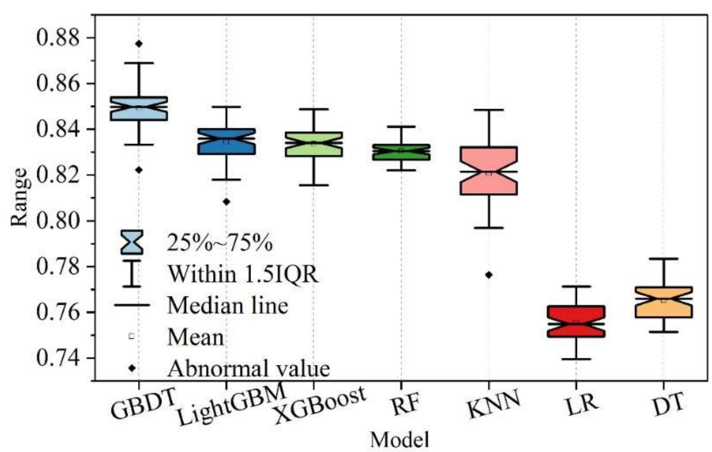
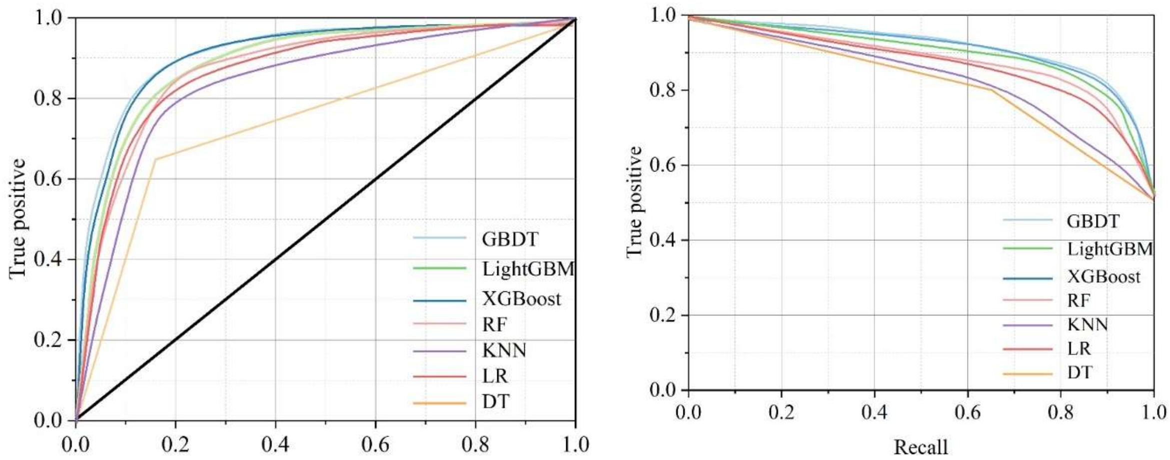


Figure 6: 10 fold cross-proof box diagram

In order to further analyze the performance of each model in depth, AUC curves and precision-recall curves are plotted, and the performance comparison between GBDT and mainstream machine learning is shown in Fig. 7, with (a) and (b) denoting the AUC curves and precision-recall curves, respectively. These two curves provide a more comprehensive view to evaluate and compare the performance of different models on the prediction task, providing a scientific basis for selecting the most suitable model. Based on the experimental data in Table 1 and Fig. 7, it can be observed that the GBDT model outperforms the other comparative models in terms of key performance metrics such as precision, accuracy, and AUC values. GBDT significantly reduces the risk of overfitting during the model training process by solving the problems of gradient bias and prediction bias, which in turn improves the overall performance of the model prediction.



(a) AUC curve (b) Accuracy - Recall curve

Figure 7: Machine learning can be compared

III. B. Oil Painting Color Analysis

In this chapter, the color rendering performance of UV light-curing inks for inkjet printing on grey-backed whiteboard paper is investigated in different gradient tones. Based on the cardboard-ink GATF color wheel diagrams of corrugated paper and grey-backed whiteboard, the CIELAB color space relationship model between grey-backed whiteboard and corrugated paper is established through the GBDT algorithm with and printing ladder scale color data.

The use of Aiseley eXact densitometer to measure the two kinds of paper brightness and color deviation index, the measurement results are shown in Table 2. Corrugated cardboard brightness is lower, only 13%, higher color deviation value of 20%, the color is yellow tends to magenta. The brightness of whiteboard paper with gray background is higher, 75%, and the color deviation value is lower, 3%, and the color is green tending to magenta.

Table 2: Paper index

Paper	Brightness/%	Color offset/%	Partial color
Corrugated paper	13	20%	Y-M
White board	75	3%	C-M

Using Acuity EY flatbed inkjet printing machine in the same printing conditions of different dot area rate of four-color ladder scale, corrugated paper and gray background whiteboard paper two kinds of printing specimens using a densitometer for color measurement, the data are shown in Tables 3 to 6.

Cyan ink color measurements at different orders as shown in Table 3, two cardboard printing cyan ink, with the increase in dot area ratio, a * and b * values show that the blue component and green component increases, gray background whiteboard paper blue-green degree is better than corrugated cardboard, corrugated cardboard due to the yellow-biased magenta base, resulting in the cyan ink printing color green component is more than the blue component.

Table 3: Color measurement results at different levels of cyan ink

Board Color value \ Order modulation (%)		15	40	60	85	100
Corrugated paper	L*	51.83	46.28	40.84	29.43	22.22
	a*	8.02	1.32	-6.09	-23.29	-28.64
	b*	23.46	17.53	12.23	-3.6	-12.7
White board	L*	84.42	76.09	68.01	50.87	41.85
	a*	-5.95	-16.44	-26.17	-37.99	-31.99
	b*	-4.99	-15.47	-26.39	-47.03	-53.87

Magenta block color measurements at different order tones are shown in Table 4. When two kinds of cardboard printing magenta ink, with the increase of dot area rate, a* and b* values show an increase in the red component and blue component, the degree of red and blue of grey-backed white board ink is good, and the color of corrugated paper magenta ink is reddish-yellow, and the printing effect is poor.

Table 4: Color measurement results at different levels of magenta ink

Order modulation (%) / Board Color value		15	40	60	85	100
Corrugated paper	L*	51.28	47.24	42.88	35.26	27.68
	a*	15.35	22.11	28.3	40.82	49.98
	b*	24.86	22.01	19.92	16.57	15.44
White board	L*	82.87	74.3	66.51	51.55	40.89
	a*	8.31	24.27	37.65	64.02	74.99
	b*	-2.02	-7.48	-10.94	-14.32	-7.01

The color measurements of the yellow blocks at different order tones are shown in Table 5. When two kinds of cardboard printing yellow ink, with the increase of dot area rate, the brightness is almost unchanged, the b* value shows that the yellow component increases more, and the ink color of gray-backed white board paper is yellowish-green, and the ink color of corrugated cardboard is reddish-yellow and the degree of yellow is not as good as that of gray-backed white board paper.

Table 5: Color measurement results at different levels of yellow ink

Order modulation (%) / Board Color value		15	40	60	85	100
Corrugated paper	L*	52.08	51.4	52.06	51.4	52.66
	a*	11.36	11.46	10.03	9.01	7.82
	b*	29.97	36.84	42.95	52.52	58.79
White board	L*	87.37	86.49	85.76	84.17	84.27
	a*	-3.12	-7.11	-8.86	-8.56	-5.9
	b*	13.69	34.84	54.59	80.48	94.77

The results of black block color measurements at different order tones are shown in Table 6. When the two cardboards were printed with black ink, the brightness was significantly reduced, and the color of the corrugated cardboard was yellowish reddish in the light, middle and dark tones, and only the field printing of black ink covered the base color completely.

Table 6: Color measurement results at different levels of black ink

Order modulation (%) / Board Color value		15	40	60	85	100
Corrugated paper	L*	47.85	38.5	32.03	18.99	2.44
	a*	10.29	8.5	6.82	4.62	0.55
	b*	23.1	19.71	16.66	10.96	0.56
White board	L*	77.57	61.26	47.49	25.96	1.06
	a*	0.1	0.39	1.12	1.05	0
	b*	1.67	2.57	3.45	3.33	-0.24

The printing effect of the paperboard is shown in Figure 8, with (a) and (b) indicating the color rendering effect of the ink on white board with gray background and on corrugated paper, respectively. A comprehensive analysis of Tables 3 to 6 and Fig. 8 shows that the same kind of ink has different color rendering effects on grey-backed white board and corrugated paper, and the color rendered by the ink on corrugated paper is a result of the joint action of the base color of the corrugated paper and the color of the ink. From the analysis of the colors of the two types of paperboard in different gradations, it can be seen that on whiteboard paper with a gray background, the green, magenta and yellow colors become darker with the increase in the area rate of the printing dot, which means that the smaller the area of the white background of the paper that washes out the ink color, the darker the ink color becomes, and it gradually tends to become saturated. Printing three-color ink on corrugated paper and gray background whiteboard paper color effect is different, due to the existence of the corrugated paper itself yellow-brown background color, affecting the ink color on top of the presentation of the effect. Gray background whiteboard paper can better show the color of the ink itself, while the color shown on corrugated paper is the result of mixing the ink color with the background color, which is a big deviation from the color of the ink itself.

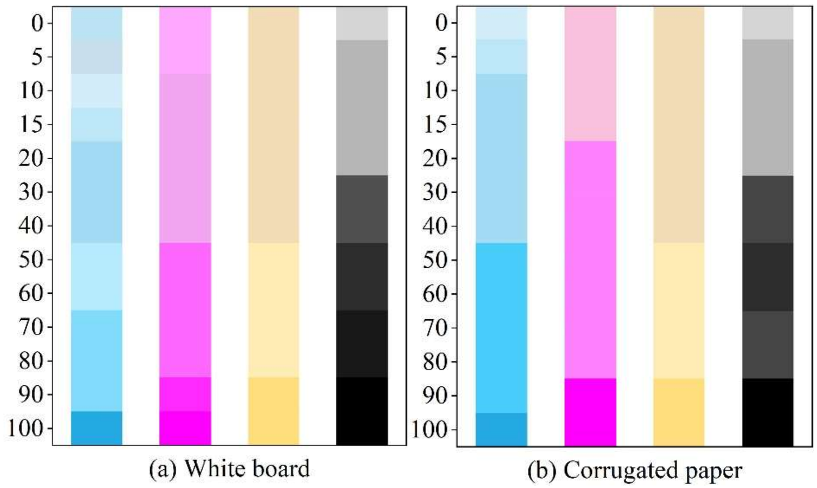


Figure 8: Printing rendering of two kinds of boards

III. C. SHAP Interpretability Analysis

Using the SHAP method to rank the importance of feature variables for the optimal prediction model GBDT, the top 10 ranked variables are shown in Figure 9. Among the top 10 variables, they can be classified into physical property factors (color particle size, color type, canvas texture, pigment layer thickness), painting technique factors (painting sequence, stroke direction and strength), environmental factors (lighting conditions, surrounding environment color), and subjective factors of the painter (painter's color preference, painter's intention to express his emotion). among the 10 variables, color particle size (V1), color type (Among the 10 variables, color particle size (V1), color type (V2), canvas texture (V3), pigment layer thickness (V4), and surrounding environment color (V8) ranked the top 5, which can be seen that their predictive ability for the color characteristics of oil paintings is significantly stronger than the other variables. The last five were painting order (V5), direction and strength of brushstrokes (V6), lighting conditions (V7), painter's color preference (V9), and painter's intention to express emotion (V10).

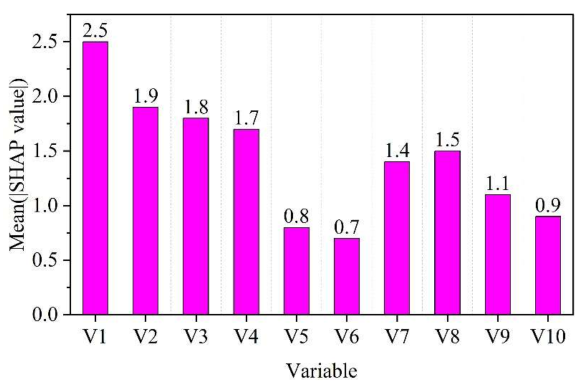


Figure 9: The characteristic variable of the first ten of the order of importance

The effect of the top 10 feature variables in terms of importance on the predicted values of the model is shown in Figure 10.The horizontal axis in the SHAP summary plot is the Shapley value, and the colors represent the magnitude of the variable's value at the sample point, where magenta indicates a larger value, and yellow is the opposite. From the SHAP summary plot, it can be seen that all other conditions being equal, when the color particle size (V1) takes on a progressively larger value (from yellow to magenta), the SHAP value therefore grows from a negative to a positive number, indicating that the probability of the color of the oil painting being predicted at the sample point is progressively increased. When the value of painting order (V5) is gradually increased, the SHAP value thus decreases from positive to negative, indicating that the probability of the oil painting color being predicted at the sample point is gradually decreasing.

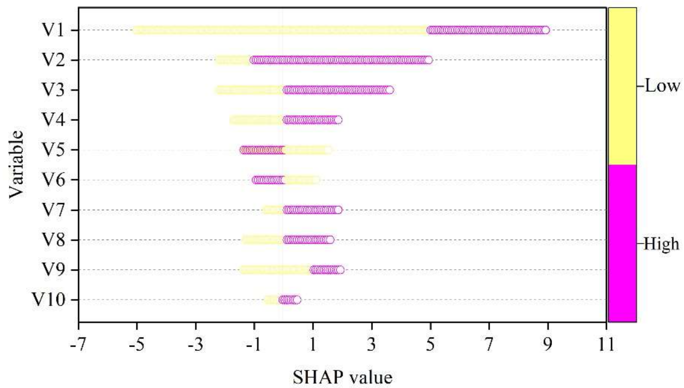


Figure 10: SHAP summary diagram

In summary, this paper compares the GBDT model with other comparative models in terms of key performance indicators such as precision, accuracy and AUC value. The CIELAB color space relationship model of whiteboard paper and corrugated paper with gray background is established to verify the model's effect on the analysis of oil painting color. The machine learning model is also interpreted using the SHAP method, which improves model transparency and facilitates model understanding by model applicators.

IV. Practical application of teaching color in oil painting for art majors

IV. A. Designing topics for course instruction

The purpose of designing the teaching theme of oil painting color course for university art majors is to give full play to the students' subjective initiative, stimulate the students' vitality and enthusiasm for learning, inject fresh blood into the oil painting color teaching for university art majors, and inspire the students' innovative thinking through the interaction between teachers and students. After the basic theoretical teaching, the teacher can specially design a problem related to the lesson for the students and let them solve it through thinking. The teacher's role in this process is that of a guide, providing students with all the specialized knowledge and creative materials they need, and guiding them to diverge their thinking in the right direction. Through this benign interaction, students are able to complete the topics designed by the teacher well, and at the same time enhance the creative ability of their works, deepen their understanding of color, and improve their skills and level of creation. By designing the teaching theme of oil painting color course, the concept of color can be better analyzed for students. Through strengthening color research and analysis, it can cultivate students' color perception ability. When designing the theme of the oil painting color course, the direction can be diversified, reflecting the pursuit of multi-level values and the expression of beauty, helping students to enhance their interest in learning and color expression, and ultimately enhancing the colorfulness of the works.

IV. B. Reasonable and scientific utilization of modern multimedia teaching

University art majors oil painting color teaching generally let students hands-on sketching practice, usually students through sketching practice, can skillfully master the color matching, color mixing, color expression and other basic knowledge and theory, in the creative practice can also basically show the color. The purpose of oil painting color teaching for university art majors is to let students systematically learn and skillfully apply color theory and optimize the combination of various colors. The introduction of advanced multimedia teaching equipment in the university art major oil painting color teaching, the traditional hand painting and modern technology combined, so that the computer becomes a way and means to serve the students' creation. The integration of the two means of manual painting and computer creation can complement each other's advantages, and through the reasonable use of computer technology, it can cultivate students' aesthetic ability and innovation ability, and can also help the teaching from a single transmission to teacher-student interaction, in order to enhance the teaching effect.

IV. C. Reform the oil painting color teaching curriculum system

Establishing a sound oil painting color teaching system for university art majors, raising the importance of color teaching, and strengthening the color perception training for students can cultivate students' color creation ability. When carrying out the reform of oil painting color course teaching, it is necessary to strengthen students' color contrast practice and color application practice, and cultivate students' color application ability. University art majors oil painting color teaching should focus on the design of the teaching mode, according to the actual situation of students and the progress of the course, the oil painting color course should be refined and optimized, and the classroom teaching content should be enriched and improved. The focus of oil painting color teaching in college art majors is to cultivate students' color expression ability. Students can better understand the importance of color matching and color optimization through teachers' well-designed color training, as well as enhance their understanding of color combinations, which is very important for them to exert their professional skills. Through continuous adjustment and exploration, a university art major oil painting color teaching system with modern teaching characteristics and meeting students' requirements is formed, which is conducive to students' comprehensive development.

V. Conclusion

Through the systematic analysis of oil painting color characteristics by the gradient boosting decision tree algorithm, it is found that physical attributes such as the size of color particles, color type, canvas texture and the thickness of the pigment layer have a decisive influence on the final color rendering effect of oil paintings. The algorithm performance comparison experiments show that the F1 value of GBDT model reaches 0.86, and the precision and recall reach 0.849 and 0.862 respectively, which is significantly better than the traditional machine learning methods. Printing color rendering experiments reveal the important role of substrate properties on color performance, the brightness of gray background whiteboard paper is 75% while corrugated paper is only 13%, resulting in the same ink presenting very different color effects on different substrates. Interpretability analysis clarifies 10 key variables affecting the color of oil paintings, among which physical property factors dominate, providing a scientific basis for artistic creation. Applying the research results to art teaching practice, students' color perception ability and creation level are significantly improved through thematic curriculum design, multimedia technology integration and teaching system optimization. The research realizes the in-depth integration of art and science and technology, provides new ideas for the innovation of the traditional art education mode, and is of great value in promoting the modernization of art education.

Funding

This paper is the phased research result of the project: "Research on Quality Training of Normal Students in Normal Universities" (Project No. SJHZDKT202412-02) of Shaanxi Education Society in 2024; The phased research results of the "Excellent Teacher Plan" art major course teaching system construction and practice of the 2024 basic education research key project in Weinan (Project No. WNYBKT202440); The phased research results of jointly constructing the course "Teaching Theory of Fine Arts Discipline" by Weinan Normal University and enterprises in 2024.

References

[1] Sutton, T. M., & Altarriba, J. (2016). Color associations to emotion and emotion-laden words: A collection of norms for stimulus construction and selection. Behavior research methods, 48(2), 686-728.

[2] Jonauskaite, D., Abu-Akel, A., Dael, N., Oberfeld, D., Abdel-Khalek, A. M., Al-Rasheed, A. S., ... & Mohr, C. (2020). Universal patterns in color-emotion associations are further shaped by linguistic and geographic proximity. Psychological Science, 31(10), 1245-1260.

[3] Cavicchio, F., Dachkovsky, S., Leemor, L., Shamay-Tsoory, S., & Sandler, W. (2018). Compositionality in the language of emotion. PloS one, 13(8), e0201970.

[4] Zhang, Y. (2017, May). Personal Emotional Expression in the Creation of Oil Painting. In 2017 International Conference on Culture, Education and Financial Development of Modern Society (ICCESE 2017) (pp. 530-532). Atlantis Press.

[5] Xu, Y. (2016, May). The emotion analysis during oil painting creation process. In 2nd International Conference on Arts, Design and Contemporary Education (pp. 37-40). Atlantis Press.

[6] Bain, J. (2012). Signifying absence: experiencing monochrome imagery in medieval painting. A Wider Trecento: Studies in 13thand 14th-century European Art Presented to Julian Gardner, Leiden, 5-20.

[7] Yuan, Y., & Chu, Y. Y. (2024). Visual impact of colours in various painting media: oil paint, watercolour, pastel, and digital art. Journal of the International Colour Association, 37, 19-35.

[8] Shi, Y., & Tsetsegdelger, D. (2024). A Multidimensional Exploration of the Mechanism of Color Emotion Elicitation in Oil Painting under the Perspective of Psycho-Aesthetics. International Journal of Education and Social Development, 1(2), 20-23.

[9] Mohamed, N. M., Orabi, E. A., Alamri, A. A., Mashlawi, A. M., El-Shabasy, A., & Taleb, T. A. A. (2025). Effect of iron rust on color layer and structural integrity of historical oil paintings. Scientific Reports, 15(1), 1-14.

[10] Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. Journal of applied science and technology trends, 2(01), 20-28.

[11] Tsai, C. M., Yeh, Z. M., & Wang, Y. F. (2011). Decision tree-based contrast enhancement for various color images. Machine Vision and Applications, 22, 21-37.

[12] Tennekes, M., & de Jonge, E. (2014). Tree colors: color schemes for tree-structured data. IEEE transactions on visualization and computer graphics, 20(12), 2072-2081.

[13] Kim, J., Jun, J. Y., Hong, M., Shim, H., & Ahn, J. (2019). Classification of oil painting using machine learning with visualized depth information. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 42, 617-623.

[14] uslimah Molo,Suttida Changsan,Lila Madares,Ruchirada Changkwanyeun,Supang Wattanasoei,Supa Vittaporn... & Sayambhu Saita. 2024). decision tree model for traffic accident prediction among food delivery riders. Epidemiology and health,e2024095.

[15] Enci Yang,Hui Liu,Shangshan Wei,Ziyao Suo,Bohan Jia & Gen Zhang. (2024). Prediction of debris flow disaster in Palongzangbu Basin based on GBDT model and ArcGIS. Journal of Globe Scientific Reports,6(3),54-62.

[16] Muhammad Amir Khan,Bilal Asad,Toomas Vaimann & Ants Kallaste. (2024). An Advanced Diagnostic Approach for Broken Rotor Bar Detection and Classification in DTC Controlled Induction Motors by Leveraging Dynamic SHAP Interaction Feature Selection (DSHAP-IFS) GBDT Methodology. Machines,12(7),495-495.