https://doi.org/10.31449/inf.v48i21.6891

Painting Style and Sentiment Recognition Using Multi-Feature Fusion and Style Migration Techniques

Jiaolong Cheng[1], Lianshuqing Yang[1], Shuyang Tong[2*]
[1]College of Art and Design, Hebei Normal University for Nationalities, Chengde 067000, China
[2]Department of Art Education, Baoding Preschool Teachers College, Baoding 072750, China
E-mail: tsyberelax@163.com
[*]Corresponding author

Keywords: computer vision, drawing; style recognition, multi-feature fusion, sentiment analysis, style migration

Received: September 20, 2024

With the continuous improvement of living standards, it has become a trend to improve one's cultural knowledge. However, there is an urgent need to find a solution to the problem of how to understand the artistic features of the pieces. In view of this, the study used computer vision technology to identify and analyze the style and mood of the paintings. First, a painting style classification recognition model based on multi-feature fusion was built by using wavelet transform to denoise and fuse color, texture, and spatial features of paintings. Then, the dataset was expanded using style migration, and style features and high-level semantic features were collected to build a painting sentiment analysis model based on style migration. The proposed model consistently achieved a classification accuracy and recall above 95%, demonstrating strong performance across tests. On the WikiArt dataset and the OilPainting dataset, the subject manipulation characteristic curves of the proposed model were able to completely cover the results of the other three models with the best overall performance. The proposed painting sentiment analysis model demonstrated good recognition performance on both the MART dataset and deviantArt dataset, with recognition accuracy above 90%. Experimental results of the study help to accurately categorize the styles and recognize the emotions of paintings, providing certain convenience for the appreciation of art works.

Povzetek: Nova metoda s pomočjo večfunkcijskega združevanja in prenosa stila omogoča kvalitetno prepoznavo slikarskega stila in čustev, kar prispeva k umetniški analitiki.

1 Introduction

With the advent of the information age, the use of search engines provides quick access to a variety of image resources on the Internet. Style is a self-contained attribute of the images themselves, usually indicating the creative characteristics of an artist or artistic group in a particular time context [1]. However, how to find the style of interest and understand the stylistic characteristics of a work from a large number of paintings is an urgent problem to be solved [2]. Painting style (PS) recognition is an image recognition technique that involves classifying and recognizing images based on the the PS. Yu et al. suggested a new less-parameter multi-channel generative adversarial network (GAN) for the style conversion problem of high-resolution semantic images. Experimental results demonstrated that the proposed model could effectively recognize the artistic styles of multiple images, and could generate clear and high quality images with high generation speed [3]. Jiang et al. proposed a convolution neural network (CNN)-based style feature extraction method for paintings in response to the low accuracy of traditional style feature extraction approaches for paintings. The outcomes indicated that the alignment rate of the proposed model could be kept above 0.7 and the accuracy rate was always above 0.9, which was highly adaptable [4]. Zhong et al. proposed a style rendering model based on improved CNN in an attempt to algorithmically create artworks that mimic the PSs of famous masters. Experimental results demonstrated that the information entropy of the proposed model was 5.58, and the average gradient value and peak signal-to-noise ratio were 22.54 and 27.81, respectively, which had a good performance of PS recognition and imitation [5]. Cascone et al. addressed the problem of categorizing PSs in murals, paintings or sculptures by collecting murals in eleven PSs ranging from prehistoric to surrealist. The outcomes revealed that the utilization of color, shape and texture feature (TF) information could be effective for style classification and help to promote the application of machine learning and computer vision (CV) techniques in artwork reconstruction work [6]. Madhu et al. introduced a perceptually based style migration training for image style conversion to address the style recognition problem of art works. Experimental results showed that the proposed method learned generic domain styles with high average precision and average recall [7]. Peng et al. proposed an image style conversion framework in order to make the effect of image style conversion closer to the characteristics of traditional Chinese paintings, and computed the gradient map by edge detection operator. According to the results, the suggested technique could translate authentic scene photographs into Chinese landscape paintings that were more sensitive to image borders [8].

Traditional classification of PSs is mainly carried out by experts based on image content, which usually takes a lot of time and energy. With the rapid development of computer technology, various intelligent technologies are gradually applied to the classification of paintings. Instead of using human eyes to detect, track, and measure the target, CV uses a camera and a computer. Zhang et al. proposed a multi-scale feature analysis using reduced units to address the lack of intrinsic inductive bias of vision transformers in some CV tasks. Experimental results demonstrated that the proposed transformer could achieve a Top-1 best classification accuracy (CA) of 91.2% [9]. Guo et al. proposed that the attention mechanism (AM) might be understood as a dynamic weight adjustment process depending upon the input visual attributes. A thorough analysis of the various AMs in CV was conducted in an attempt to explore the possible uses of the AM in the field of CV. This analysis led to a methodical classification. Tasks related to vision, including image generation, classification, self-supervised learning, and multimodal tasks, were resolved with the aid of this study [10]. Jang et al. proposed a style transfer network that can perform local style editing on images. By integrating adaptive instance normalization and attention modules, it could transfer global and local features of styles and facilitate style transfer, which helped to capture the style of dynamic images [11]. Tang et al. constructed an attention-free network based on an existing multilayer perceptron-based vision model in an attempt explore the role of self-AM in Transformer. Experimental results showed that the proposed model achieved 81.9% accuracy using only 24M parameters, so the self-AM does not necessarily facilitate the solution of CV tasks [12]. A summary of the most recent research on deep multimodal learning was provided by Bayoudh et al. A number of benchmark datasets are also suggested by the study for use in problem-solving across different vision domains [13]. With the goal of enhancing the naturalness of human-computer interaction, Qi et al. examined the technique of gesture recognition on images predicated on CV technology. The main process of the method included data acquisition, gesture detection and segmentation, feature extraction and gesture classification. Gesture recognition algorithms for human-computer interaction were also studied which could be used for effective and efficient human-computer interaction [14]. Gururaj et al. addressed the problem of low accuracy of automatic product grading system based on image recognition. According to the findings, the suggested approach's recognition accuracy was 93.23%, demonstrating its viability and efficacy [15]. The summary table of the relevant studies mentioned above is shown in Table 1.

Table 1: The summary table of the relevant studies

Study	Method	Result	Insufficient
Yu Y	Multi channel generative adversarial network	Identify artistic styles of multiple images	The dataset is not rich enough
Jiang H	Convolution neural network (CNN)	Registration rate above 0.7, accuracy above 0.9	Lack of comparison with more advanced methods
Zhong Y	Improving CNN	The average gradient value and peak signal-to-noise ratio are 22.54 and 27.81, respectively	The dataset is not rich enough
Madhu P	Perception based style transfer training	Has high average accuracy and average recall rate	Low operational efficiency
Peng X	Image style conversion framework	Good sensitivity to image edges	Low recognition accuracy
Zhang Q	Using reduced units for multi-scale feature analysis	The optimal classification accuracy can reach 91.2%	Low operational efficiency
Jang DK	Adaptive instance normalization and attention	High accuracy in capturing the style of dynamic images	Low operational efficiency
Tang C	Visual model based on multi-layer perceptron	An accuracy rate of 81.9% was achieved using only 24M parameters	Low recognition accuracy
Gururaj N	Random forest classifier	The recognition accuracy is 93.23%	Lack of comparison with more advanced methods

In summary, although PS categorization and CV techniques have been widely researched, existing image categorization techniques still suffer from low accuracy and neglect the emotional factors of paintings [16]. However, existing CNN-based feature extraction and GAN models have multiple hyperparameters that need to be tuned, and the amount of training data also limits the performance of the model. In this context, the study proposes a PS classification and recognition model (PSCRM) based on multi-feature fusion (MFF) and a painting emotion analysis model based on style migration. The ability to denoise images while integrating color, texture, and spatial features of paintings can better capture the diversity of PSs. It also comprehensively considers the style characteristics and emotional information of paintings, thereby enhancing the analytical scope of painting works. The study aims to effectively classify and recognize the style and emotion of paintings, and then help the related people to quickly understand the style characteristics of the works. The innovation of the study lies in the use of wavelet transform for denoising, and then Fisher Score and tandem fusion are used to combine color, texture and spatial features to realize the style classification of paintings.

2 Methods and materials

To identify and analyze paintings, the study build a PSCRM based on MFF by combining color, texture and spatial features of paintings through Fisher Score and tandem fusion strategies. Then the dataset is expanded using style migration. Moreover, the style features and high-level semantic features are collected to build a painting sentiment analysis model based on style migration.

2.1 Construction of PSCRM based on MFF

PS classification helps to understand and appreciate the diversity and depth of painting art from multiple perspectives. To identify and classify the styles of painting images more efficiently and accurately, the study will use the color, texture and spatial features of painting images to build a PS classification model. First, the painting images need to be preprocessed, including image normalization and denoising [17-18]. The acquired painting images will contain noise due to the effects of the surrounding environment and the equipment used in the process. Therefore, the study uses wavelet transform to denoise the image. Wavelet transform is a specific window function used to determine the convolution which provides a mathematical framework for decomposing an image into different scale compositions [19]. In Figure 1, the particular structure is displayed.

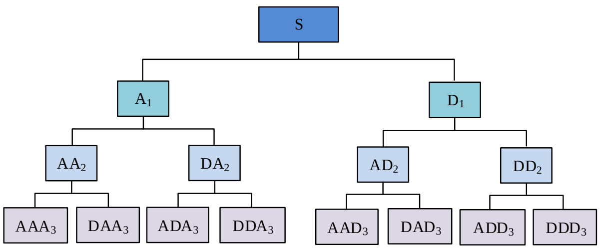


Figure 1: Structure diagram of wavelet transform.

Next, color features are extracted. The most popular visual feature in image retrieval is color, which is frequently very relevant to the objects or situations in the image. Common color spaces (CS) include RGB, HSV and CMYK. Among them, HSV CS contains hue, saturation and brightness features, which are more in line with the visual intuition of the human eye [20]. Therefore, HSV CS is chosen to calculate its color histogram features. The hue H in HSV CS is divided into 8 parts, saturation S and brightness V are divided into 3 parts respectively. Moreover, H, S and V are synthesized into a one-dimensional feature vector (FV) I as shown in Equation (1).

$$I = Q_S Q_V H + Q_V S + V \quad (1)$$

In Equation (1), Q_S and Q_V denote the quantization grades of S and V, respectively, which take the value of 3. The mean μ, variance σ^2, skewness μ_s, cragness μ_k, and energy μ_N are calculated as shown in Equation (2).

$$\begin{cases} \mu = \sum_{i=0}^{72} iP(i) \\ \sigma^2 = \sum_{i=0}^{72} (i-\mu)^2 P(i) \\ \mu_s = \dfrac{1}{\sigma^3} \sum_{i=0}^{72} (i-\mu)^3 P(i) \\ \mu_k = \dfrac{1}{\sigma^4} \sum_{i=0}^{72} (i-\mu)^4 P(i) \\ \mu_N = \sum_{i=0}^{72} P^2(i) \end{cases} \quad (2)$$

In Equation (2), $P(i)$ denotes the probability that the i th color appears in the oil painting image. The color features f_c on the H, S and V components are shown in Equation (3).

$$\begin{cases} f_c = (c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8) \\ c_i = (\mu_i, \sigma_i^2, \mu_{si}, \mu_{ki}, \mu_{Ni}), i = 1, 2, ..., 8 \end{cases} \quad (3)$$

Then, the TFs of the painting images are extracted. TF is a visual feature that describes the homogeneity of an image. For the extraction of image TFs, the study uses gray level co-occurrence matrix (GLCM). By examining the statistical distribution of gray levels and the spatial interaction between image pixels, this technique determines the TFs of a given image. Corner second order moments A, contrast C, entropy E, inverse difference matrix H and correlation coefficient R are chosen as evaluation metrics for the study. The calculation of each index is shown in Equation (4).

$$\begin{cases} A = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P^2(i,j) \\ C = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i-j)^2 P(i,j) \\ E = -\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i,j) \lg P(i,j) \\ H = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \dfrac{P(i,j)}{1+(i-j)^2} \\ R = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \dfrac{ijP(i,j) - \mu_1\mu_2}{\sigma_1\sigma_2} \end{cases} \quad (4)$$

In Equation (4), i and j denote the pixel points with corresponding gray level layers. $P(i,j)$ denotes the joint probability.

When $\mu_1=\sum_{i=0}^{L-1} i\sum_{j=0}^{L-1}P(i,j)$, $\mu_2=\sum_{i=0}^{L-1} j\sum_{j=0}^{L-1}P(i,j)$, $\sigma_1=\sum_{i=0}^{L-1}(i-\mu_1)^2\sum_{j=0}^{L-1}P(i,j)$, and $\sigma_2=\sum_{i=0}^{L-1}(j-\mu_2)^2\sum_{j=0}^{L-1}P(i,j)$,

then the texture characteristics of each region of the image are shown in Equation (5).

$$\begin{cases} f_t=(t_1,t_2,t_3,t_4,t_5,t_6,t_7,t_8) \\ t_i=(\mu_i^A,\mu_i^C,\mu_i^E,\mu_i^R,\sigma_i^A,\sigma_i^C,\sigma_i^E,\sigma_i^R), i=1,2,...,8 \end{cases} \quad (5)$$

Finally, the spatial features of the image are extracted. Commonly used methods include the nine-grid method and the upper-middle-lower chunking method. However, these two methods may cause a certain degree of damage to the integrity of the image, and the extraction effect of spatial features is more limited. Therefore, this study proposes an improved spatial feature extraction method, and the specific schematic diagram is shown in Figure 2.

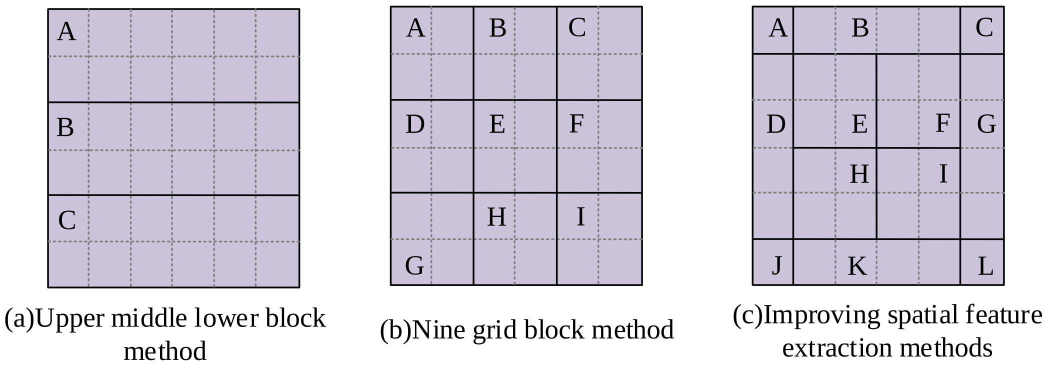


Figure 2: Schematic diagram of three spatial feature extraction methods

In Figure 2, the main area included in the nine-grid method is too small and the edge parts are wide, making it easy to miss important feature information. The upper middle lower block method is only applicable to oil paintings with horizontal composition. The improved spatial feature extraction method proposed by the research combines the advantages of the nine grid method and the upper middle lower block method, including a more complete picture subject. After the features are extracted, the color and TFs need to be normalized for subsequent analysis. Assuming that there are M paintings, the corresponding FV of each painting is $P=(P_1,P_2,...,P_M)$. Assuming that the FV conforms to Gaussian distribution, the mean μ_j and standard deviation σ_j are calculated as shown in Equation (6).

$$\begin{cases} \mu_j=(\sum_{i=0}^{M}F_{ij})/M \\ \sigma_j=\sqrt{\dfrac{\sum_{i=0}^{M}(F_{ij}-\mu_j)^2}{M-1}} \end{cases} \quad (6)$$

In Equation (6), F_{ij} denotes the feature matrix. After obtaining the eigenvectors, they are normalized to the [-1,1] interval as shown in Equation (7).

$$f_{ij}^N=\frac{f_{ij}-\mu_i}{\sigma_j} \quad (7)$$

In Equation (7), f_{ij} denotes the j th feature element of the i th row of the feature matrix. After completing the feature normalization, the features can be further fused. The study utilizes Fisher Score to evaluate the feature correlation values. The weight W_{f_i} of the corresponding feature A_i of f_i is shown in Equation (8).

$$W_{f_i}=\frac{\sum_{c=1}^{N}\left(\bar{T}_{f_i}-\bar{T}_{f_{i,c}}\right)}{\sum_{c=1}^{N}\sigma_{f_{i,c}}^2} \quad (8)$$

In Equation (8), N denotes the number of categories. $\bar{T}_{f_i}$ denotes the mean of the eigenvalue (MOE) f_i in the training set (TS). $\bar{T}_{f_{i,c}}$ denotes the MOE f_i in the TS of category c. $\sigma_{f_{i,c}}^2$ denotes the variance of the feature value f_i in the TS of category c. The study uses the weighted fusion method to fuse the color and TFs as shown in Equation (9).

$$F=\omega_1 f_c+\omega_2 f_t \quad (9)$$

In Equation (9), F denotes the color texture fusion feature. ω_1 and ω_2 denote the correlation values of color and TFs, respectively. Since color and texture are 2D features and spatial features are 3D features, feature fusion needs to be performed again. F is fused in series with the spatial features to form the fusion feature $F_f = (F_1, F_2, ..., F_i)$. The Naive Bayes classifier (NBC) relies on the premise of conditional independence of features and is based on Bayes' theorem [21]. The study utilizes NBC to classify PSs and votes on the classification results of NBC using spatial voting method in order to make the classification results more focused on the subject area of the painting. The voting function is constructed as shown in Equation (10).

$$f_i(r_i) = \begin{cases} \frac{1}{i}, r_i = i \\ 0, r_i \neq i \end{cases}, i = 1, 2, ..., 8 \quad (10)$$

In Equation (10), $f_i(r_i)$ denotes the number of votes the painting received in category i. When $i = 1, 2, 3, 4$, r_i is the classification result of the painting's center region feature. When $i = 5, 6, 7, 8$, r_i denotes the classification result of the features of the background region of the painting. Each region of the painting image is classified using NBC. The classification result is $r = (r_1, r_2, ..., r_8)$. The final PS classification result R is shown in Equation (11).

$$\begin{cases} A_1 = \sum_{i=1}^{8} r_i f_1(r_i) \\ B_1 = \sum_{i=1}^{8} r_i f_2(r_i) \\ C_1 = \sum_{i=1}^{8} r_i f_3(r_i) \\ R = \max\{A_1, B_1, C_1\} \end{cases} \quad (11)$$

In Equation (11), A_1, B_1 and C_1 denote the number of votes for each category. In summary, the framework of the MFF-based PSCRM built by the study is shown in figure 3.

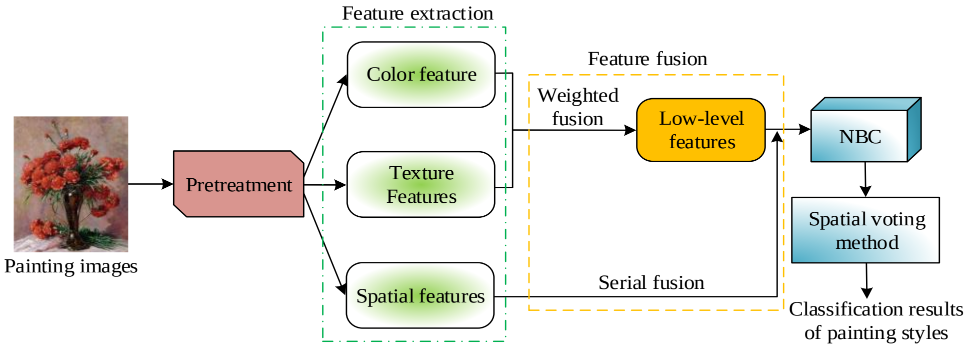


Figure 3: Frame diagram of PSCRM

2.2 Construction of a painting sentiment analysis model based on style migration

The PSCRM based on MFF can only recognize and classify the PS. To further analyze the painting images, the study will identify the emotion of the paintings. Neural network models need to be trained on large-scale datasets to learn better label prediction ability. Using data augmentation techniques can greatly expand the training samples (TS) in the dataset and prevent the model from overfitting. Style migration data augmentation is an important method of generative data augmentation that improves the model's ability to recognize high-level semantic information from disturbances [22]. Therefore, the study uses style migration data augmentation to expand the dataset. The study uses MetaStyle pre-trained model for style migration. The flow of style migration data enhancement is shown in figure 4.

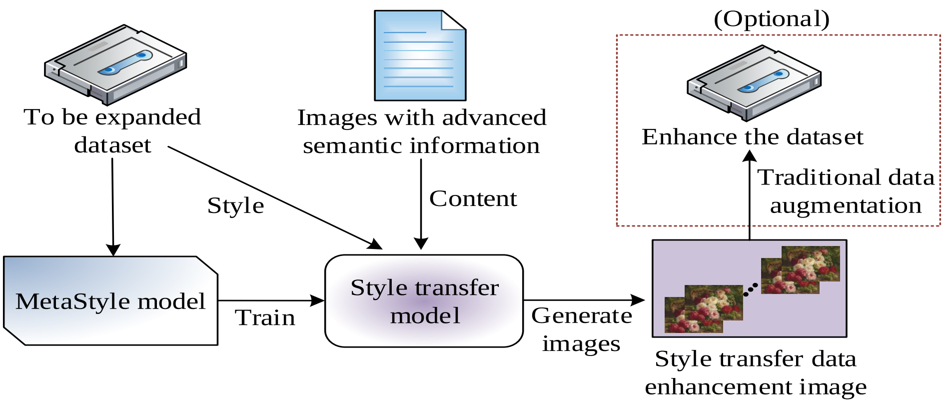


Figure 4: The process of style transfer and data enhancement

Style migration is an optimization technique that is mainly used to acquire two images and blend them to generate a new image. The VGG network's output determines the image's content, and the Euclidean distance is utilized to calculate the difference in content between the two images. The style migration perceptual loss consists of a combination of content loss $loss_c$ and style loss $loss_s$ of the images as shown in Equation (12).

$$\begin{cases} loss_c(I_c, I_x) = \frac{1}{N_i}\left\|\varphi_i(I_c) - \varphi_i(I_x)\right\|_2^2 \\ loss_s(I_s, I_x) = \sum_{i\in S}\left\|G\left(\varphi_i(I_s)\right) - G\left(\varphi_i(I_x)\right)\right\|_F^2 \end{cases} \quad (12)$$

In Equation (12), I_c denotes the content image and I_x denotes the best generated image. N_i denotes the number of features in layer i, and φ_i is the feature map in layer i of the VGG network. I_s denotes the style image and S denotes the feature space. $G(x)$ denotes Gram matrix and F denotes Frobenius paradigm. A weighted summation of $loss_c$ and $loss_s$ yields the style migration perceptual loss. To reduce the number of training sessions, a two-layer optimization (TLO) framework is introduced in the MetaStyle model. The TLO problem is formulated as Equation (13).

$$\begin{aligned} &\underset{\theta}{\text{minimize}}\, E_{c,s}\left[loss(I_c, I_s, M(I_c; w_{s,T}))\right] \\ &\text{subject to } \; w_{s,0} = \theta \\ &\qquad\qquad w_{s,t} = w_{s,t-1} - \delta\nabla E_c\left[loss(I_c, I_s, M(I_c; w_{s,t-1}))\right] \end{aligned} \quad (13)$$

In Equation (13), E denotes the external problem. M denotes the style migration network model. $w_{s,T}$ denotes the style adjusted parameters. T is the maximum steps. θ is the initialization parameters of the model. δ is the learning rate (LR) of the internal target. For the stylistic features and high-level objective semantic features of the images, the study uses a pre-trained VGG19 network, and a combination of squeeze-excitation (SE) AM and residual convolutional blocks for extraction, respectively. The study also uses a global average pooling operation for downsampling the information. The VGG19 network is a classical deep CNN model divided into five convolutional groups. Its connects a maximum pooling layer at the end of each convolutional group. The study uses deep residual network (ResNet) in residual convolutional blocks, which effectively mitigates the gradient vanishing problem. ResNet is a deep neural network structure. Its core idea is to transfer information from the lower layers of the neural network directly to the higher layers by introducing jump connections. The calculation process of residual block is shown in Equation (14).

$$\begin{cases} y_n = g(x_n) + H(x_n, w_n) \\ x_{n+1} = s(y_n) \end{cases} \quad (14)$$

In Equation (14), y denotes the output. $g(\cdot)$ denotes the mapping relation. x denotes the input. $H(\cdot)$ denotes the residual function. $s(\cdot)$ denotes the activation function (AF). Define the difference between the input and the output of the residual module as $T_n = x_n - y_n$. Constant mapping in ResNet means connecting the input directly to the output without any transformation or dimension reduction process, which can reduce the effect brought by redundant layers. In addition, to make the network focus on the focal information of the channel dimension, the study also introduces the SE AM. The SE AM is an adaptive weight adjustment method for deep neural networks. By adjusting the channel dimensions for attention, it improves the performance of the model [23]. For the feature map F of dimension $H\times W\times C$, the SE AM processes it in three stages of squeezing, excitation and assignment of values as shown in Equation (15).

$$\begin{cases} z = f_{squeeze}(F) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}F(i,j) \\ s = f_{excitation}(z,W) = \sigma\left(g(z,W)\right) = \sigma\left(W_2\xi(W_1 z)\right) \\ X = f_{scale}(F,s) = sF \end{cases} \quad (15)$$

In Equation (15), z denotes the vector of $1\times 1\times C$. $\sigma(\cdot)$ denotes the sigmoid AF. $\xi(\cdot)$ denotes the ReLU AF. s denotes the weight information of each channel. In summary, the proposed painting sentiment analysis model based on style migration is studied to extract the stylistic features and high-level semantic features of the image respectively and perform feature fusion. Then the fused features are analyzed for sentiment analysis. Figure 5 depicts the model's precise structure.

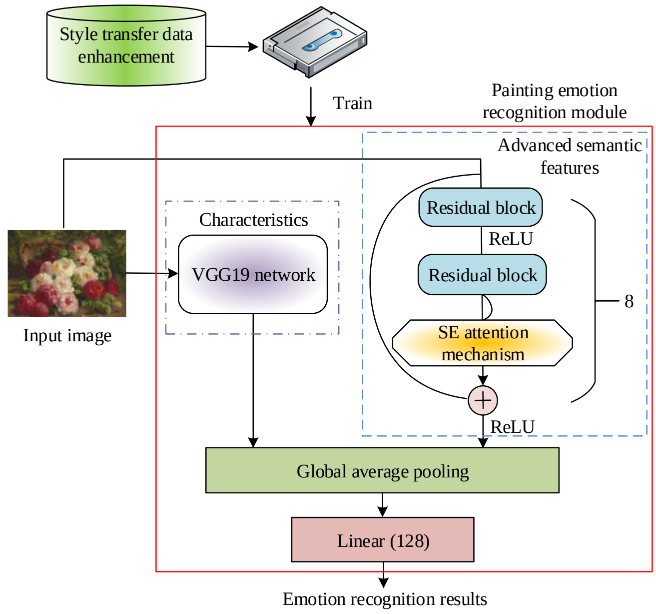


Figure 5: Structure diagram of painting emotion analysis model based on style transfer

3 Results

The study builds a PSCRM based on MFF and a painting sentiment analysis model based on style migration, but its performance has to be further verified. The study mainly analyzes from two aspects, firstly, the validity of the PSCRM based on MFF is analyzed. Then the feasibility of the painting sentiment analysis model based on style migration is verified.

3.1 Validity of a model for categorizing and identifying PSs

To validate the effectiveness of the PSCRM based on MFF, the study conducts simulation experiments based on Windows 10 operating system, Intel Core i7-4790 CPU processor, and Matlab R2018a simulation environment mechanical energy. The study uses WikiArt dataset for testing and divides it into TS and testing set in accordance with the ratio of 8:2. The WikiArt dataset is a large collection of paintings by 195 artists, primarily used for the development of digital art archiving and retrieval systems. The study sets the fusion weight of color features to 0.32 and the fusion weight of TFs to 0.68. The parameters of NBC are calculated by grid search algorithm and the search interval is set to (0,40]. The experiment is repeated for 10 times. Figure 6 displays the suggested model's accuracy, recall, F1 value, and area under the curve (AUC) test results. The proposed PSCRM demonstrates good style classification performance in all 10 tests. The CA and recall are always at a high level, both above 95%, and the AUC value stays above 90%. The outcomes display that the PSCRM based on MFF has better style classification performance, and has certain feasibility and effectiveness.

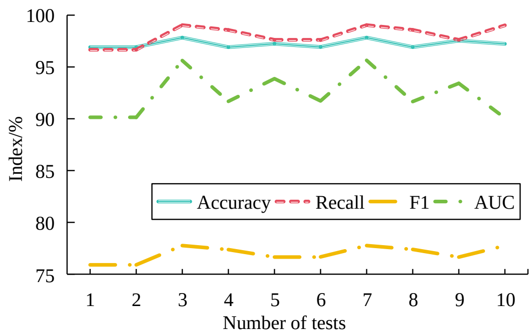


Figure 6: Classification and recognition results of PSs using the proposed model

The study uses ablation experiments to confirm that the suggested technique works. The study compares the accuracy and recall of the model (A) using the traditional nine-cell blocking method, the model without Fisher Score (B), the model without spatial vote selection (C) and the complete model. The confidence level is 95%. Figure presents the findings. Among the four models in Fig. 7(a), the complete model has the highest CA of 96.68% and the model without Fisher Score has the lowest CA of 87.84%. The improvement strategy proposed by the research has the potential to enhance the CA of the model, thereby substantiating the assertion that Fisher Score is the most effective method for optimizing model performance. In Fig. 7(b), the complete model still has the highest recall of 95.83%, followed by the model without spatial voting method with 91.05%. Experimental results show that all the improvement strategies proposed in the study can effectively improve the performance of the PSCRM, which has certain feasibility and effectiveness. However, these improvement strategies have also increased the complexity of the model. Moreover, for complex models, especially those with a large number of parameters, small parameter changes may lead to significant performance changes.

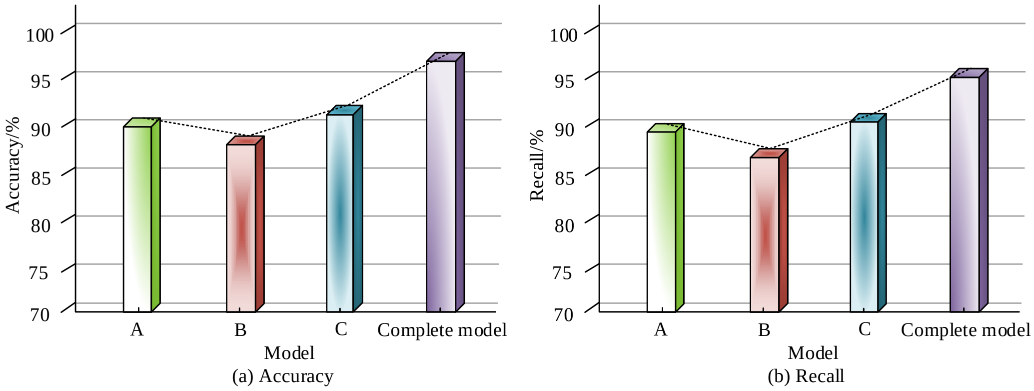


Figure 7: Results of ablation experiment

To further validate the superiority of the MFF-based PSCRM, the study uses the WikiArt dataset and the OilPainting dataset for testing. The OilPainting dataset comprises a collection of oil PSs, encompassing six distinct types, each represented by 30 images and corresponding prompts. These are employed to train models capable of generating images exhibiting the characteristics of a given style. It is compared with three more popular models of fusion nearest neighbor-genetic algorithm (NN-GA), Fisher Score-support vector machine (FS-SVM) and NBS combined with Fisher Score (FS-NBS). The comparison results of receiver operating characteristic (ROC) curves are shown in figure 8. The suggested model performs better overall on the WikiArt dataset shown in Fig. 8(a) since its ROC curves can fully encompass the outcomes of the other three models. This is followed by the NN-GA model, and the FS-SVM model has the worst performance. The suggested model continues to have the greatest overall performance when using the OilPainting dataset in Fig. 8(b) manner. The findings indicate that the PSCRM based on MFF performs better in classification and exhibits some superiority.

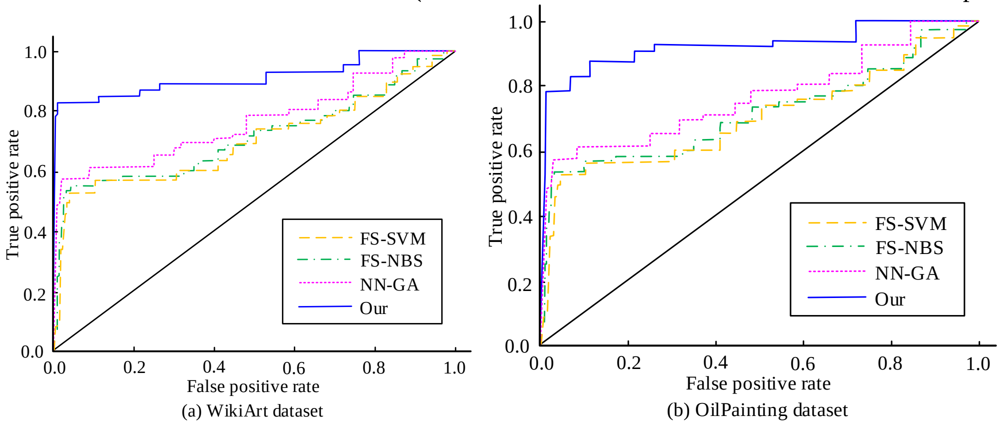


Figure 8: Comparison of P-R and ROC curves for four models

To verify the computational efficiency of the proposed model, the study compared it with baseline models such as CNN, GAN, and LSTM using runtime as the evaluation metric. The results are presented in Table 2. The GAN and LSTM models have higher efficiency on both datasets, with running time not exceeding 20 seconds. The running time of the models proposed by the research is the longest, but still within an acceptable range. Experimental results show that although the proposed model improves the recognition performance, it also increases the complexity of the model to some extent, resulting in a longer running time, but still within an acceptable range. Therefore, it still has advantages overall.

Table 2: Comparison of running time

Model	Data set	
	WikiArt	OilPainting
CNN	26.53s	27.66s
GAN	13.49s	14.18s
LSTM	14.76s	15.31s
Our	28.47s	29.42s

3.2 Feasibility analysis of a sentiment analysis model for drawing

The study is tested using MART dataset and deviantArt dataset. The MART dataset and deviantArt dataset are binary classification datasets used for abstract image emotion recognition, each containing 500 abstract images. The images in the MART dataset are collected from the Museum of Modern Art in Trento and Rovereto, while the images in the deviantArt dataset are collected from the online art community Deviant Ar. Among them, the MART dataset has 131 images labeled as negative emotions and 369 images labeled as positive emotions. The deviantArt dataset has 138 images labeled as negative emotions and 362 images labeled as positive emotions. In style migration data augmentation, the study uses the FI dataset to expand the TSs in the dataset. The study sets the batch size to 16, the epoch to 100, and the LR to 0.0001. To investigate the effectiveness of the proposed MetaStyle based style migration data augmentation strategy, three models are used: a three-layer fully connected network (FCN), SVM, and decision tree. The training is conducted on the original MART dataset, datasets expanded by traditional data augmentation methods (scaling, translation, flipping), datasets expanded by GAN style migration method, and datasets expanded by the proposed style migration data augmentation strategy. Table 3 displays the three models' categorization performance. On the dataset obtained by studying the proposed style migration data enhancement strategy, all three models perform better in terms of metrics. Among them, the SVM model has higher accuracy and recall, which are 78.83% and 78.11%, respectively. Experimental results demonstrate that the proposed strategy for enhancing data regarding style migration can effectively enrich the dataset, thereby improving the recognition performance of the models. This approach is both feasible and effective.

Table 3: Classification performance of models on different datasets

Model	Index	Original dataset	Traditional data augmentation	GAN	Our
FCN	Accuracy/%	59.81	71.62	72.87	73.12
	Recall/%	59.26	70.89	72.17	72.54
SVM	Accuracy/%	74.23	74.35	78.72	78.83
	Recall/%	73.69	73.47	78.04	78.11
Decision tree	Accuracy/%	72.16	73.24	78.15	78.21
	Recall/%	71.68	72.77	77.49	77.53

To investigate the feasibility of a style migration-based painting sentiment analysis model, the study conducts ablation experiments. The recognition accuracy and recall of the model without style migration data enhancement module (A), the model without VGG19 network (B), the model without SE AM (C), and the full model are compared. The confidence level is 95%. Figure 9 presents the findings. In Fig. 9(a), the full model has the highest recognition accuracy of 95.02%. This is followed by the model without SE AM, with a recognition accuracy of 92.84%. In Fig. 9(b), the full model still has the highest recall rate of 93.86%. Experimental results show that the data enhancement module, style features and AM proposed in the study are all effective in improving the performance of the painting sentiment analysis model with certain feasibility and effectiveness.

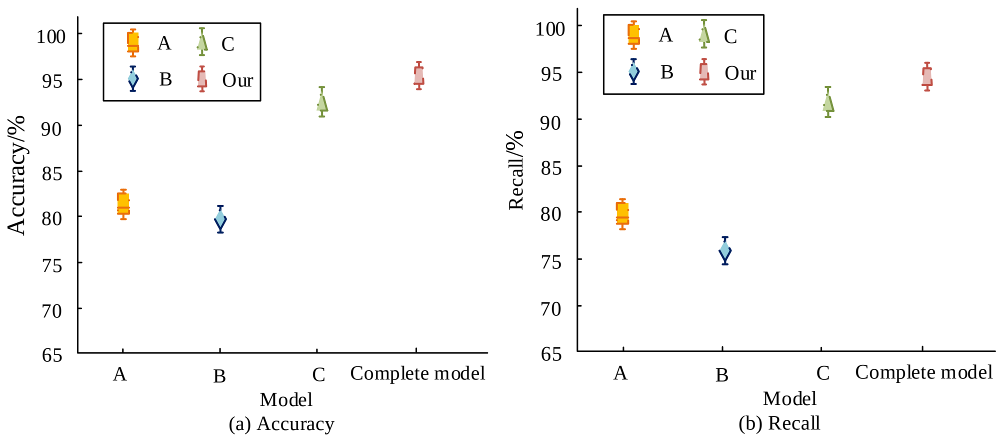
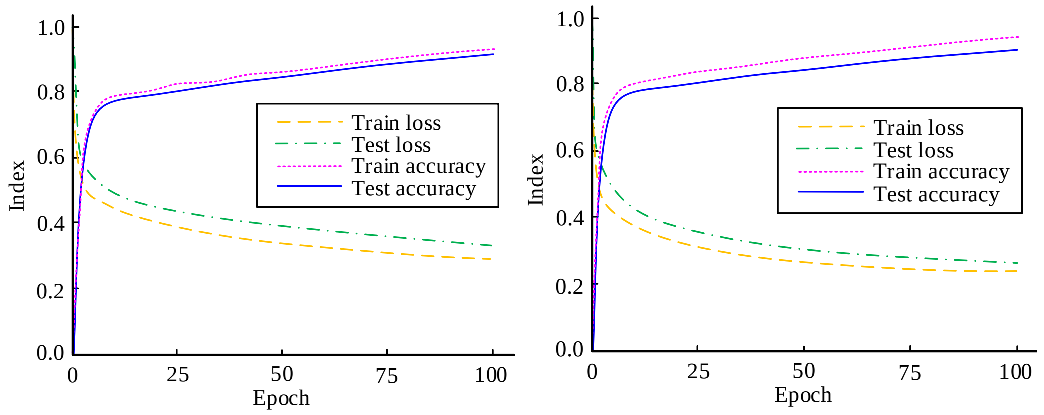


Figure 9: Experimental results of emotion analysis model ablation

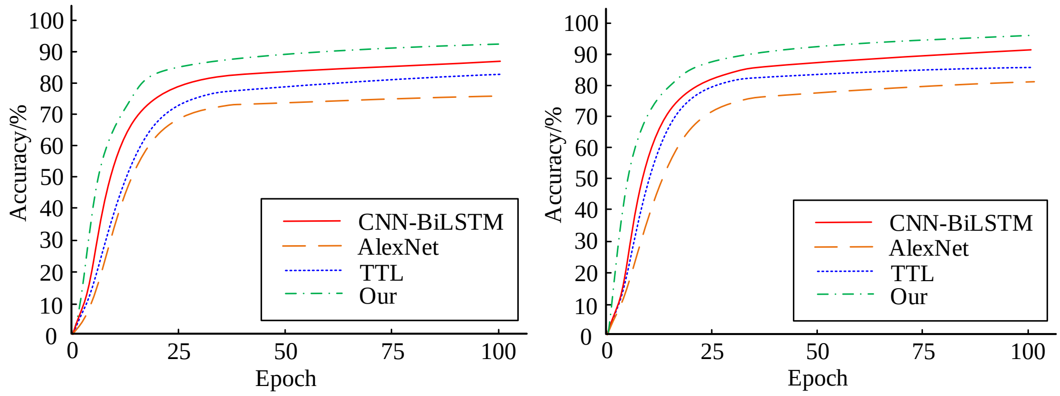
The study uses two datasets for training and testing, respectively, to examine the style migration-based painting sentiment analysis model's performance in sentiment recognition. Figure 10 illustrates how the suggested model's accuracy and loss value changed. The proposed model demonstrates good recognition performance on both datasets, with recognition accuracy above 90%.



(a) MART dataset (b) DeviantArt dataset

Figure 10: Training and testing curves of the model

To further validate the superiority of the style migration-based painting sentiment analysis model, the study compares it with the more advanced convolutional neural network-bidirectional long short term memory (CNN-BiLSTM), AlexNet and two layers transfer learning (TTL). The sentiment recognition accuracy results of the four models are shown in figure 11. The suggested model has a greater accuracy than the comparison model, 93.26%, on the MART dataset, as shown in Fig. 11(a). The suggested model continues to have the highest accuracy (95.02%) in Fig. 11(b) when applied to the deviantArt dataset. The CNN-BiLSTM model comes in second, and the AlexNet model does the worst. Experimental results show that the painting sentiment analysis model based on style migration has high painting sentiment recognition accuracy and demonstrates certain superiority.



(a) MART dataset (b) DeviantArt dataset

Figure 11: The accuracy of emotion recognition for four models

4 Discussion and conclusion

To effectively identify and analyze the PSs of art works, the study constructs a PSCRM based on MFF and a painting sentiment analysis model based on style migration. Experimental results demonstrated that the CA and recall of the proposed PSCRM always remained above 95% in 10 tests, and the AUC value remained above 90%. The full model exhibited the highest CA of 96.68%. The model without Fisher Score had the lowest CA of 87.84%. The full model still had the highest recall of 95.83%, followed by the model with the spateless ballot method with 91.05%. On the WikiArt dataset and the OilPainting dataset, the ROC curves of the proposed model were able to completely cover the results of the other three models with the best overall performance. On the dataset obtained from the style migration data enhancement strategy proposed in the study, all three models performed better in terms of metrics. The complete painting sentiment analysis model achieved the highest recognition accuracy and recall, 95.02% and 93.86%, respectively. The painting sentiment analysis model demonstrated better recognition performance on both datasets, with recognition accuracy above 90%. The test had high comprehensiveness and accuracy. Compared to references [8] and [12], the proposed model had a higher recognition accuracy of over 95% and better recognition performance. This may be attributed to the utilization of the wavelet transform for denoising purposes, coupled with the integration of color, texture, and spatial features inherent to the painting. This approach had the potential to enhance the precision and resilience of the PS classification and recognition model. Like references [7], [9], and [10], this research model not only improved the recognition performance, but also become more complex, resulting in longer running time of the model. However, the running time of this research model remained within an acceptable range, and it continued to offer advantages overall.

In summary, the PSCRM and painting sentiment analysis model constructed by the study demonstrate good classification and recognition performance, and have certain feasibility and effectiveness. However, the study only focuses on the picture composition of the painting among the spatial features when performing style recognition. Therefore, future research should further investigate how to more effectively integrate color, texture, and spatial features in paintings. Deep learning models should be introduced to learn the complex relationships between different features. Moreover, transformers should be used for more advanced feature extraction to classify paintings more accurately.

Funding information

This study is supported by PhD early development program of Hebei Normal University for Nationalities in 2022 under Grant No. DR2022011.

References

[1] Liu S. Research on the classification method of artistic painting image style based on naive Bayesian. International Journal of Information and Communication Technology. 2022, 21(4):398-411. https://doi.org/10.1504/IJICT.2022.126491.

[2] Chung CY, Huang SH. Interactively transforming Chinese ink paintings into realistic images using a border enhance generative adversarial network. Multimedia Tools and Applications. 2023 Mar;82(8):11663-11696. https://doi.org/10.1007/s11042-022-13684-4.

[3] Yu Y, Li D, Li B, Li N. Multi-style image generation based on semantic image. The Visual Computer. 2024, 40(5):3411-3426. https://doi.org/10.1007/s00371-023-03042-2.

[4] Jiang H, Yang T. Research on the extraction method of painting style features based on convolutional neural network. International Journal of Arts and Technology. 2022, 14(1):40-55. https://doi.org/10.1504/IJART.2022.122448.

[5] Zhong Y, Huang X. A painting style system using an improved CNN algorithm. IEIE Transactions on Smart Processing & Computing. 2022, 11(5):332-342. https://doi.org/10.5573/IEIESPC.2022.11.5.332.

[6] Cascone L, Nappi M, Narducci F, Russo SL. Classification of fragments: recognition of artistic style. Journal of Ambient Intelligence and Humanized Computing. 2023, 14(4):4087-4097. https://doi.org/10.1007/s12652-022-04472-x.

[7] Madhu P, Villar-Corrales A, Kosti R, Bendschus T, Reinhardt C, Bell P, Maier A, Christlein V. Enhancing human pose estimation in ancient vase paintings via perceptually-grounded style transfer learning. ACM Journal on Computing and Cultural Heritage. 2022, 16(1):1-7. https://doi.org/10.1145/3569089.

[8] Peng X, Peng S, Hu Q, Peng J, Wang J, Liu X, Fan J. Contour-enhanced CycleGAN framework for style transfer from scenery photos to Chinese landscape paintings. Neural computing and applications. 2022, 34(20):18075-18096. https://doi.org/10.1007/s00521-022-07432-w.

[9] Zhang Q, Xu Y, Zhang J, Tao D. Vitaev2: Vision transformer advanced by exploring inductive bias for image recognition and beyond. International Journal of Computer Vision. 2023, 131(5):1141-1162. https://doi.org/10.1007/s11263-022-01739-w.

[10] Guo MH, Xu TX, Liu JJ, Liu ZN, Jiang PT, Mu TJ, Zhang SH, Martin RR, Cheng MM, Hu SM. Attention mechanisms in computer vision: A survey. Computational visual media. 2022, 8(3):331-368. https://doi.org/10.1007/s41095-022-0271-y.

[11] Jang DK, Park S, Lee SH. Motion puzzle: Arbitrary motion style transfer by body part. ACM Transactions on Graphics (TOG). 2022 Jun 9;41(3):1-6. https://doi.org/10.1145/3516429.

[12] Tang C, Zhao Y, Wang G, Luo C, Xie W, Zeng W. Sparse MLP for image recognition: Is self-attention really necessary? In Proceedings of the AAAI conference on artificial intelligence 2022, 36(2); 2344-2351. https://doi.org/10.1609/aaai.v36i2.20133.

[13] Bayoudh K, Knani R, Hamdaoui F, Mtibaa A. A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets. The Visual Computer. 2022, 38(8):2939-2970. https://doi.org/10.1007/s00371-021-02166-7.

[14] Qi J, Ma L, Cui Z, Yu Y. Computer vision-based hand gesture recognition for human-robot interaction: a review. Complex & Intelligent Systems. 2024, 10(1):1581-1606. https://doi.org/10.1007/s40747-023-01173-6.

[15] Gururaj N, Vinod V, Vijayakumar K. Deep grading of mangoes using convolutional neural network and computer vision. Multimedia Tools and Applications. 2023, 82(25):39525-39550. https://doi.org/10.1007/s11042-021-11616-2.

[16] Ashtiani F, Geers A J, Aflatouni F. An on-chip photonic deep neural network for image classification. Nature. 2022, 606(7914):501-506. https://doi.org/10.1038/s41586-022-04714-0.

[17] Hasanvand M, Nooshyar M, Moharamkhani E, Selyari A. Machine learning methodology for identifying vehicles using image processing. InArtificial Intelligence and Applications 2023, 1(3):170-178. https://doi.org/10.47852/bonviewAIA3202833.

[18] Chunxia T. Research on the Multilevel Security Authorization Method Based on Image Content. Acta Informatica Malaysia. 2017; 1(2):17-19. http//:doi.org/ 10.26480/aim.02.2017.17.19

[19] Deb PP, Bhattacharya D, Chatterjee I, Chatterjee P, Zavadskas EK. An Intuitionistic Fuzzy Consensus WASPAS Method for Assessment of Open-Source Software Learning Management Systems. Informatica. 2023, 34(3):529-556. https://doi.org/10.15388/23-INFOR523.

[20] Ali S, Agrawal J. Automated segmentation of brain tumour images using deep learning-based model VGG19 and ResNet 101. Multimedia Tools and Applications. 2024, 83(11):33351-33370. https://doi.org/10.1007/s11042-023-16828-2.

[21] Wabang K, Nurhayati OD. Application of the naive bayes classifier algorithm to classify community complaints. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi). 2022, 6(5):872-876. https://doi.org/10.29207/resti.v6i5.4498.

[22] Vélez Bedoya JI, González Bedia MA, Castillo Ossa LF, Arango López J, Moreira F. Causal Inference Applied to Explaining the Appearance of Shadow Phenomena in an Image. Informatica. 2023, 34(3):665-677. https://doi.org/10.15388/23-INFOR526.

[23] Shu X, Yang J, Yan R, Song Y. Expansion-squeeze-excitation fusion network for elderly activity recognition. IEEE Transactions on Circuits and Systems for Video Technology. 2022, 32(8):5281-5292. https://doi.org/10.1109/TCSVT.2022.3142771.