IEICE TRANSACTIONS

on Fundamentals of Electronics, Communications and Computer Sciences

DOI:10.1587/transfun.2024SML0003

Publicized:2024/08/21

A PUBLICATION OF THE ENGINEERING SCIENCES SOCIETY
The Institute of Electronics, Information and Communication Engineers
Kikai-Shinko-Kaikan Bldg., 5-8, Shibakoen 3 chome, Minato-ku, TOKYO, 105-0011 JAPAN

LETTER

Aesthetic Evaluation of Chinese Calligraphy Using TabNet: Interpretability and Novel Features for Improved Accuracy

Soh YOSHIDA[†a)], *Member*, Nozomi YATOH[†], *Nonmember*, *and* Mitsuji MUNEYASU[†], *Fellow*

SUMMARY The aesthetic evaluation of Chinese calligraphy, an art form with deep cultural roots and subjective interpretations, poses significant challenges in artificial intelligence. In this paper, we extend the methodology introduced in previous work using TabNet, a deep learning approach, to enhance the accuracy and interpretability of assessing the aesthetic qualities of Chinese calligraphy. Our study incorporates an expanded feature set: we add 10 new characteristics to the previously established 22 global shape features. This comprehensive feature ensemble captures the subtleties of Chinese calligraphy in accordance with its traditional artistic standards. Using TabNet, well known for its interpretability within deep learning frameworks, we aim to predict aesthetic scores with increased precision. We performed a rigorous evaluation using the Chinese Handwriting Aesthetic Evaluation Database. Our approach improved accuracy and elucidated the underlying reasoning behind the model's predictions, thereby enhancing transparency.

key words: *Chinese calligraphy, aesthetic evaluation, TabNet, regression*

1. Introduction

Chinese calligraphy, a significant cultural heritage with a long history [1], is important in artistic research, education [2], and the archaeological examination of ancient text [3]. However, assessing the aesthetic value of Chinese characters objectively, accurately, and interpretably using artificial intelligence is challenging. Research on the aesthetic evaluation of images is aimed at scoring and predicting the impressions they elicit. This field of research, which spans photography [4], Western painting [5], ink painting [6], and web design [7], often uses both handcrafted features and deep learning methodologies [8]. By contrast, the study of Chinese characters primarily relies on handcrafted features. This preference stems from the belief that the aesthetic value of these characters is deeply rooted in traditional Chinese calligraphy principles [9], [10].

Sun *et al.* [11] introduced character features based on these calligraphy principles by combining global shape features and radical layout features to assess aesthetics using the Chinese Handwriting Aesthetic Evaluation Database (CHAED), as illustrated in Fig. 1, and multi-layer perceptron (MLP). However, this approach encounters obstacles in evaluation transparency and capturing fine character nuances. The complexity of machine learning models makes it difficult to discern the effect of specific features on predictions, which limits their practical application. Additionally, global shape features mirror broad calligraphic rules and may overlook subtle stylistic differences crucial to character impressions.

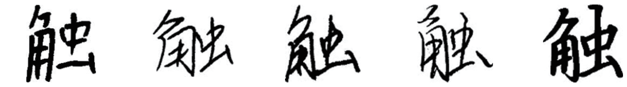


Fig. 1: Examples of bad to good aesthetic qualities of the same Chinese characters in CHAED.

In this paper, we propose an aesthetic evaluation method that leverages TabNet, which is renowned for its interpretability. By selecting key character features for prediction using an attention-based deep learning approach, we aim to enhance both accuracy and interpretability (**Novelty 1**). Furthermore, we refine and introduce new global shape features, thereby enriching the feature set to enable a more comprehensive evaluation (**Novelty 2**). Through rigorous testing and comparison with other regression models, we demonstrated the effectiveness of our approach in evaluating the aesthetic quality of handwritten Chinese characters and addressed the limitations of existing methods.

2. Previous Aesthetic Evaluation Method [11]

In Sun's method, character images are processed as black and white binary images, from which 22 global shape features are derived based on pixel information. Additionally, the characters are divided into radicals and 10 radical layout features are designed to quantify their positional relationships. However, because radical layout features require calibration based on dictionary lookup, they are less suited for processing the diverse range of characters spanning from ancient to modern times. Therefore, we focus on global shape features. These features are established from three perspectives: stability, distribution of whitespace, and density of strokes.

As Chinese characters are typically arranged vertically, an appropriately aligned character must maintain stability on both sides. Each image contains a single character, with the bounding box (BBox) around the character defined by its width W_b and height H_b.

- f_1 (Convex Hull Rectangularity): defined by the ratio $f_1 = P_c/P_b$, where P_c is the perimeter of the convex hull (the smallest polygon enclosing all points of a character) and P_b is the perimeter of the character's BBox. A value of 1 indicates a perfect rectangle and lower values suggest instability.

[†]Faculty of Engineering Science, Kansai University, Yamate-cho 3–3–35, Suita-shi, Osaka, 564-8680 Japan

a) E-mail: sohy@kansai-u.ac.jp (Corresponding author)

- f_2, f_3 (Axis Slope and Intercept): the character's axis is modeled as a line $y = kx + b$. The slope k and intercept b are determined using least squares regression, with k representing the axis tilt ($f_2 = k$) and b normalized by the BBox width ($f_3 = b/W_b$).
- f_4, f_5 (Center of Gravity): the center of gravity (x_g, y_g) is calculated using $(x_g, y_g) = 1/C \sum_{i=1}^{W_b} \sum_{j=1}^{H_b} (x_i, y_j) \times I_{i,j}$, where C is the total number of black pixels in the character image and $I_{i,j} = 1$ for black pixels. The normalized coordinates are $f_4 = x_g/H_b$ and $f_5 = y_g/W_b$.

The distribution of the margins indicates whether the black pixels in the image are dense.

- f_6 (Convexity): a classical measure for the distribution of whitespace, calculated as $f_6 = C/A_{convex}$, where A_{convex} is the area of the convex hull and C is the total count of black pixels.
- f_7 (Convex Hull Cut Ratio): reflects the area ratio of the left part of the convex hull when split by the character axis, determined from f_2 and f_3. It is calculated as $f_7 = A_{left}/A_{convex}$, where A_{left} denotes the area of the left section post-cut.
- f_8–f_{11} (Quadrant Pixel Distribution Ratio): the character's convex hull is divided into four quadrants using the center (x_c, y_c) as the origin. For each quadrant i $(i = 1, 2, 3, 4)$, the ratio of pixel count C_i to the quadrant's convex hull area $A_{convex(i)}$ is computed as $f_{7+i} = C_i/A_{convex(i)}$.
- f_{12}–f_{17} (Mesh Layout): the mesh divides the black pixels of the character image evenly in both the horizontal and vertical directions. A 4 × 4 mesh is used. The i^{th} $(i = 1, 2, 3)$ vertical line's x-coordinate is $x_{vLine(i)}$ and the j^{th} $(j = 1, 2, 3)$ horizontal line's y-coordinate is $y_{hLine(j)}$. For normalization, the positions of the vertical and horizontal lines are divided by H_b and W_b, yielding $f_{11+i} = x_{vLine(i)}/H_b$ and $f_{14+j} = y_{hLine(j)}/W_b$, respectively.

The density of the stroke count is an important feature of Chinese characters, which are composed mainly of straight lines.

- f_{18} (Maximum Fill Ratio): designed based on shape density analysis, this feature fills in the gaps within characters by drawing lines between black pixels at the edges of rows and columns. It describes the distance between strokes and achieves posture invariance by rotating the character image once to fill gaps, calculated as $f_{18} = \max_\alpha C_{gap(\alpha)}/(C_{gap(\alpha)} + C)$, where $\alpha = 1°, 2°, \ldots, 90°$ denotes the rotation angle and $C_{gap(\alpha)}$ is the number of pixels that fill the gaps when the image is rotated by α degrees.
- f_{19}–f_{22} (Pixel Projection Variance): reflects the presence of the four most common types of strokes in Chinese characters: horizontal, vertical, left oblique, and right oblique. They are observed by projecting black pixels onto the x-axis rotated by $\alpha = 0°, 45°, 90°, 135°$, respectively. The variance of the distribution after projection, δ_α, defines these features: $f_{18+i} = \delta_\alpha$ $(i = 1, 2, 3, 4)$.

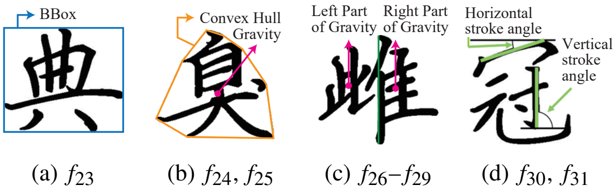
These extracted features serve as inputs for training the aesthetic evaluation model. A regression model using a four-layer MLP, with f_1–f_{22} as input features and the mean square error (MSE) as the loss function, is constructed for aesthetic evaluation. Note that radical layout features were added in the original literature.

3. Proposed Method

3.1 Proposed Global Shape Features

In this study, we enhance our approach by refining four principal global shape features from the original set of 22 using conventional machine learning techniques: extra tree [12], random forest [13], LightGBM [14], and XGBoost [15], along with TabNet. The specifics of our experimental setup are detailed in Section 4.1. Through ensemble evaluation, the axis slope (f_2) emerged as the most significant feature, which highlights the critical role of bilateral symmetry in aesthetic assessments. Next in importance, the convexity measure (f_6) underscored the substantial influence of stroke density on overall impressions. The pixel projection variance at 90° (f_{21}) illuminated the contribution of stroke angle stability, inherent to Chinese characters, to creating a clean and neat impression. Additionally, the center of gravity along the x-axis (f_4) stressed the importance of stability in aesthetic judgment. Leveraging these insights, we introduce 10 new features, as shown in Fig. 2, to provide a more nuanced understanding of character aesthetics.

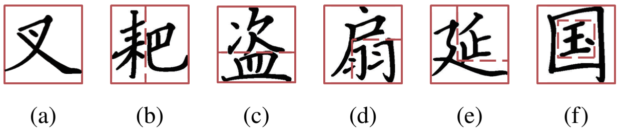
- f_{23} (Aspect Ratio): reflects the width to height ratio of a character's BBox, calculated as $f_{23} = H_b/W_b$. It indicates that well-balanced characters tend to have a near-square BBox.
- f_{24}, f_{25} (Convex Hull Centroid): represents the centroid within a character's convex hull and calculated from the entire convex hull, including margins. After the hull's margins are replaced with black pixels, the centroid (x_{cg}, y_{cg}) is computed using the same procedure as f_4, f_5, and normalized by the BBox dimensions, yielding $f_{23} = x_{cg}/H_b$, $f_{24} = y_{cg}/W_b$.
- f_{26}–f_{29} (Bisected Image Centroids): evaluate symmetry by dividing the image along the character's centroid, and computing the centroids (x_{gright}, y_{gright}) and (x_{gleft}, y_{gleft}) for the right and left halves, respectively. These are normalized by the BBox dimensions to obtain $f_{26} = x_{gright}/H_b$, $f_{27} = y_{gright}/W_b$, $f_{28} = x_{gleft}/H_b$, $f_{29} = y_{gleft}/W_b$.
- f_{30}, f_{31} (Stroke Angles): directly measure the horizontality and verticality of strokes by thinning the character image, detecting the longest vertical and horizontal strokes, and calculating their angles relative to the horizontal line as f_{30} and f_{31}. Characters that lack vertical or horizontal strokes are assigned an initial value of 360°.



(a) f_{23} (b) f_{24}, f_{25} (c) f_{26}–f_{29} (d) f_{30}, f_{31}

Fig. 2: Proposed global character shape features.

- f_{32} (Structure Category): classifies character structure into six types based on radicals, thereby offering simplified structural information without specific radical positioning. Fig. 3 shows the structure categories: (a) single-element characters, (b) characters divisible horizontally with radicals, (c) vertically divisible characters with crown or foot radicals, (d) characters with "hanging" radicals that allow division around and within, (e) characters with "enclosing" radicals distinguishable around and within, and (f) characters with "embracing" radicals distinguishable around and within.



(a) (b) (c) (d) (e) (f)

Fig. 3: Proposed layout categorical feature: f_{32}.

3.2 Aesthetic Evaluation Using TabNet

The TabNet encoder consists of two types of transformers: the feature transformer (FT) layer for feature determination and the attentive transformer (AT) layer for selecting important features. To enhance interpretability, an attention mask (AM) layer calculates and aggregates feature contributions. TabNet operates on sequential decision steps and selects features for each i^{th} step ($i = 1, \dots, N_{steps}$) through masking, therby aggregating processed feature representations for interpretability. N_{steps} represents the total number of steps. The encoding and learning processes from the input to the output layer are detailed as follows:

Input layer: feature vector $f \in \mathbb{R}^{B\times D}$ undergoes normalization via a batch normalization (BN) layer before proceeding to the FT layer, where B represents the batch size and D ($= 32$) is the number of dimensions.

FT layer: comprises four blocks of fully connected (FC) layers, BN layers, and gated linear units (GLUs) [16], where the GLU serves as the activation function that determines feature passage. The first two blocks share weights across decision steps, whereas the last two have independent weights for each step. Each block's output, normalized by $\sqrt{0.5}$, connects to the next block's input and output via skip connections. The FT output splits into $d[i] \in \mathbb{R}^{N_d}$ and $a[i] \in \mathbb{R}^{N_a}$, where N_d and N_a represent each of the dimensions, and they are inputs for the following layers' $(i+1)^{th}$ step, respectively. Note that the FT layer connected after the input layer is used to obtain the initial values, whose outputs are denoted by $d[0]$ and $a[0]$. Counting decision steps starts at the next AT layer.

AT layer: decides which features to select and which to ignore. Following an FC layer and BN layer, a scale is applied. A learnable mask $M[i] \in \mathbb{R}^{B\times D}$ at the i^{th} decision step selects features of f as $M[i] \cdot f$, where $\sum_{j=1}^{D} M[i]_{b,j} = 1$. $M[i]_{b,j}$ represents the (b, j) component of $M[i]$. The scale $P[i] \in \mathbb{R}^{B\times D}$ reflects the emphasis placed on each feature in the previous steps, updated as $P[i] = \prod_{j=1}^{i} P[i](\gamma - M[j])$. Initially, $P[0] = \mathbf{1}^{B\times D}$, where γ is a relaxation parameter. Mask $M[i]$ is determined by $M[i] = \text{sparsemax}(P[i] \cdot h_i(a[i-1]))$, where h_i is a layer that combines the FC and BN layers at step i. Sparsemax [17] is an activation function that adds sparsity to outputs, similar to softmax. The output of the AT layer is then forwarded to the AM layer.

AM layer: for interpretability, the AT layer's generated mask calculates the feature contribution degrees, forming an aggregated feature importance mask. If $M[i]_{b,j} = 0$, the feature dimension j of instance b does not contribute to the i^{th} step. The contribution degree $\eta_b[i]$ is calculated as $\eta_b[i] = \sum_{c=1}^{N_d} \text{ReLU}(d[i]_{b,c})$. The importance mask for aggregated features is represented by $M_{b,j}^{agg} = \sum_{i=1}^{N_{steps}} \eta_b[i] M_{b,j}[i] \Big/ \sum_{j=1}^{D} \sum_{i=1}^{N_{steps}} \eta_b[i] M_{b,j}[i]$ .The output that passes through the AM layer is fed into the FT layer.

Output layer: the output $d[i]$ from the FT layer at the i^{th} step is aggregated using the ReLU activation function as $d = \sum_{i=1}^{N_{steps}} \text{ReLU}(d[i])$. This aggregated output passes through an FC layer to produce the prediction. The proposed method defines the loss function as the MSE, by comparing the actual aesthetic score with the predicted score. Additionally, a regularization term $L_{sparse} = \sum_{i=1}^{N_{steps}} \sum_{b=1}^{B} \sum_{j=1}^{D} -M[i]_{b,j} \log(M[i]_{b,j} + \epsilon)/(N_{steps} \cdot B)$ is included with a regularization parameter λ_{sparse} to enhance the sparsity of feature selection, where ϵ is a small number added to maintain numerical stability.

Tabular self-supervised learning: before TabNet is trained for aesthetic score prediction, a self-supervised pre-training phase is conducted to enhance predictive performance. In this phase, the encoder's output is processed by the FT and FC layers. A binary mask $S \in \mathbb{R}^{B\times D}$, based on Bernoulli sampling, is applied to the input feature vector f for reconstruction, where each $S_{b,j} \in \{0, 1\}$ indicates whether a feature is retained or masked. The reconstruction error is defined as $L_{recon} = \sum_{b=1}^{B} \sum_{j=1}^{D} |(\hat{f}_{b,j} - f_{b,j})| \cdot S_{b,j} / \sqrt{\sum_{b=1}^{B} (f_{b,j} - 1/B \sum_{b=1}^{B} f_{b,j})}|^2$, where $\hat{f}_{b,j}$ and $f_{b,j}$ represent the predicted and input values of the feature vector at position (b, j), respectively.

4. Experiments

4.1 Dataset and Experimental Setup

Dataset. CHAED reported in the literature [11] was used to validate the effectiveness of the proposed method through comparative experiments. CHAED comprises 100 types of characters, each with 10 handwritten Chinese characters representing different impressions. Each of 1,000 character images is assigned an aesthetic impression score based on surveys from 33 evaluators via crowdsourcing. After 10 unusable images were excluded, 990 images were processed to extract global shape features, with 891 designated as training data and 99 as test data.

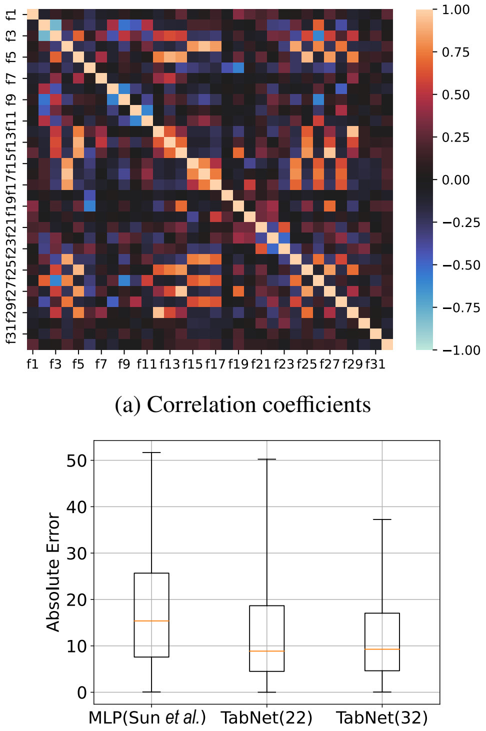
Experimental condition. In accordance with the size of the dataset and the dimensions of the proposed features, the TabNet configuration included the settings of $N_d = 8$, $N_a = 16$, $N_{steps} = 3$, $B = 16$, $\gamma = 1.5$, and $\lambda_{sparse} = 1.0 \times 10^{-3}$. The Adam optimizer was used with an initial learning rate of 0.02, which was decreased by 0.1 every 10 epochs if there was no reduction in loss. The comparative methods were linear regression [18], k–nearest neighbors (k–NN) [19], support vector regression (SVR) [20], extra trees, random forest, LightGBM, and XGBoost, with optimal parameters determined through grid search. The mean absolute error (MAE), which is the average absolute difference between the predicted values and actual scores within the dataset, was used as the evaluation metric.

4.2 Results

Table 1 shows the results, which indicate that TabNet, using 32 global shape features, achieved the lowest MAE. This outcome suggests that the proposed method most accurately reflects human subjective evaluations. Specifically, the addition of features improved the predictive accuracy of the proposed method. However, accuracy declined for some comparative methods. To analyze this, Fig. 4(a) illustrates the strength of the relationship between two variables based on Pearson's correlation coefficient. Notably, features f_{24}–f_{29}, designed based on insights from the existing 22 features, showed significant correlations, which led to prediction instability caused by multicollinearity [21]. By contrast, feature selection techniques such as TabNet and decision tree variants proved effective when the proposed features were added. Specifically, TabNet uses AT layers to select relevant features at each decision step, which mitigates the effect of multicollinearity. Decision tree-based methods such as extra tree and random forest make splitting decisions based on individual features, which makes them less sensitive to correlations between features. This explains why these methods improved performance when the expanded feature set was used, whereas models such as linear regression, k-NN, and MLP either maintained or decreased performance. Fig. 4(b) shows a comparison of the distribution of scoring errors between previous methods and the proposed method, and demonstrates the enhanced robustness with our approach because median predictions were within a 10-point error margin and the outliers are significantly suppressed.

Table 1: Comparison of MAE for various methods.

Method \ Feature	f_1–f_{22}	f_1–f_{32}
Linear regression	16.69	16.61
k–NN	17.62	18.72
SVR	16.59	15.89
Extra tree	15.91	15.33
Random forest	16.39	15.71
LightGBM	16.61	15.89
XGBoost	17.03	16.76
MLP (Sun *et al.*) [11]	17.85	19.96
TabNet (**ours**)	12.58	11.96



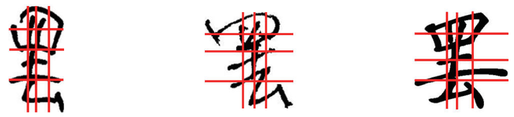
(a) Correlation coefficients

(b) Absolute error

Fig. 4: Statistical analysis.

Fig. 5 shows the character images 罢, along with the corresponding scores and features that contributed to predictions, as identified by AM. The contributing features (f_5, f_{12}, f_{17}, f_{21}, f_{24}, f_{27}) focus on stroke density and the character's center of gravity, which indicates that character balance influences the aesthetic impression. In this case study, we indicated features f_{12}–f_{17} using red lines on the different images to enable us to discuss character balance. We analyze three characters with low, medium, and high aesthetic scores to provide a comprehensive view of how the model evaluates characters across different quality levels and to identify which features are consistently important or change in importance based on the aesthetic score.

- The consistent importance of f_{21} (pixel projection variance, $\alpha = 90°$) and f_{24} (x-coordinate of the convex hull's center of gravity) across all score ranges suggests that they have a fundamental role in evaluating the structure of 罢. It is likely that f_{21} assesses the arrangement



(a) (15.15, **33.75**) f_{17},f_{21},f_{24},f_{27}

(b) (49.96, **41.10**) f_{17}, f_{21}, f_{24}, f_{27}

(c) (96.96, **85.97**) f_5,f_{12},f_{21},f_{24}

Fig. 5: Case study: the caption indicates the target score, predicted score (bold), and contributing features.

of vertical strokes in 罒 and 去, whereas f_{24} evaluates the left-right balance. The prominence of f_{24} indicates that left-right symmetry and the balance between 罒 and 去 significantly affect the aesthetic evaluation. This feature is likely to reflect the balance between the mesh-like structure of 罒 and the asymmetrical shape of 去.

- For low and medium scores, f_{17} (mesh layout of the horizontal line) and f_{27} (y-coordinate of the right half's center of gravity) appear to be important, which suggests that they have role in identifying character imbalance.
- For high scores, f_5 (y-coordinate of the center of gravity) and f_{12} (mesh layout of the vertical line) are significant. Given that Chinese character strokes typically progress from top-left to bottom-right, this suggests that, to be aesthetically pleasing, 罢 characters need to have on overall vertical balance and appropriate left-side arrangement. The well-structured 罒 and harmoniously arranged 去 are likely to contribute to high aesthetic evaluations.

This analysis also highlights a limitation of the proposed method: its tendency to make conservative predictions, particularly at the score extremes. This means that the model may lean toward the dataset's average for very high or low true scores, thereby leading to prediction errors. Addressing this limitation requires refining the model's sensitivity to features in future work, particularly for accurately handling extreme cases.

5. Conclusions

In this paper, we presented a novel method to assess the aesthetic impressions of Chinese characters leveraging TabNet. Our experimental findings underscore the proposed method's effectiveness by demonstrating notable performance and interpretability.

Acknowledgment

This work was supported by JSPS, KAKENHI Grant Number 22K18007, Japan. We thank Edanz (https://jp.edanz.com/ac) for editing a draft of this manuscript.

References

[1] Y. Xu and R. Shen, "Aesthetic evaluation of chinese calligraphy: a cross-cultural comparative study," Current Psychology, vol.42, pp.23096–23109, 2023.

[2] M. Wang, Q. Fu, X. Wang, Z. Wu, M. Zhou, *et al.*, "Evaluation of chinese calligraphy by using dbsc vectorization and icp algorithm," Mathematical Problems in Engineering, vol.2016, 2016.

[3] T. Fujita, "A basic consideration for the handwrighting analysis of the han woodslips," Essays on the Occation of the 70th Anniversary of the Institute of Oriental and Occidental Studies, Kansai University, pp.357–376, 2013.

[4] L. Li, H. Zhu, S. Zhao, G. Ding, H. Jiang, and A. Tan, "Personality driven multi-task learning for image aesthetic assessment," Proceedings of the IEEE International Conference on Multimedia and Expo, pp.430–435, 2019.

[5] C. Li and T. Chen, "Aesthetic visual quality assessment of paintings," IEEE Journal of Selected Topics in Signal Processing, vol.3, no.2, pp.236–252, 2009.

[6] A. Sartori, V. Yanulevskaya, A.A. Salah, J. Uijlings, E. Bruni, and N. Sebe, "Affective analysis of professional and amateur abstract paintings using statistical analysis and art theory," ACM Transactions on Interactive Intelligent Systems, vol.5, no.2, 2015.

[7] J. Zhang, Y. Miao, J. Zhang, and J. Yu, "Inkthetics: A comprehensive computational model for aesthetic evaluation of chinese ink paintings," IEEE Access, vol.8, pp.225857–225871, 2020.

[8] K. Saira, U. Muhammad, and H. Ullah, "A survey of hand crafted and deep learning methods for image aesthetic assessment," CoRR, vol.abs/2103.11616, 2021.

[9] W. Li, Y. Song, and C. Zhou, "Computationally evaluating and synthesizing chinese calligraphy," Neurocomputing, vol.135, pp.299–305, 2014.

[10] M. Sun, X. Gong, H. Nie, M.M. Iqbal, and B. Xie, "Srafe: Siamese regression aesthetic fusion evaluation for chinese calligraphic copy," CAAI Transactions on Intelligence Technology, vol.8, no.3, pp.1077–1086, 2023.

[11] R. Sun, Z. Lian, Y. Tang, and J. Xiao, "Aesthetic visual quality evaluation of chinese handwritings," Proceedings of the International Conference on Artificial Intelligence, pp.2510–2516, 2015.

[12] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," Machine learning, vol.63, pp.3–42, 2006.

[13] R. Genuer, J.M. Poggi, R. Genuer, and J.M. Poggi, Random forests, Springer, 2020.

[14] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," Advances in Neural Information Processing Systems, 2017.

[15] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.785–794, 2016.

[16] S. Narang, H.W. Chung, Y. Tay, W. Fedus, T. Fevry, M. Matena, K. Malkan, N. Fiedel, N. Shazeer, Z. Lan, Y. Zhou, W. Li, N. Ding, J. Marcus, A. Roberts, and C. Raffel, "Do transformer modifications transfer across implementations and applications?," CoRR, vol.abs/2102.11972, 2021.

[17] A.F.T. Martins and R.F. Astudillo, "From softmax to sparsemax: a sparse model of attention and multi-label classification," Proceedings of the International Conference on Machine Learning, pp.1614–1623, 2016.

[18] F. Galton, "Regression towards mediocrity in hereditary stature," The Journal of the Anthropological Institute of Great Britain and Ireland, vol.15, pp.246–263, 1886.

[19] A. Mucherino, P.J. Papajorgji, and P.M. Pardalos, k-Nearest Neighbor Classification, Springer New York, 2009.

[20] H. Drucker, C.J.C. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," Proceedings of the International Conference on Neural Information Processing Systems, pp.155–161, 1996.

[21] G. Smith, "10-multiple regression," in Essential Statistics, Regression, and Econometrics (Second Edition), pp.301–337, Academic Press, 2015.