



# OPEN Predict the writer's trait emotional intelligence from reproduced calligraphy

Ruimin Lyu<sup>1,2</sup>, Wen Sun<sup>1</sup>, Yongle Cheng<sup>1</sup>, Yifei Shi<sup>1</sup>, Ning Wang<sup>1</sup>, Joydeep Bhattacharya<sup>3,4</sup> & Guoying Yang<sup>1,3</sup>✉

Trait emotional intelligence (EI) describes an individual's ability to control their emotions. In Chinese calligraphy, there is a saying that “the character reflects the person.” This raises a hypothesis: is it possible to predict a writer's trait EI from their calligraphy reproductions? To test this hypothesis, we propose a predictive method that integrates deep learning with aesthetic features of calligraphy. First, a hard pen calligraphy reproduction dataset was constructed, consisting of 48,826 reproduced characters from 191 participants, with corresponding trait EI scores and reproduction skill score ratings. A Siamese neural network was then used to extract deep feature differences between the reproduction characters and the reference characters, which were further combined with handcrafted features for regression-based predictions. Experimental results show that, using Mean Absolute Error (MAE), Mean Squared Error (MSE) and Pearson Correlation Coefficient (PCC) as evaluation metrics, this method's ability to predict the writer's trait EI from calligraphy reproductions (MAE: 0.463, MSE: 0.462, PCC: 0.730) significantly outperforms human evaluative abilities (MAE: 1.006, MSE: 1.740, PCC: 0.145), confirming that calligraphy reproductions indeed contain latent information about the writer's trait EI.

**Keywords** Trait EI prediction, Psychological projection experiment, Calligraphy psychology, Siamese neural network, Computational aesthetics, Computer assisted assessment

Trait Emotional Intelligence (EI) refers to an individual's ability to perceive, express, regulate, and utilize emotions<sup>1</sup>. This personality trait is generally divided into four dimensions: Well-being, self-control, emotion, and socialability<sup>2,3</sup>. Trait EI has been associated with a variety of outcomes, including decision making, interpersonal functioning, and mental health. Current methods for measuring the EI trait rely mainly on self-report questionnaires<sup>4,5</sup>, which are limited by issues of subjectivity, social desirability bias, and constrained design features<sup>6</sup>, therefore making it difficult to fully capture the complexity and subtle variations of this emotional trait.

Chinese calligraphy, through concepts such as “writing as the painting of the heart” and “the character reflects the person,” suggests that calligraphy characters may reflect the inner emotions and psychological traits of the writer. Based on this, we propose the hypothesis: **Can calligraphy characters, particularly reproduced characters, reveal the trait emotional intelligence (EI) of the writer?** The theory of “psychological projection tests” in psychology offers an analogous perspective. Psychological projection tests, such as the Rorschach inkblot test<sup>7</sup>, the thematic apperception test<sup>8</sup>, and the sentence completion test<sup>9</sup>, induce individuals to unconsciously project emotions, motivations, and personality traits from their subconscious onto external stimuli by strictly controlling the test procedures, themes, and content. The reactions of participants to these controlled stimuli, especially the differences in their responses, can reveal their underlying psychological traits. Similarly, the practice of “reproduction” in calligraphy, as a distinct form of writing, shares similarities with the principles of psychological projection tests. In the process of reproduction, the writer is required to follow a fixed form and strokes but may unconsciously incorporate personal emotions, style, and understanding, thereby reflecting their psychological state in the subtle differences from the reference characters. Compared to freehand writing, reproduction provides a strictly controlled framework in which the writer is more likely to unconsciously express underlying personality traits. Therefore, we hypothesize that reproduced works are likely to accurately reflect the trait emotional intelligence of the writer. For example, Figure 1 shows the results of three writers reproducing the same character by Yan Zhenqing. The red represents the reference character. The blue indicates the areas of difference between the reproduced character and the reference character, and the black

<sup>1</sup>Jiangnan University, Wuxi 214122, China. <sup>2</sup>Jiangsu Key University Laboratory of Software and Media Technology under Human-Computer Cooperation, Jiangsu, Wuxi, China. <sup>3</sup>Goldsmiths, University of London, London, UK, England. <sup>4</sup>Academy of Music, School of Creative Arts, Hong Kong Baptist University, Hong Kong, China. ✉email: guoyingyang@jiangnan.edu.cn



**Fig. 1.** Do the differences in the reproductions by three writers reveal their trait emotional intelligence? (Red: reference character, Blue: different areas of the reference and reproduced character, Black: overlap areas of the reference and reproduced character).

represents the overlapping parts of the two. Although the three writers reproduced the same character, their works show noticeable differences. Based on the preceding discussion, our core hypothesis can be formulated as a modeling problem: **Through data modeling, can we achieve accurate prediction of a calligraphy learner's trait emotional intelligence by analyzing the visual discrepancies between reference calligraphic characters and their reproduced counterparts?**

Although handwriting analysis has historically been used to infer psychological traits from handwriting features such as shape, strokes, curves, or spatial layout<sup>10,11</sup>, such approaches have lacked standardization and empirical rigor, limiting their acceptance in mainstream psychology. Recent advances in machine learning have enabled more robust and scalable analysis of handwriting features, with applications in detecting depression<sup>12</sup>, anxiety<sup>13</sup>, and personality traits such as the Big Five<sup>14</sup>. However, the relationship between handwriting, particularly in the constrained format of calligraphic reproduction, and the EI trait remains underexplored.

Current modeling technologies also exhibit inherent limitations. While existing handwriting analysis methods have primarily relied on handcrafted feature extraction and traditional machine learning algorithms, these approaches often struggle to capture nuanced patterns in handwriting due to their limited representational capacity. Deep learning, on the other hand, excels at automatically learning discriminative features from raw data, yet its inherent opacity poses challenges for psychological research, where interpretability is crucial.

To bridge this gap, we propose a hybrid framework that synergizes the strengths of deep learning with domain-specific handcrafted features, enabling both high-dimensional pattern recognition and meaningful feature interpretation. Our method leverages a Siamese neural network to extract differences between reproduced and reference characters and integrates these with 48 handcrafted features derived from calligraphic aesthetics. These features are used to train a regression model on LightGBM, enabling prediction of trait EI scores across its multiple dimensions. This method enhances the depth of feature extraction and improves the model's interpretability through handcrafted features, thereby further increasing prediction accuracy. To train the model, we constructed a large-scale dataset comprising 191 participants, 605 reference characters, and a total of 48,826 reproduced characters, with each participant's trait EI scores measured by questionnaire and reproduction skill ratings from calligraphy experts. Modeling results validate that calligraphy reproduction characters indeed contain implicit information about the writer's trait EI, thus supporting the hypothesis of this study.

## Related works

Trait emotional intelligence (EI), a relatively stable internal characteristic, has been shown in multiple studies to significantly impact individuals' daily lives. Smith et al.<sup>15</sup> indicated that individuals' multidimensional assessment of events or situations can trigger unique emotional responses, further influencing their behavior and decision-making abilities. The study by Petrides et al.<sup>16</sup> confirmed the validity of the trait EI concept and found that individuals with high emotional intelligence demonstrate significant advantages in emotional processing and responsiveness. Karimi<sup>17</sup> specifically emphasized the role of trait EI in enhancing job performance, while Austin et al.<sup>18</sup> found a negative correlation between trait EI and psychological distress and depressive symptoms, suggesting that higher EI contributes to maintaining mental health. Together, these studies underscore the practical significance of research on trait emotional intelligence.

However, traditional EI measurements mainly depend on scales and surveys. Although Bru-Luna et al.<sup>19</sup> enhanced scale reliability by assessing internal consistency and proposed a linear model-based tool for EI measurement, and Schutte et al.<sup>20</sup> examined the EI model proposed by Salovey and Mayer, which covers key areas like emotion appraisal, regulation, and application in problem-solving, these questionnaire-based studies still encounter methodological challenges. Surveys are particularly vulnerable to the influences of respondent subjectivity and social desirability biases. Despite Siegling's<sup>21</sup> efforts to enhance EI assessment tools, no comprehensive method has yet fully addressed these limitations.

In addition to questionnaires, projective tests are widely used in psychological assessments, with writing and drawing being common forms of projection. Buck<sup>22</sup> proposed the "House-Tree-Person" drawing test as a projective method, analyzing drawing content to reflect an individual's emotional and psychological state. Handler<sup>23</sup> further analyzed the effectiveness of this method in revealing emotional and psychological traits, while Guo et al.<sup>24</sup> suggested its utility in screening and classifying mental disorders. These studies have not fully recognized that writing, especially in the form of reproducing, can serve as an effective projective technique. Research specifically on using projective methods to assess EI is scarce. However, the use of drawing and writing in previous projective experiments suggests the potential feasibility of "reproducing" as a projective technique.

In recent years, the rapid development of machine learning and deep learning technologies has significantly progressed psychological assessment. Dwyer et al.<sup>25</sup> highlighted that machine learning can effectively process multidimensional correlated data and shows considerable potential for predicting individual psychological traits. Mekhaznia et al.<sup>26</sup> employed artificial neural networks to extract and classify features from handwritten samples to identify various personality traits. Zeng et al.<sup>27</sup> utilized the LightGBM model to classify drivers' psychological states, while Li<sup>28</sup> combined the BERT model with LightGBM to predict users' emotional tendencies. In the area of trait emotional intelligence (EI), the study by Dollmat et al.<sup>29</sup> demonstrate the substantial potential of machine learning in EI research. Nevertheless, no studies have yet combined deep learning and machine learning to investigate the relationship between trait EI and individual behavior, although existing research reveals the potential of applying machine learning in detecting trait EI.

Zhang et al.<sup>30</sup> utilized CNN to extract deep features from Chinese ink paintings, achieving an effective quality assessment of these artworks. Yan et al.<sup>31</sup> employed VGG16 to capture primary features in calligraphy, developing a Siamese network for evaluation, though the absence of handcrafted features limited the model's interpretability. Sun et al.<sup>32</sup>, effectively incorporated handcrafted features with Siamese neural network to assess the aesthetics of calligraphy. These studies indicate that the Siamese neural network structure is especially suitable for data involving reference characters and reproduced characters, as is often the case with calligraphy assessment. Handcrafted features greatly enhance model interpretability, given the unique demands of calligraphy evaluation. Therefore, the combination of Siamese neural networks with handcrafted features is the most suitable approach for this study, while models with other structures, despite their advantages, lack compatibility with our specific research context.

In light of the specific characteristics of this study, we propose a new method for the task of evaluating calligraphy reproductions. This approach combines deep features extracted by a Siamese neural network with handcrafted features, utilizing LightGBM for regression prediction. Deep learning is used to extract deep feature differences between reference characters and reproduced characters, which are then complemented by handcrafted features to minimize information loss and enhance the model's predictive accuracy. This approach not only addresses the limitations of current methods in feature extraction and information integration but also provides new insights and applications for interdisciplinary research between calligraphy and psychology.

## Methods

**Written informed consent:** The experimental protocol of this study has been approved by the Medical Ethics Committee of Jiangnan University (Reference Number: JNU202409RB0029).

All methods of this study were conducted according to the guidance and regulations of the Medical Ethics Committee of Jiangnan University. The study was performed in accordance with the Declaration of Helsinki.

All methods of this study have obtained the informed consent of all subjects.

## Experimental framework

In the calligraphy reproducing behavior, which aims to reproduce the reference as accurately as possible, is the trait emotional intelligence(EI) of the writer projected into the reproducing work? To verify the hypothesis, the experimental framework of this study is shown in Figure 2

## Handwriting calligraphy reproduction dataset

In this study, a dataset was constructed of hard pen calligraphy reproductions, including reference characters, participants' reproductions, expert ratings of reproduction skill scores, and survey results regarding participants' trait emotional intelligence (EI).

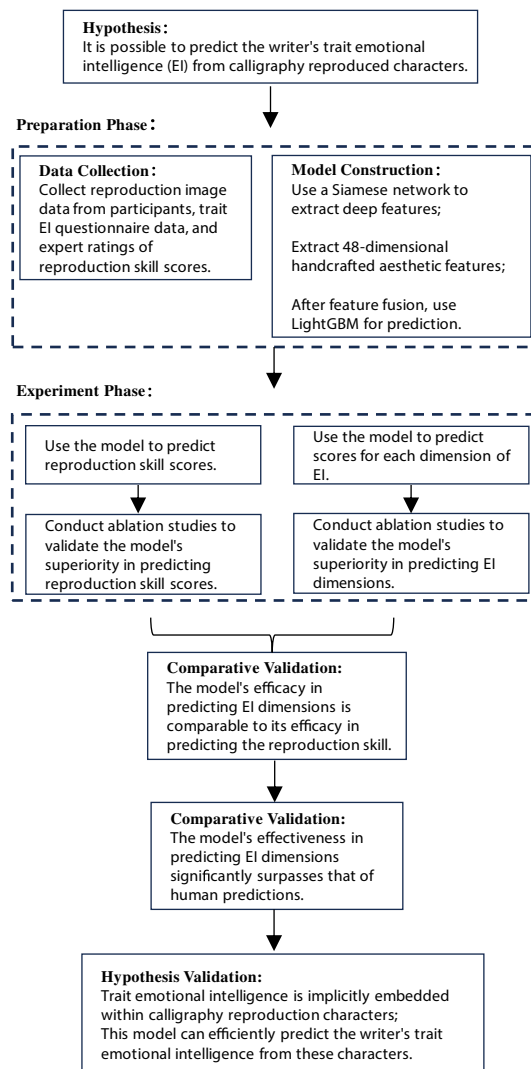
## Data collection

A total of 55 commonly used Chinese characters were selected as reproduction references, reflecting various stroke complexities and common structures (Figure 3). These characters were produced by 11 renowned calligraphers, each with a unique style: Chu Suiliang, Liu Gongquan, Ouyang Xun, Emperor Huizong of Song, Wang Chong, Wang Xizhi, Wen Zhengming, Yan Zhenqing, Yu Shinan, Zhao Mengfu, and Zhong Shaojing. In total, there are 605 reference characters, ensuring both diversity and representativeness in the dataset.

Data collection occurred in university calligraphy classes, where participants were randomly assigned works by calligraphers as reference templates. Reproduction tasks were conducted biweekly, and the entire data collection process spanned 8 weeks.

A total of 191 participants were recruited for the study, all of whom were students enrolled in hard pen calligraphy classes. All the participants were first-year undergraduate students majoring in Teacher Education and Educational Technology. The average age of the participants was 18 years, comprising 141 females and 50 males. Due to the specific demands of their academic programs, these students are expected to possess relatively strong handwriting skills. They are required to pass a standardized handwriting proficiency test by their second year, and therefore engage in additional handwriting training beyond the regular curriculum. We also surveyed their prior calligraphy experience: 51% had received formal extracurricular calligraphy training before the age of 10, and 18% between the ages of 10 and 15. Combined with qualitative evaluations from their calligraphy instructors, these findings suggest that the participants' handwriting abilities and experience are slightly above the population average, though significantly below the level of professional calligraphy students. Thus, they serve as a reasonable representation of the general, non-expert population.

Throughout a 8-week hard-pen calligraphy course (with sessions held biweekly), participants' calligraphy reproduction samples were collected at 2-week intervals. At the beginning of the course, each participant selected a preferred model script—authored by one of eleven calligraphers—from a set of imitation practice



**Fig. 2.** Research design.

sheets we designed. This chosen exemplar served as their consistent training material throughout the semester-long copying practice. Through electronic scanning and preprocessing, a total of 48,826 images of reproduced characters were collected. In the first class, all participants completed an online EI questionnaire<sup>33</sup> consisting of 30 questions. This questionnaire, based on a 7-point scale, assessed participants' EI levels along with four EI sub-dimensions: Well-being, Self-control, Emotionality and Sociability.

After collecting the calligraphy reproduction data, we invited three calligraphy experts to evaluate the reproduction skill score using a 7-point scale. The experts based their ratings on a comparison between the reference characters and the reproduced characters. Both the ratings and the questionnaire results were standardized.

Table 1 displays samples from the dataset. The first column shows the reference characters, while the second column presents the participants' reproductions. Each pair of reference characters and reproduced characters includes the expert rating as well as the participant's scores across each EI dimension. Scores close to zero indicate moderate ability, higher positive values signify stronger ability, and negative values reflect weaker ability.

### Image preprocessing

To address the variations in size and positioning among the reproduced characters that impact the overall quality of the dataset, a preprocessing pipeline for images was established in this study to ensure the consistency of all characters. This pipeline adheres to the following four criteria:

- All character images must be binarized.
- All character images must have standardized lengths or widths.
- All character images must have aligned centroids.
- All character images must maintain consistent dimensions.

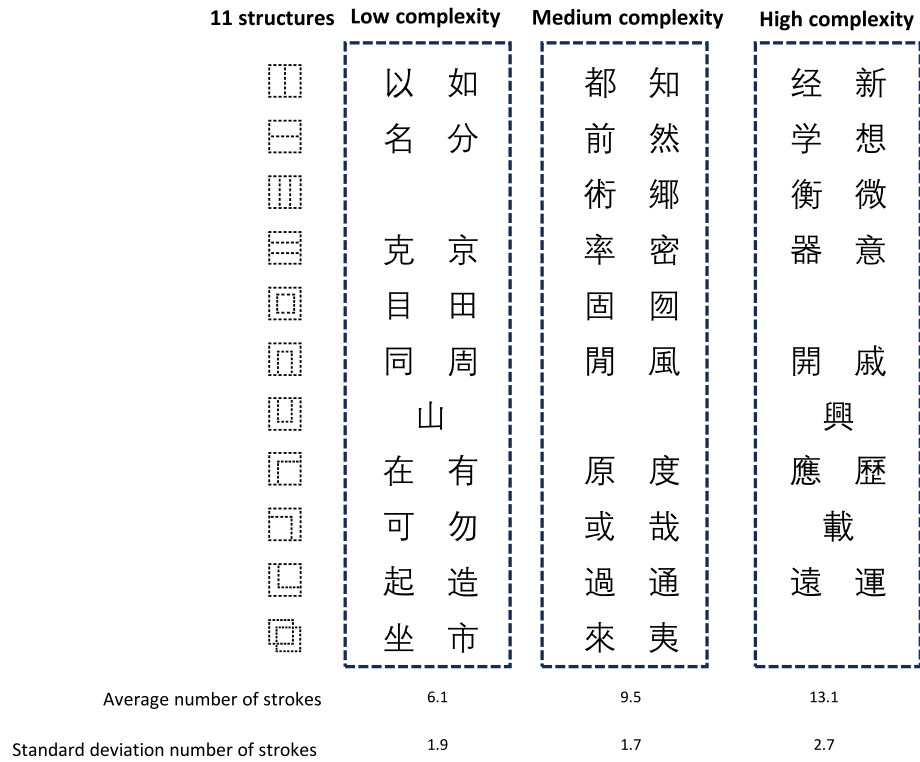


Fig. 3. The pre-selected 55 high-frequency Chinese characters have different stroke numbers, covering different levels of complexity and 11 common structures.

Reference	Reproduced	Reproduction skill	EI	Well-being	Self-control	Emotionality	Sociability
然	然	-0.271	1.818	1.228	1.342	2.404	0.803
經	經	-1.171	0.819	-0.064	1.081	1.128	0.211
意	意	0.552	-0.291	-0.064	0.558	-0.785	-0.085
周	周	1.229	-0.513	-0.433	0.034	-0.625	-1.270
器	器	2.352	-0.679	-0.802	-0.227	-0.147	-1.270

Table 1. Part of the data set images and corresponding z-scores of each dimension sample display. Scores close to zero indicate moderate ability, higher positive values signify stronger ability, and negative values reflect weaker ability.

### Cropping and binarization

The first step in preprocessing involves binarizing the scanned images (Figure 4(a)). This is done using a threshold value of 127 to eliminate background noise and highlight the ink strokes. Next, the images are cropped using the minimum bounding rectangle, determined by the black pixels. This step preserves the character information while removing any extraneous noise. Figure 4(b) illustrates the cropping effect, where the green box represents the minimum bounding rectangle.

### Centroid alignment

The centroid of a character is calculated using the following formula:

$$(x, y) = \frac{1}{N} \sum_{i=1}^{\text{width}} \sum_{j=1}^{\text{height}} (x_i, y_j) I_{ij}$$

$N$  represents the total number of pixels within the minimum bounding rectangle, while  $(x_i, y_j)$  denotes the pixel coordinates, and  $I_{ij}$  signifies the binarized pixel value. As shown by the red dot in Figure 4(b), this dot marks the centroid of the black pixels within the minimum bounding rectangle. Each adjusted character is then positioned on a 224x224 pixel white background, ensuring the character's centroid aligns with the center of the canvas, as shown in Figure 4(c).

### z-Score transformation

Although both the expert ratings and the EI questionnaire scores use a 7-point scale, their distributions and contexts differ significantly. To facilitate the comparison of model predictions regarding reproduction skill and each EI dimension, both types of scores were converted into z-scores. The z-score is calculated using the following formula:

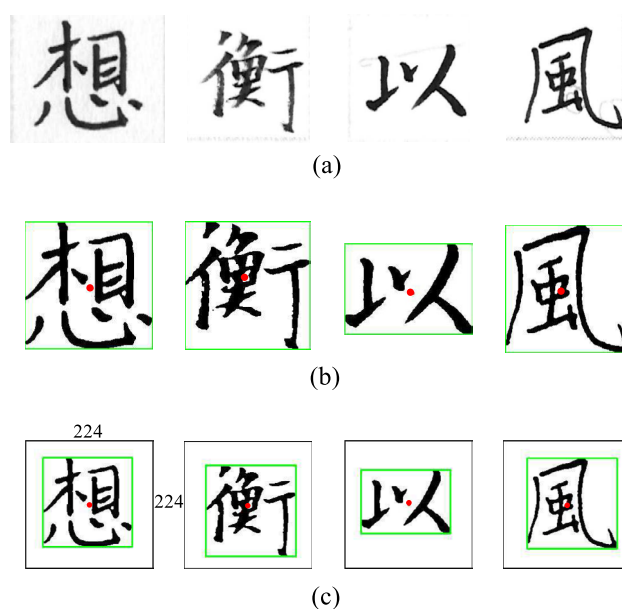
$$z = \frac{(x - \mu)}{\sigma}$$

Where  $x$  is the original score,  $\mu$  is the mean and  $\sigma$  is the standard deviation.

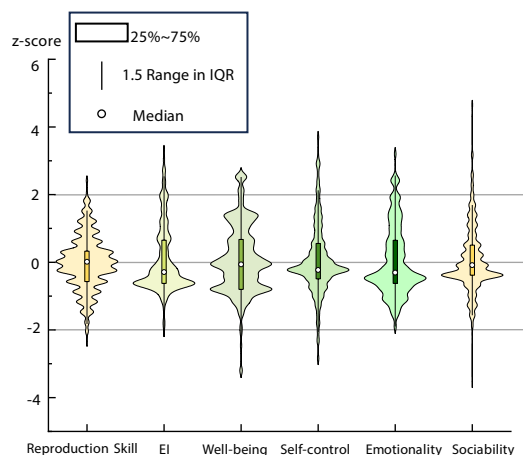
In this study, the z-scores for EI dimensions were applied to all reproduced characters based on the participant's score, while the z-scores for the reproduction skill were calculated individually for each task. A higher positive z-score indicates stronger ability, while a negative value indicates weaker ability. Figure 5 presents the data distribution of each dimension after z-score transformation through violin plots.

### Model construction

This section explores the construction of a model that combines handcrafted features with deep learning techniques for evaluating calligraphy reproductions. This evaluation requires simultaneous processing of both the reference characters and the reproduced characters to quantify their differences. A Siamese neural network<sup>34</sup> is well-suited for this task, as it processes multiple image inputs in parallel while sharing weights. After extracting the deep feature differences between the reference and reproduced characters, these differences are combined with the handcrafted feature differences. Subsequently, regression analysis is conducted using LightGBM to



**Fig. 4.** The preprocessing of image data in data set.



**Fig. 5.** The violin diagram for score distribution of each dimension data after conversion to z-score.

predict the reproduction skill score and the dimensions of EI, respectively. The overall structure of the model is illustrated in Figure 6.

### Deep feature extraction based on Siamese neural network

The model adopts a Siamese neural network structure, in which both the reference and reproduced characters are input simultaneously through a shared backbone network. This backbone network extracts key features from the characters and reduces their dimensionality to 50. The dimensionality is reduced to 50 to match the similar feature quantity of the subsequent 48 handcrafted features.

The deep feature differences between the reference and reproduced characters are calculated using a subtraction operation. The regression layer then produces the prediction results of the Siamese network. The entire network calculates the loss through the loss function and optimizes the network via backpropagation, ensuring it accurately captures the difference in deep features between the two characters.

### Extract 48-dimensional handcrafted feature differences

To ensure the comprehensiveness of feature representation in this study and address potential limitations of deep features in adequately capturing aesthetic and psychological nuances embedded in calligraphic works, this study selects 48 handcrafted features based on research in computational aesthetics and computer vision<sup>35–41</sup>. These features are utilized to quantify the aesthetic characteristics of characters. Below is a brief description of the 48 handcrafted features:

*h1**h2*: The length and width of the minimum bounding rectangle of the character.

*h3*: The aspect ratio of the character's minimum bounding rectangle.

*h4*: Rectangularity, defined as the ratio of the perimeter of the character's convex hull to the perimeter of its minimum bounding rectangle.

*h5**h6*: The values of the major and minor axes of the character's minimum enclosing ellipse.

*h7*: Circularity, representing the degree to which the character's convex hull approximates a circle. The formula is below:

$$h_7 = \frac{4\pi C}{P_{con}^2}$$

where  $P_{con}$  is the perimeter of the character's convex hull, and  $C$  is the character area.

*h8*: Eccentricity of the character's enclosing ellipse, calculated by the formula below:

$$h_8 = \sqrt{1 - \left(\frac{b}{a}\right)^2}$$

where  $a$  and  $b$  are the values of the ellipse's major and minor axes.

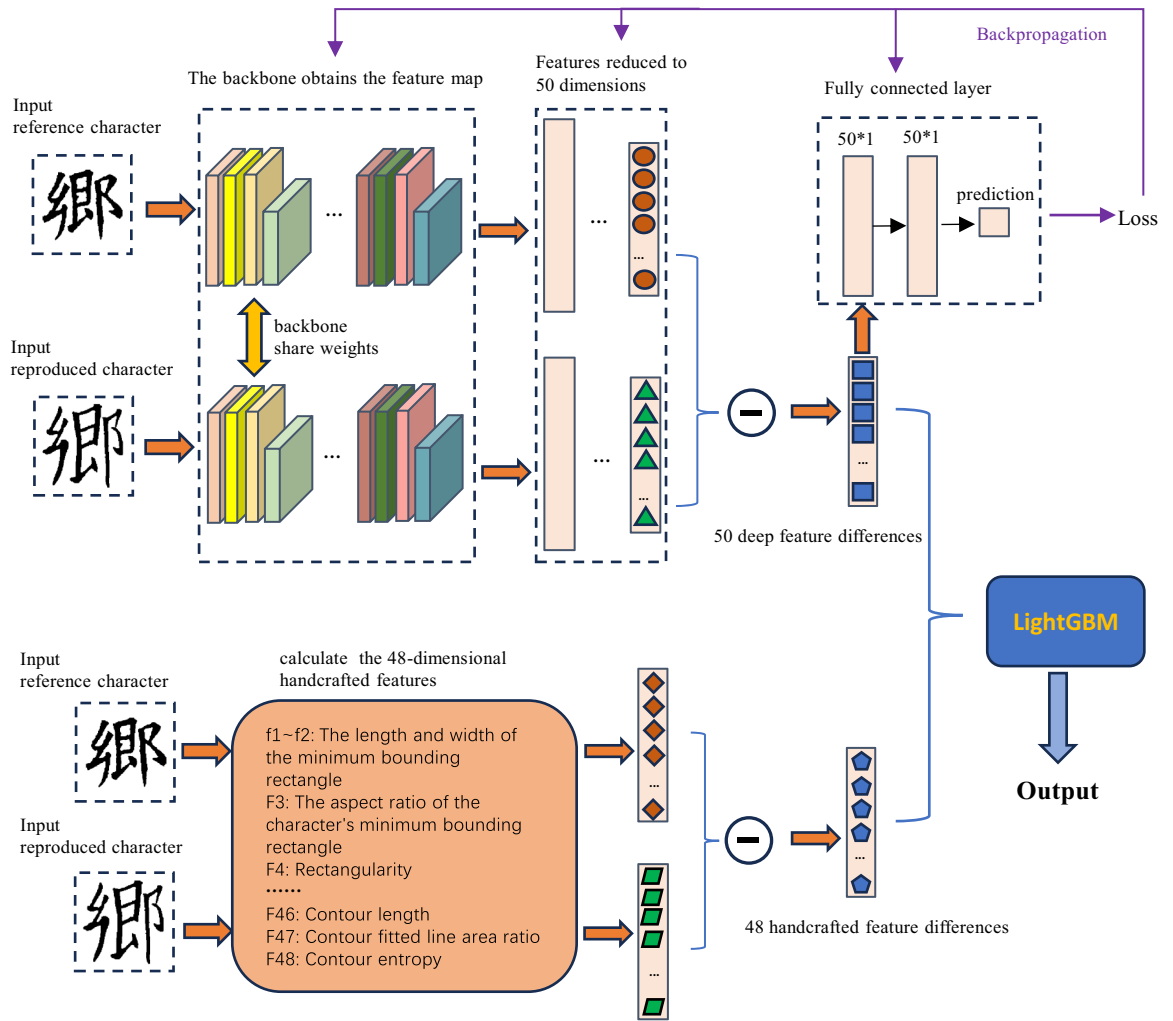
*h9**h10*: The slope and intercept of the fitted line. The least squares method is used to find the best-fit line for the character, providing the line's slope and intercept.

*h11**h14*: Mean, median, mode, and variance of the character's width. The width for each pixel row is defined as the distance from the leftmost to the rightmost black pixel.

*h15*: The number of enclosed spaces within the character.

*h16*: Proportion of the area occupied by enclosed spaces within the character. The formula is below:

$$h_{16} = \frac{C_n}{A}$$



**Fig. 6.** Prediction model architecture diagram.

where  $C_n$  represents the area of the enclosed regions, and  $A$  is the area of the convex hull.

$h17$ : Space utilization rate, defined as the ratio of the character's convex hull area to the area occupied by black pixels.

$h18h21$ : Variance of the pixel vertical projections of the character in four directions:  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$ .

$h22h25$ : Proportion of black pixel distribution across four quadrants.

$h26$ : Variance in the distribution ratios of black pixels across quadrants, calculated as the variance of  $h22h25$ .

$h27$ : Area ratio based on the division of black pixels by the fitted line. The fitted line divides the convex hull of black pixels into two regions, and the ratio of the smaller region to the larger region is calculated.

$h28h33$ : Coordinate positions of elastic grid lines. The ink strokes of the character are divided into four equal parts both vertically and horizontally using three lines, and the position of each line is calculated relative to the character's width or height.

$h34$ : Central region. The character's minimum bounding rectangle is evenly divided into a  $3 \times 3$  grid (9 cells in total), and the ratio of black pixels within the central cell to the total black pixels is calculated.

$h35$ : Average distance to the central axis. After fitting the central axis of the character, the average distance from all contour points of the character to this axis is calculated.

$h36$ : Number of stroke intersections.

$h37h38$ : Horizontal and vertical coordinates of the centroid within the character contour. The calculation formula is the same as calculating the centroid of the black pixel in the minimum bounding rectangle, with the area restricted to the character's contour rather than its minimum bounding rectangle.

$h39h41$ : Ratios of stroke initiation, progression, and termination areas. The character is evenly divided into three parts from left to right, and the proportions of black pixels in the first, second, and third parts relative to the total number of black pixels are calculated.

$h42$ : Texture energy, representing the uniformity of element values in the gray-level co-occurrence matrix (GLCM) of the character image. A higher value indicates a smoother character texture. The calculation formula is below:

$$h_{42} = \sum_{i,j} P(i,j)^2$$

Where  $P(i, j)$  is an element value in the GLCM.

$h_{43}$ : Texture contrast, indicating the intensity of gray-level variations in the character's GLCM. Higher values denote a rougher texture. The calculation formula is below:

$$h_{43} = \sum_{i,j} P(i,j)(i-j)^2$$

Where  $P(i, j)$  represents an element value in the GLCM.

$h_{44}$ : Texture homogeneity, reflecting the similarity in local gray-level distribution within the GLCM. A higher value indicates greater similarity in local gray levels. The calculation formula is below:

$$h_{44} = \sum_{i,j} \frac{P(i,j)}{1+(i-j)^2}$$

Where  $P(i, j)$  represents an element value in the GLCM.

$h_{45}$ : Number of endpoints within the character.

$h_{46}$ : Contour length, defined as the number of black pixels along the character contour.

$h_{47}$ : Contour fitted line area ratio. A line fitted to the character's contour divides it into two regions; the area ratio of the smaller region to the larger region is calculated.

$h_{48}$ : Contour entropy. The contour directions are assigned numbers from 0 to 7, corresponding to angles of  $0^\circ, 45^\circ, \dots, 315^\circ$ , to represent the contour sequence. Contour entropy is calculated using the formula below:

$$h_{48} = - \sum_{i=0}^7 p(i) \log_2 p(i)$$

where  $p(i)$  represents the probability of each assigned number  $i$  occurring.

### Feature difference fusion and LightGBM regression prediction

LightGBM<sup>42</sup> is built on the gradient-boosting decision tree framework and uses a histogram-based optimization algorithm, making it well-suited for large-scale data handling and the creation of high-quality predictive models.

In the prediction tasks for the reproduction skill and each dimension of EI, the 50-dimensional deep feature differences between the reference and reproduced characters are extracted using the Siamese network, respectively. These differences are then combined with 48-dimensional handcrafted feature differences. Afterward, regression prediction is performed using LightGBM, resulting in the model's predictive scores for the reproduction skill and each dimension of EI.

### Evaluation indicators

This study selects Mean Absolute Error (MAE) as the loss function and primary evaluation indicator. With its intuitive and easily computed characteristics, MAE serves as an ideal metric to measure the deviation between predicted values and actual values. A lower MAE value indicates that the model can more accurately capture the subtle differences between the reference and reproduced characters, thereby achieving precise assessment. The formula for MAE is as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

where  $N$  represents the number of samples,  $\hat{y}_i$  is the predicted value for the  $i$ -th sample, and  $y_i$  is the actual value. In this study, MAE measures the difference between the predicted results and the actual z-scores, with a smaller MAE value indicating better model performance.

As an additional evaluation metric, the Mean Squared Error (MSE) was selected due to its quadratic computation mechanism that induces disproportionate amplification of substantial prediction deviations, thereby yielding a sensitive indicator of model predictive capability in critical error domains. The formula for MSE is as follows:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

Additionally, this study introduces the Pearson Correlation Coefficient (PCC) as an evaluation metric to verify the linear correlation between predicted and actual values. To validate the robustness of Pearson correlation coefficient (PCC) computations, 95% confidence intervals were rigorously calculated and systematically integrated into the tabular presentation of results, constituting a statistical verification framework for measurement reliability. The formula for PCC is as follows:

Evaluation Indicator	VGG16	VGG19	Resnet50	Resnet101	Densenet121
MAE	<b>0.390</b>	0.398	0.529	0.563	0.596
MSE	<b>0.251</b>	0.265	0.461	0.525	0.567
PCC	<b>0.770</b>	0.755	0.575	0.498	0.523
95% CI of PCC	<b>[0.758, 0.781]</b>	[0.743, 0.767]	[0.556, 0.593]	[0.477, 0.519]	[0.502, 0.543]

**Table 2.** The mean absolute error(MAE), Mean Squared Error (MSE), Pearson correlation(PCC) and 95% confidence intervals comparison of Siamese networks for predicting reproduction skill scores under different backbone networks in the test set. The lower the MAE and MSE, the smaller the difference between the predicted value and the measured value; The higher the PCC, the higher the linear correlation between the predicted value and the measured value.

Prediction Dimension	Exp	Deep	Handcrafted	LightGBM	MAE	MSE	PCC	95% CI of PCC
Reproduction Skill	1	✓	✗	✗	0.390	0.251	0.770	[0.758, 0.781]
	2	✓	✗	✓	0.384	0.248	0.771	[0.759, 0.782]
	3	✗	✓	✓	0.437	0.340	0.696	[0.658, 0.689]
	4	✓	✓	✓	<b>0.378</b>	<b>0.244</b>	<b>0.776</b>	<b>[0.765, 0.787]</b>

**Table 3.** Ablation experiments verify the high efficiency of the model in predicting the reproduction skill score.

$$PCC = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Where  $x_i$  and  $y_i$  are the predicted and actual values of the  $i$ -th sample,  $\bar{x}$  and  $\bar{y}$  are their mean values.

## Experiment

All network training and testing in this study were performed in an environment equipped with an NVIDIA GTX 1080 Ti and PyTorch 1.8. The dataset was divided into training, validation, and test sets in an 8:1:1 ratio. VGG16, VGG19<sup>43</sup>, ResNet50, ResNet101<sup>44</sup>, and DenseNet121<sup>45</sup> were used as the backbone networks for the Siamese network.

VGG is commonly utilized due to its straightforward architecture and strong feature extraction capabilities. ResNet effectively addresses the vanishing gradient problem through the use of residual connections. DenseNet improves feature reuse and enhances information flow efficiency via dense connections.

To ensure comprehensive experimentation, comparative experiments were conducted with these networks to identify the optimal backbone network. The training process employed the Adam optimizer with a learning rate of  $1e-5$ .

To ensure a fair comparison environment for the selected backbone networks, we refrain from using pre-trained models. Prior to the commencement of training, the parameters of the backbone networks are randomized while preserving their core architectures without modification. Only the output layer is adjusted according to the aforementioned design to reduce its dimensionality to 50 features.

## Model performance in predicting the reproduction skill score

Predicting the reproduction skill score is closely aligned with our understanding of calligraphy. This prediction serves as an objective benchmark for our research. By comparing the model's ability to predict both EI dimensions and the reproduction skill score, we can assess its effectiveness across different tasks and evaluate its accuracy in predicting EI dimensions.

Table 2 presents the performance of the Siamese network in predicting the reproduction skill score using different backbone networks. Notably, the Siamese network that employs VGG16 as its backbone exhibits the best performance in terms of Mean Absolute Error (MAE), Mean Squared Error (MSE), and Pearson Correlation Coefficient (PCC). This suggests that it has the smallest deviation from the actual measured values and demonstrates the strongest linear correlation with them.

After selecting VGG16 as the backbone network, we extracted the deep feature differences between the reference and reproduced characters using the Siamese network model. Next, we computed 48-dimensional handcrafted feature differences between them. To further validate the effectiveness of the proposed model, we conducted ablation experiments. As shown in Table 3, Exp 1 presents the prediction results from the optimal Siamese network. Exp 2 illustrates the predictions made by LightGBM based on the deep feature differences extracted by the Siamese network. Exp 3 details the LightGBM predictions based on the 48-dimensional handcrafted feature differences. Finally, Exp 4 showcases the proposed model from this study, which integrates both deep and handcrafted feature differences for LightGBM predictions.

Reference	Reproduced	Measured Value	Predicted Value	Reference	Reproduced	Measured Value	Predicted Value
閒	閒	-2.071	-1.453	以	以	0.329	-0.178
然	然	-1.771	-1.399	微	微	0.929	0.551
知	知	-1.171	-0.818	夷	夷	1.229	1.677
達	達	-0.871	-0.275	園	園	1.829	1.242
想	想	-0.294	-0.874	興	興	2.428	1.780

**Table 4.** 10 pairs of samples were selected in the test set to show the prediction results of the proposed model for the reproduction skill scores.

Dimension	VGG16	VGG19	Resnet50	Resnet101	Densenet121
EI	<b>0.476</b>	0.515	0.668	0.669	0.658
Well-being	<b>0.470</b>	0.601	0.723	0.718	0.703
Self-control	<b>0.481</b>	0.550	0.667	0.660	0.656
Emotionality	<b>0.466</b>	0.602	0.731	0.729	0.689
Sociability	<b>0.473</b>	0.669	0.624	0.619	0.611

**Table 5.** Comparison of MAE values of EI dimensions predicted by Siamese Networks under different backbone networks in the test set.

Dimension	VGG16	VGG19	Resnet50	Resnet101	Densenet121
EI	<b>0.469</b>	0.533	0.950	0.989	0.955
Well-being	<b>0.405</b>	0.625	0.860	0.911	0.836
Self-control	<b>0.479</b>	0.591	0.990	0.950	0.926
Emotionality	<b>0.410</b>	0.647	1.029	1.069	0.853
Sociability	<b>0.440</b>	0.885	0.787	0.801	0.777

**Table 6.** Comparison of MSE values of EI dimensions predicted by Siamese Networks under different backbone networks in the test set.

The results indicate that Exp 4 achieved an MAE of 0.378, MSE of 0.244 and a PCC of 0.776, outperforming both Exp 2 (MAE 0.384, MSE 0.248, PCC 0.771) and Exp 3 (MAE 0.437, MSE 0.340, PCC 0.696). This improvement highlights the benefits of combining deep and handcrafted feature differences.

Furthermore, when comparing Exp 2 to Exp 1, there was a slight decrease in MAE, MSE and a minor increase in PCC, which demonstrates the effectiveness of the deep feature differences extracted by the Siamese network in LightGBM regression predictions. Additionally, the comparison between Exp 4 and Exp 3 shows that the deep feature differences effectively address the limitations of handcrafted features, allowing LightGBM to utilize these combined features more effectively for predictions.

10 pairs of reference and reproduced characters were selected based on their increasing z-scores for the reproduction skill. Table 4 displays the model's prediction results. As shown, the predicted values closely align with the measured scores, further confirming the method's excellent performance and its ability to generalize in predicting reproduction skill scores.

### Model performance in predicting EI dimensions

Tables 5, 6 and 7 demonstrate that when utilizing different backbone networks in the Siamese network to predict EI and its sub-dimensions, VGG16 outperforms VGG19, ResNet50, ResNet101 and DenseNet121 across all EI dimensions. It achieves the lowest Mean Absolute Error (MAE), Mean Squared Error (MSE), and the highest

Dimension	VGG16	VGG19	Resnet50	Resnet101	Densenet121
EI	<b>0.725 [0.712, 0.738]</b>	0.678 [0.662, 0.692]	0.315 [0.289, 0.340]	0.306 [0.280, 0.331]	0.338 [0.313, 0.363]
Well-being	<b>0.767 [0.755, 0.778]</b>	0.606 [0.588, 0.623]	0.392 [0.368, 0.416]	0.356 [0.331, 0.380]	0.455 [0.433, 0.477]
Self-control	<b>0.750 [0.737, 0.762]</b>	0.661 [0.645, 0.676]	0.283 [0.257, 0.309]	0.340 [0.315, 0.364]	0.358 [0.333, 0.382]
Emotionality	<b>0.785 [0.773, 0.795]</b>	0.619 [0.601, 0.636]	0.292 [0.266, 0.317]	0.304 [0.278, 0.329]	0.465 [0.443, 0.487]
Sociability	<b>0.707 [0.692, 0.720]</b>	0.021 [-0.015, 0.041]	0.360 [0.335, 0.384]	0.321 [0.295, 0.346]	0.349 [0.324, 0.374]

**Table 7.** Comparison of PCC values and 95% confidence intervals of EI dimensions predicted by Siamese Networks under different backbone networks in the test set.

Dimension	Exp	Deep	Handcrafted	LightGBM	MAE	MSE	PCC	95% CI of PCC
EI	1	✓	✗	✗	0.476	0.469	0.725	[0.712, 0.738]
	2	✓	✗	✓	0.474	0.471	0.723	[0.711, 0.737]
	3	✗	✓	✓	0.570	0.591	0.687	[0.627, 0.659]
	4	✓	✓	✓	<b>0.463</b>	<b>0.462</b>	<b>0.730</b>	<b>[0.717, 0.743]</b>
Well-being	1	✓	✗	✗	0.470	0.405	0.767	[0.755, 0.778]
	2	✓	✗	✓	0.463	0.405	0.767	[0.756, 0.779]
	3	✗	✓	✓	0.579	0.589	0.680	[0.632, 0.665]
	4	✓	✓	✓	<b>0.451</b>	<b>0.396</b>	<b>0.773</b>	<b>[0.762, 0.785]</b>
Self-control	1	✓	✗	✗	0.481	0.479	0.750	[0.737, 0.762]
	2	✓	✗	✓	0.474	0.480	0.748	[0.735, 0.760]
	3	✗	✓	✓	0.542	0.574	0.712	[0.674, 0.703]
	4	✓	✓	✓	<b>0.469</b>	<b>0.474</b>	<b>0.752</b>	<b>[0.741, 0.765]</b>
Emotionality	1	✓	✗	✗	0.466	0.410	0.785	[0.773, 0.795]
	2	✓	✗	✓	0.461	0.408	0.782	[0.772, 0.794]
	3	✗	✓	✓	0.590	0.625	0.683	[0.627, 0.659]
	4	✓	✓	✓	<b>0.452</b>	<b>0.405</b>	<b>0.786</b>	<b>[0.777, 0.798]</b>
Sociability	1	✓	✗	✗	0.473	0.440	0.707	[0.692, 0.720]
	2	✓	✗	✓	0.466	0.438	0.705	[0.691, 0.719]
	3	✗	✓	✓	0.517	0.529	0.675	[0.619, 0.652]
	4	✓	✓	✓	<b>0.455</b>	<b>0.429</b>	<b>0.712</b>	<b>[0.699, 0.726]</b>

**Table 8.** Ablation experiments verify the high efficiency of the model in predicting all dimensions of EI.

Pearson Correlation Coefficient (PCC) values. This suggests that using VGG16 as the backbone network results in predictions with the smallest error and the strongest linear correlation to the measured scores.

After selecting VGG16 as the backbone network, we extracted the deep feature differences between the reference characters and reproduced characters across EI dimensions. These were then fused with the 48-dimensional handcrafted feature differences to make the final predictions using LightGBM. The results from the ablation experiments are presented in Table 8. Exp 1 showcases the prediction results from the optimal Siamese network. Exp 2 displays the LightGBM predictions based on deep feature differences, while Exp 3 presents LightGBM predictions based on handcrafted feature differences. Finally, Exp 4 shows the results of our model, which combines both deep and handcrafted feature differences for LightGBM predictions.

The results indicate that Exp 4 outperformed both Exp 2 and Exp 3 in all dimensions of EI, further confirming the superiority of the proposed model. Additionally, when compared to Exp 1, Exp 2 showed improved prediction performance across all dimensions. This implies that the deep feature differences extracted by the Siamese network are effective, resulting in reduced information loss in LightGBM regression predictions. Furthermore, Exp 4 demonstrated significantly better predictive performance than Exp 3 in all EI dimensions. This suggests that the deep feature differences can compensate for the limitations of handcrafted features, allowing LightGBM to fully utilize their complementarity, thereby enhancing the ability to predict each EI dimension.

Table 9 presents 5 pairs of reference and reproduced characters selected from the test set, ordered by ascending EI scores to illustrate the model's final prediction results for each dimension of EI.

Table 10 illustrates the selection of various reproduced characters from the same reference character for each EI dimension. These characters were categorized into low-level and high-level groups based on a z-score of 0. This categorization visually highlights the differences in reproduced results across the EI dimensions and demonstrates the effectiveness of the model's predictions.

### Comparison of model prediction ability with human prediction ability

To compare the predictive capabilities of humans and models, our core approach involves randomly selecting a subset of handwriting imitation samples from the test set. A group of 15 postgraduate students were recruited to

Reference	Reproduced		EI	Well-being	Self-control	Emotionality	Sociability
密	密	Measured	-1.734	-3.202	-2.320	1.288	-0.677
		Predicted	-1.272	-2.960	-1.266	1.266	-0.382
想	想	Measured	-0.957	-0.802	-0.489	-1.103	-0.381
		Predicted	-0.472	-0.152	-0.404	-1.018	0.099
如	如	Measured	0.320	0.490	-0.489	0.172	-0.085
		Predicted	-0.021	0.501	-0.531	-0.133	-0.399
新	新	Measured	1.097	1.413	0.296	0.331	1.395
		Predicted	0.661	0.154	-0.087	0.085	0.566
在	在	Measured	2.428	1.597	1.342	2.404	2.580
		Predicted	1.881	1.498	0.611	1.655	1.809

**Table 9.** 10 pairs of samples were selected in the test set to show the prediction results of the proposed model for each dimension score of EI.

serve as human evaluators, all of whom are first- or second-year students in science and engineering disciplines. While they possess varying levels of exposure to calligraphy, none have received formal or prolonged training in fine arts, psychology, or handwriting analysis. Thus, they can be regarded as representative non-expert evaluators with typical calligraphic aesthetic sensibilities. Each evaluator was asked to subjectively assess the reproduction ability and trait EI of the imitators by comparing their works to exemplar scripts. All evaluators first completed a self-assessment of their Trait EI using a standardized scale, followed by a 15-minute focused training session to ensure a clear understanding of the concept. They then assessed the same set of 48 test samples using a 7-point Likert scale, with the sample order randomized for each evaluator to control for order effects. Simultaneously, the best-performing model from our experiment conducted the same evaluations, enabling a comparative analysis between human and model assessments.

The Fleiss' Kappa coefficients and corresponding p-values for all dimensions, as evaluated by the 15 human evaluators, are presented in Table 11. Values of Fleiss' Kappa approaching 1 indicate stronger inter-rater agreement. Notably, the consistency of ratings across EI dimensions was systematically lower than that observed for handwriting reproduction skill scores. This finding is statistically meaningful, as it suggests that raters could not reliably infer a writer's trait EI from reproduction works. For predictive accuracy assessment, complementary metrics—including MAE and PCC—must be integrated into the analysis.

Mean Absolute Error (MAE) and Pearson Correlation Coefficient (PCC) were calculated for human evaluations in each dimension. The comparison results, presented in Table 12, show that the model's predictions for the reproduction skill score significantly exceed those made by humans. Additionally, humans struggled to accurately predict the EI dimensions from calligraphic characters. This finding further highlights the value of this model in predicting EI.

### Contribution analysis across all dimensions of handcrafted features in predictive modeling

As demonstrated in Tables 3, 8 and 12 the handcrafted feature-based model (exp 3), although exhibiting inferior performance compared to feature fusion models (exp 4) and neural network architectures (exp 1), still surpasses human evaluation capabilities in predictive accuracy. Nevertheless, the specific handcrafted features contributing substantially to predictions across different data dimensions remain undetermined. Exp 3, a LightGBM-based model constructed using handcrafted features, was subjected to SHAP analysis to quantify feature contribution patterns across diverse data dimensions.

SHAP (SHapley Additive exPlanations)<sup>46</sup> is a model interpretation method grounded in cooperative game theory's Shapley values. Its fundamental principle involves calculating features' marginal contributions across all possible feature subset combinations, subsequently deriving feature importance through weighted averaging. The mathematical formulation is expressed as:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f(S \cup \{i\}) - f(S)]$$

Dimension	Reference		Reproduced1	Reproduced2	Reproduced3	
EI	想	High-level	想	想	想	
			Measured	1.818	1.097	0.653
		Low-level	Predicted	1.720	0.591	0.963
			想	想	想	
		High-level	Measured	-1.401	-1.068	-0.735
			Predicted	-1.102	-0.707	-0.358
Well-being	都	High-level	都	都	都	
			Measured	2.520	1.228	0.305
		Low-level	Predicted	1.828	0.741	0.198
			都	都	都	
		High-level	Measured	-1.356	-0.987	-0.617
			Predicted	-0.644	-0.555	-0.371
Self-control	分	High-level	分	分	分	
			Measured	1.081	0.819	0.558
		Low-level	Predicted	0.554	0.662	0.353
			分	分	分	
		High-level	Measured	-1.536	-1.012	-0.751
			Predicted	-0.945	-0.783	-0.423
Emotionality	新	High-level	新	新	新	
			Measured	1.926	1.447	0.810
		Low-level	Predicted	0.476	0.909	0.600
			新	新	新	
		High-level	Measured	-1.901	-1.422	-0.785
			Predicted	-1.116	-0.825	-0.371
Sociability	度	High-level	度	度	度	
			Measured	1.691	1.395	1.099
		Low-level	Predicted	1.097	0.867	0.662
			度	度	度	
		High-level	Measured	-1.270	-0.677	-0.381
			Predicted	-0.424	-0.298	-0.223

**Table 10.** Under the same EI dimension, when reproducing the same reference character, the reproduced character samples and predicted scores with different levels are displayed.

	Reproduction skill	EI	Well-being	Self-control	Emotionality	Sociability
Fleiss's Kappa	0.031	0.015	0.005	0.029	0.001	-0.011
p Value	<0.001	0.043	0.48	<0.001	0.935	0.122

**Table 11.** Fleiss' Kappa coefficients were computed to assess agreement among the 15 raters across all evaluation dimensions. The Kappa statistic ranges from -1 to 1, where values approaching 1 indicate near-perfect agreement, while those tending toward -1 signify systematic disagreement among raters.

Dimension	MAE		MSE		PCC	
	Human	Model	Human	Model	Human	Model
Reproduction Skill	0.507	<b>0.378</b>	0.463	<b>0.244</b>	0.574 [0.346, 0.738]	<b>0.776 [0.765, 0.787]</b>
EI	1.006	<b>0.463</b>	1.740	<b>0.462</b>	0.145 [-0.145, 0.412]	<b>0.730 [0.717, 0.743]</b>
Well-being	0.734	<b>0.451</b>	0.931	<b>0.396</b>	0.168 [-0.122, 0.432]	<b>0.773 [0.762, 0.785]</b>
Self-control	1.089	<b>0.469</b>	2.092	<b>0.474</b>	0.001 [-0.284, 0.285]	<b>0.752 [0.741, 0.765]</b>
Emotionality	0.894	<b>0.452</b>	1.274	<b>0.405</b>	0.086 [-0.203, 0.361]	<b>0.786 [0.777, 0.798]</b>
Sociability	0.869	<b>0.455</b>	1.265	<b>0.429</b>	0.234 [-0.054, 0.486]	<b>0.712 [0.699, 0.726]</b>

**Table 12.** Comparison of the performance of this model's predictions with human predictions in all data dimensions.

	Reproduction Skill	EI	Well-being	Self-control	Emotionality	Sociability
1	h17 (0.473)	h23 (0.045)	h35 (0.058)	h21 (0.076)	h23 (0.038)	h3 (0.044)
2	h43 (0.007)	h21 (0.040)	h23 (0.045)	h35 (0.048)	h35 (0.038)	h21 (0.044)
3	h22 (0.006)	h16 (0.039)	h18 (0.036)	h24 (0.048)	h10 (0.034)	h4 (0.032)
4	h23 (0.005)	h35 (0.037)	h16 (0.035)	h44 (0.046)	h17 (0.034)	h42 (0.032)
5	h46 (0.004)	h24 (0.033)	h25 (0.034)	h23 (0.041)	h3 (0.033)	h17 (0.030)
6	h21 (0.004)	h3 (0.033)	h24 (0.033)	h16 (0.038)	h20 (0.032)	h9 (0.028)
7	h35 (0.003)	h2 (0.031)	h21 (0.033)	h20 (0.035)	h16 (0.032)	h22 (0.027)
8	h6 (0.003)	h18 (0.030)	h19 (0.031)	h43 (0.034)	h1 (0.031)	h25 (0.027)

**Table 13.** The top 8 contributing handcrafted features with corresponding SHAP values across predictive dimensions. The higher the absolute SHAP values, the stronger predictive influence.

$$f(x) = \phi_0 + \sum_{i=1}^M \phi_i$$

where  $F$  represents the complete feature set,  $S$  denotes a subset excluding feature  $G$ , and  $f$  corresponds to the model's prediction function. A distinctive property of SHAP lies in its additivity: the summation of SHAP values for all features equals the deviation between model predictions and baseline values ( $\phi_0$ ). For any given model, features with higher absolute SHAP values indicate stronger predictive dependence.

The top eight handcrafted features with highest predictive contributions across dimensions were systematically quantified (Table 13). For reproduction skill prediction, feature h17 (character space utilization rate), demonstrated predominant contributions, aligning with calligraphic principles of ink control precision and structural regularity. Emotional intelligence (EI) dimension predictions exhibited differentiated feature hierarchies with notable feature overlap. Recurrent predictors including h35 (mean distance from contour to central axis), h23 (pixel distribution in quadrant II), h16 (enclosed area ratio), and h21 (135° projection variance) demonstrated cross-dimensional significance, indicating their capacity to concurrently capture multi-dimensional emotional intelligence traits through subconscious writing behaviors. Systematic variations in these aesthetic characteristics across writers with distinct emotional intelligence profiles enabled precise trait EI level prediction through regression modeling, with feature value differentials serving as reliable discriminative markers.

## Discussion

This study investigated whether reproduced calligraphy, which is a structured, visually rich, and culturally meaningful form of motor expression, can be used to predict an individual's trait emotional intelligence (EI). By combining deep visual features derived from a Siamese network with handcrafted aesthetic features rooted in traditional calligraphy analysis, we developed a predictive model that significantly outperformed human evaluators across all five EI dimensions. These results support the long-standing notion that handwriting reflects psychological disposition and further show that computational models can extract and utilise latent emotional signals embedded in motor reproduction tasks.

Several points are worth highlighting. First, our computational model was able to predict trait EI with a high degree of accuracy, particularly for the global EI score (PCC = 0.730, MAE = 0.463), demonstrating both statistical robustness and psychological relevance. This performance was substantially better than that of human raters (PCC = 0.145, MAE = 1.006), who viewed the same calligraphic material. The results suggest that subtle visual-motor cues, imperceptible or inconsistently used by human judges, can be leveraged by machine learning systems to infer affective traits with considerable precision.

Second, the success of the model is attributable in part to the feature fusion strategy, which combined interpretable handcrafted features, such as stroke density, spatial balance, and symmetry, with non-linear, high-

dimensional features learned via a Siamese neural network. Notably, when comparing backbone architectures within the Siamese network, VGG16 consistently outperformed VGG19, ResNet50, ResNet101, and DenseNet121 across all EI dimensions. This may be attributed to the architectural simplicity of VGG16, including its moderate depth and smaller convolutional kernels, which proved particularly effective for capturing the fine-grained stylistic nuances embedded in calligraphic strokes, structure, and spatial layout. In contrast, the deeper or more complex networks imposed additional computational burden without improving predictive performance in this context.

Third, the use of LightGBM as the regression model was advantageous not only due to its predictive performance but also because of its compatibility with post hoc interpretability techniques such as SHAP (Shapley Additive Explanations). SHAP values provide additive, locally faithful explanations for individual predictions, enabling us to understand which features contributed most to each prediction. The SHAP analysis revealed that handcrafted features, particularly those encoding spatial symmetry, curvature, stroke density, and balance, exhibited dimension-specific predictive utility. However, some degree of convergence in feature importance across EI subdimensions was also observed, suggesting that trait EI may be encoded through a core set of discriminative visual-motor features. Interestingly, while handcrafted features alone (Experiment 3) yielded the best performance in predicting reproduction skill scores, their utility was limited when predicting EI dimensions. Conversely, combining handcrafted and deep features (Experiment 4) improved EI prediction beyond either feature set alone, indicating complementary contributions: deep features capture subtle structural deviations, while handcrafted features preserve psychologically interpretable stylistic cues. Among the EI subdimensions, Sociability exhibited the highest predictive accuracy (PCC = 0.622), while Self-control yielded the lowest (PCC = 0.497), potentially reflecting greater variability or reduced expressivity in this trait's manifestation through motor-reproductive features.

Fourth, although SHAP provided useful insight into model interpretability, the combined model still operates as a high-dimensional system where feature interactions and dependencies can obscure straightforward causal inference. Prior work has shown that Shapley-value-based explanations may misattribute importance when features are correlated or interact non-linearly<sup>47,48</sup>. Future work could explore causal feature perturbation, attention-based visualisation techniques, or symbolic regression approaches to disentangle these interdependencies and to provide more interpretable and potentially causal explanations of the relationships between calligraphic features and psychological traits.

Fifth, given the widespread availability of high-performance pretrained models based on VGG, ResNet, and ViT, continued pretraining or fine-tuning from these architectures may reveal that models beyond VGG possess greater potential, possibly achieving performance superior to that of the model presented in this study. From an engineering perspective, leveraging high-quality pretrained models as a foundation for further training is generally a more effective and practical strategy.

Finally, we compared the model's predictive performance with that of human raters. While human judges were able to infer limited aspects of EI from handwriting samples, their performance lagged significantly behind the model across all EI dimensions. This finding aligns with longstanding notions that “writing reflects the mind” and “handwriting reveals personality,” but it also demonstrates the added value of computational modeling in detecting latent affective traits that are not readily accessible to human intuition.

Despite the intriguing findings, several limitations must be acknowledged. First, the generalizability of our results is constrained by the homogeneity of sample: all our participants were young, right-handed, Chinese university students. It remains to be seen whether similar predictive accuracy could be achieved across different cultures, age groups, writing systems or levels of calligraphy proficiency. In particular, the cultural familiarity and implicit aesthetic norms associated with Chinese calligraphy may play a role in the observed associations. Therefore, an interesting future work is to extend this framework to diverse cultural contexts, including calligraphy practices in English, Arabic, Japanese, Korean or other languages. Such exploration is necessary to validate its broad applicability and generalizability across diverse cultural contexts. Second, while our model predicted self-reported trait EI scores, it is important to consider the limitations of self-report measures. Trait EI questionnaires are known to be subject to social desirability bias, self-perception inaccuracies, and context effects<sup>49</sup>. Thus, the model may reflect the participants' self-perceptions of emotional ability rather than objectively verifiable emotional competencies. Future work could include behavioral or physiological indicators of emotional regulation and expression to triangulate the results. Third, although we used a controlled reproduction task to reduce stylistic freedom, variability in attention, motivation, or fatigue may have influenced the motor output of participants. Without concurrent recordings (e.g., pen pressure, writing dynamics, or real-time affect), we cannot fully rule out state-related confounding factors. Moreover, while the model predicts trait scores based on aggregate data, it is not currently suitable for real-time or single-character inference, limiting its applicability to high-frequency, low-resource settings. Fourth, although SHAP provided useful insight into model interpretability, the combined model still operates as a high-dimensional system where feature interactions and dependencies can obscure straightforward causal inference. Indeed, prior work has shown that Shapley-value-based explanations may misattribute importance when features are correlated or when interaction effects exist, potentially misleading users about causal relationships<sup>51</sup>. To address this, future work could explore causal feature perturbation, attention-based visualization techniques, or symbolic regression approaches to disentangle interdependent contributions and enhance interpretability. Finally, the application of machine learning to infer psychological traits from biometric or artistic data raises significant ethical concerns. Although this study was conducted with full informed consent and academic intent, any real-world deployment must be overseen by rigorous ethical review. This includes robust attention to data privacy, mitigating algorithmic bias, and preventing misuse in evaluative contexts such as hiring or educational assessment. While interpretable models like LightGBM and SHAP offer avenues for transparency, effective deployment must occur within comprehensive responsible-AI frameworks—ensuring accountability, human oversight, and protection against potential harms.

## Conclusion

This study shows that an individual's trait emotional intelligence can be predicted with high accuracy from their reproduced calligraphy, using a hybrid model that combines deep visual representations with psychologically meaningful handcrafted features. LightGBM served not only as an effective predictive engine but also enabled interpretable modeling via SHAP analysis, revealing how specific visual-motor properties relate to distinct affective traits. The findings support the long-standing notion that handwriting reveals psychological traits and offer a new computational approach for exploring the embodied expression of emotion and personality. Further, the findings also contribute to the emerging intersection of affective computing, personality psychology, and computational aesthetics, suggesting new avenues for non-verbal psychological assessment. Future work should aim to extend this framework across cultures and modalities, while ensuring ethical transparency and scientific rigor in applied contexts.

## Data availability

The data set of this study can be downloaded from the website: <https://zenodo.org/records/14166945>

Received: 17 November 2024; Accepted: 23 July 2025

Published online: 06 August 2025

## References

- Petrides, K. V. & Furnham, A. Trait emotional intelligence: Psychometric investigation with reference to established trait taxonomies. *Eur. journal personality* **15**, 425–448. <https://doi.org/10.1002/per.416> (2001).
- Petrides, K. V., Pita, R. & Kokkinaki, F. The location of trait emotional intelligence in personality factor space. *Br. journal psychology* **98**, 273–289. <https://doi.org/10.1348/000712606X120618> (2007).
- Petrides, K. V. Psychometric properties of the trait emotional intelligence questionnaire (teique). In *Assessing emotional intelligence: Theory, research, and applications*, 85–101, [https://doi.org/10.1007/978-0-387-88370-0\\_5](https://doi.org/10.1007/978-0-387-88370-0_5) (Springer, 2009).
- Conte, J. M. A review and critique of emotional intelligence measures. *J. organizational behavior* **26**, 433–440. <https://doi.org/10.1002/job.319> (2005).
- O'Connor, P. J., Hill, A., Kaya, M. & Martin, B. The measurement of emotional intelligence: A critical review of the literature and recommendations for researchers and practitioners. *Front. psychology* **10**, 1116. <https://doi.org/10.3389/fpsyg.2019.01116> (2019).
- Freudenthaler, H. H. & Neubauer, A. C. Emotional intelligence: The convergent and discriminant validities of intra- and interpersonal emotional abilities. *Pers. Individ. Differ.* **39**, 569–579. <https://doi.org/10.1016/j.paid.2005.02.004> (2005).
- Vernon, P. E. The rorschach ink-blot test 1. i. *Br. J. Med. Psychol.* **13**, 90–118, <https://doi.org/10.1111/j.2044-8341.1933.tb01094.x> (1933).
- Murray, H. Thematic apperception test (1943).
- Sheehan, K. M. & Mislevy, R. J. An inquiry into the nature of the sentence-completion task: Implications for item generation. *Ets Res. Rep.* **2001**, i 51 (2001).
- Aulia, M. R., Djamal, E. C. & Bon, A. T. Personality identification based on handwritten signature using convolutional neural networks. In *Proceedings of the 5th NA International Conference on Industrial Engineering and Operations Management Detroit*, 10–14, <https://doi.org/10.1109/ic-ETITE58242.2024.10493494> (2020).
- Maliki, I. & Sidik, M. Personality prediction system based on signatures using machine learning. In *IOP Conference Series: Materials Science and Engineering* **879**, 012068. <https://doi.org/10.1088/1757-899X/879/1/012068> (IOP Publishing 2020).
- Raimo, G. et al. Handwriting and drawing for depression detection: a preliminary study. In *International Conference on Applied Intelligence and Informatics*, 320–332, [https://doi.org/10.1007/978-3-031-24801-6\\_23](https://doi.org/10.1007/978-3-031-24801-6_23) (Springer, 2022).
- Nolazco-Flores, J. A., Faundez-Zanuy, M., Velázquez-Flores, O. A., Cordasco, G. & Esposito, A. Emotional state recognition performance improvement on a handwriting and drawing task. *IEEE Access* **9**, 28496–28504. <https://doi.org/10.1109/ACCESS.2021.3058443> (2021).
- Gavrilescu, M. & Vizireanu, N. Predicting the big five personality traits from handwriting. *EURASIP J. on Image and Video Processing* **1–17**, 2018. <https://doi.org/10.1186/s13640-018-0297-3> (2018).
- Smith, C. A. & Lazarus, R. S. Appraisal components, core relational themes, and the emotions. *Cogn. & emotion* **7**, 233–269. <https://doi.org/10.1080/02699939308409189> (1993).
- Petrides, K. V. & Furnham, A. Trait emotional intelligence: Behavioural validation in two studies of emotion recognition and reactivity to mood induction. *Eur. journal personality* **17**, 39–57. <https://doi.org/10.1002/per.466> (2003).
- Karimi, L. et al. Emotional intelligence: predictor of employees wellbeing, quality of patient care, and psychological empowerment. *BMC psychology* **9**, 93. <https://doi.org/10.1186/s40359-021-00593-8> (2021).
- Austin, E. J., Saklofske, D. H. & Egan, V. Personality, well-being and health correlates of trait emotional intelligence. *Pers. Individ. Differences* **38**, 547–558. <https://doi.org/10.1016/j.paid.2004.05.009> (2005).
- Bru-Luna, L. M., Martí-Vilar, M., Merino-Soto, C. & Cervera-Santiago, J. L. Emotional intelligence measures: A systematic review. In *Healthcare* **9**, 1696. <https://doi.org/10.3390/healthcare9121696> (2021) (MDPI).
- Schutte, N. S. et al. Development and validation of a measure of emotional intelligence. *Pers. individual differences* **25**, 167–177. [https://doi.org/10.1016/S0191-8869\(98\)00001-4](https://doi.org/10.1016/S0191-8869(98)00001-4) (1998).
- Siegling, A., Petrides, K. & Martskvishvili, K. An examination of a new psychometric method for optimizing multi-faceted assessment instruments in the context of trait emotional intelligence. *Eur. journal personality* **29**, 42–54. <https://doi.org/10.1002/per.197> (2015).
- Buck, J. N. The htp test. *J. Clin. psychology* **4** (1948).
- Handler, L. The clinical use of drawings. *Major psychological assessment instruments* **2**, 206–293 (1996).
- Guo, H. et al. Analysis of the screening and predicting characteristics of the house-tree-person drawing test for mental disorders: A systematic review and meta-analysis. *Front. Psychiatry* **13**, 1041770. <https://doi.org/10.3389/fpsyg.2022.1041770> (2023).
- Dwyer, D. B., Falkai, P. & Koutsouleris, N. Machine learning approaches for clinical psychology and psychiatry. *Annu. review of clinical psychology* **14**, 91–118. <https://doi.org/10.1146/annurev-clinpsy-032816-045037> (2018).
- Mekhaznia, T., Djeddi, C. spsamps Sarkar, S. Personality traits identification through handwriting analysis. In *Pattern Recognition and Artificial Intelligence: 4th Mediterranean Conference, MedPRAI 2020, Hammamet, Tunisia, December 20–22, 2020, Proceedings* **4**, 155–169, [https://doi.org/10.1007/978-3-030-71804-6\\_12](https://doi.org/10.1007/978-3-030-71804-6_12) (Springer, 2021).
- Zeng, H. et al. A lightgbm-based eeg analysis method for driver mental states classification. *Comput. intelligence and neuroscience* **2019**, 3761203. <https://doi.org/10.1155/2019/3761203> (2019).
- Li, R. spsamps Dai, J. Emotion prediction and analysis of weibo users combined with portraits. In *2022 3rd International Conference on Management Science and Engineering Management (ICMSEM 2022)*, 292–300, [https://doi.org/10.2991/978-94-6463-038-1\\_27](https://doi.org/10.2991/978-94-6463-038-1_27) (Atlantis Press, 2022).

29. Dollmat, K. S. & Abdullah, N. A. Machine learning in emotional intelligence studies: a survey. *Behav. & Inf. Technol.* **41**, 1485–1502. <https://doi.org/10.1080/0144929X.2021.1877356> (2022).
30. Zhang, J., Miao, Y., Zhang, J. & Yu, J. Inkthetics: a comprehensive computational model for aesthetic evaluation of chinese ink paintings. *IEEE Access* **8**, 225857–225871. <https://doi.org/10.1109/ACCESS.2020.3044573> (2020).
31. Yan, F., Lan, X., Zhang, H. & Li, L. Intelligent evaluation of chinese hard-pen calligraphy using a siamese transformer network. *Appl. Sci.* **14**, 2051. <https://doi.org/10.3390/app14052051> (2024).
32. Sun, M., Gong, X., Nie, H., Iqbal, M. M. & Xie, B. Srafe: Siamese regression aesthetic fusion evaluation for chinese calligraphic copy. *CAAI Transactions on Intell. Technol* **8**, 1077–1086. <https://doi.org/10.1049/cit2.12095> (2023).
33. Cooper, A. & Petrides, K. V. A psychometric analysis of the trait emotional intelligence questionnaire-short form (teique-sf) using item response theory. *J. personality assessment* **92**, 449–457. <https://doi.org/10.1080/00223891.2010.497426> (2010).
34. Bertinetto, L., Valmadre, J., Henriques, J. F., Vedaldi, A. sampsamps Torr, P. H. Fully-convolutional siamese networks for object tracking. In *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8–10 and 15–16, 2016, Proceedings, Part II 14*, 850–865. [https://doi.org/10.1007/978-3-319-48881-3\\_56](https://doi.org/10.1007/978-3-319-48881-3_56) (Springer, 2016).
35. Li, C. & Chen, T. Aesthetic visual quality assessment of paintings. *IEEE J. selected topics in Signal Processing* **3**, 236–252. <https://doi.org/10.1109/JSTSP.2009.2015077> (2009).
36. Mallon, B., Redies, C. & Hayn-Leichsenring, G. U. Beauty in abstract paintings: perceptual contrast and statistical properties. *Front. human neuroscience* **8**, 161. <https://doi.org/10.3389/fnhum.2014.00161> (2014).
37. Sartori, A. et al. Affective analysis of professional and amateur abstract paintings using statistical analysis and art theory. *ACM Transactions on Interact. Intell. Syst. (TiiS)* **5**, 1–27. <https://doi.org/10.1145/2768209> (2015).
38. Rajan, V. How handwriting evolves: an initial quantitative analysis of the development of indic scripts. In *17th Biennial Conference of the International Graphonomics Society* (2015).
39. Rajan, V. Quantifying scripts: Defining metrics of characters for quantitative and descriptive analysis. *Digit. Scholarsh. Humanit.* **32**, 602–631. <https://doi.org/10.1093/lc/fqw030> (2017).
40. Xiao, X. & Li, C. Research progress on evaluation methods of handwritten chinese characters. *Comput. Eng. Appl* **58**, 27–42 (2022).
41. Xuexiong, D., Jingtao, L. & Mu, L. Comouter evaluation of imitation in chinese calligraphy. *J. Graph.* **35**, 899–904 (2014).
42. Ke, G. et al. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. in neural information processing systems* **30** (2017).
43. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)<https://doi.org/10.48550/arXiv.1409.1556> (2014).
44. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778. <https://doi.org/10.1109/CVPR.2016.90> (2016).
45. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708. <https://doi.org/10.1109/CVPR.2017.243> (2017).
46. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. *Adv. in neural information processing systems* **30** (2017).
47. Verdinelli, I. & Wasserman, L. Feature importance: A closer look at shapley values and loco. *Stat. Sci.* **39**, 623–636 (2024).
48. Basu, I. & Maji, S. Multicollinearity correction and combined feature effect in shapley values. arxiv (2011).
49. Day, A. L. & Carroll, S. A. Faking emotional intelligence (ei): comparing response distortion on ability and trait-based ei measures. *J. Organ. Behav.* **29**, 761–784. <https://doi.org/10.1002/job.485> (2008). <https://onlinelibrary.wiley.com/doi/pdf/10.1002/job.485>.
50. Tett, R. P., Freund, K. A., Christiansen, N. D., Fox, K. E. & Coaster, J. Faking on self-report emotional intelligence and personality tests: Effects of faking opportunity, cognitive ability, and job type. *Pers. Individ. Differ.* **52**, 195–201. <https://doi.org/10.1016/j.paid.2011.10.017> (2012).
51. Kumar, I. E., Venkatasubramanian, S., Scheidegger, C. & Friedler, S. Problems with shapley-value-based explanations as feature importance measures. In III, H. D. & Singh, A. (eds.) *Proceedings of the 37th International Conference on Machine Learning*, vol. 119 of *Proceedings of Machine Learning Research*, 5491–5500 (PMLR, 2020).

## Author contributions

R.L. designed experiments, organized data collection, guided the overall experimental ideas, acted as an expert in the reproduction skill scoring, and assisted in writing papers. W.S. designed and organized experiments, collected data, wrote programs, and wrote papers. Y.C. designed a program with 48 dimensions of handcrafted features. Y.S. organized the collection of data and acted as an expert in the reproduction skill scoring. N.W. acted as an expert in the reproduction skill scoring. J.B. shared the knowledge in the field of trait emotional intelligence, and assisted in writing papers. G.Y. shared the knowledge in the field of trait emotional intelligence, and organized the project. All authors reviewed the manuscript.

## Declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to G.Y.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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

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

© The Author(s) 2025